A Textual Analysis of Bank of England Growth Forecasts

Jacob T. Jones, Tara M. Sinclair, and Herman O. Stekler
The George Washington University

Original Version: November 15, 2018

This Version Updated on May 2, 2019

A Textual Analysis of Bank of England Growth Forecasts [1]

Jacob T. Jones, Tara M. Sinclair[2] and Herman O. Stekler
The George Washington University
Department of Economics and Research Program on Forecasting
Washington, DC 20037

This Version:
May 2, 2019

Abstract: The Bank of England publishes a quarterly Inflation Report that provides numerical forecasts and text discussion of its assessment of the UK economy. Previous research has evaluated the *quantitative* forecasts included in these Reports, but we focus on the *qualitative* discussion of output growth. We use an in-sample textual analysis procedure to convert these qualitative assessments into a score for each Report for the period 2005-2014. We also construct out-of-sample scores for Reports before and after this period. We compare the scores to real-time output growth data as well as to the corresponding quantitative projections published by the Bank. We find that overall developments in the UK economy were accurately represented in the text of the Inflation Report. Furthermore, efficiency regressions suggest that there is information in the text that could improve the Bank of England's quantitative nowcasts and one-quarter ahead forecasts.

Keywords: Forecast Evaluation, Qualitative Forecasting, Great Recession

JEL Classification: C53, E37, E58

---

1. Introduction

Evaluation of economic forecasts has typically concentrated on the quality of numerical forecasts. Focusing on the quantitative forecasts neglects the substantial amount of text that often accompanies them, particularly in publications released by central banks. The text is included to provide context and nuance which can also be evaluated using textual analysis. Furthermore, numerical forecasts are not always available to the public, but text in reports, meeting minutes, and speeches may reveal insights on the assessment of the current and future economic conditions.

The literature has established the value of textual analysis, as well as a general methodology for converting text into quantitative scores that can be compared to numerical forecasts and realized data. Goldfarb et al. (2005) developed a scoring procedure to create an index of qualitative forecasts made during the Great Depression based on a textual analysis of contemporary newspaper articles. Lundquist and Stekler (2012) applied the Goldfarb et al. procedure to evaluate qualitative assessments made by business economists during the Great Recession. More recently, Stekler and Symington (2016) used the procedure to examine the minutes of Federal Open Market Committee (FOMC) meetings, and Mathy and Stekler (2018) adjusted the scoring procedure to incorporate outlooks that are more negative. They also proposed an evaluation methodology involving the use of correlation coefficients to compare scores to a benchmark series. Catalfamo (2018) incorporated both of these innovations into her examination of French nowcasts of the US economy during the Great Recession.

This paper applies a similar textual analysis procedure to the quarterly Inflation Reports released by the Bank of England over the period 2005 to 2014.[3] Each Report in-sample is scored on a scale between −1.5 and +1. The "Bank speak" words (see Appendix) were selected in-sample, and then scores for the Reports for the periods 1998-2004 and 2015-2018 were produced as (pseudo) out-of-sample exercises. Identifying words based on the middle part of the available text allows us to consider the role of shifting vocabulary and structure by comparing the fit both before and after.

Examining Inflation Reports allows us to study a case where text forecasts and quantitative forecasts are included in the same document. Each Report contains a comprehensive analysis of demand, output, price levels, financial market conditions, and monetary policy in the United Kingdom. Our examination focuses on the *Overview* section, in which the Bank provides qualitative assessments of the current and future outlook for real output growth. Using a textual analysis scoring methodology, we convert these assessments to quantitative values and construct a time series index of the Bank's assessment of the current and near term forecast of UK output growth implied by the text in the Report. We compare this time series to both the quantitative nowcasts and one-quarter-ahead forecasts since the text blends together current conditions and near term forecasts, and we include words describing both in our analysis.

To evaluate the accuracy of the qualitative assessments, we compare the time series constructed by scoring the text of the Inflation Report to real-time output growth data in the United Kingdom. The results of this analysis indicate that the score series is very highly correlated with output growth in the publication quarter in-sample, and remains relatively highly

---

[3] Clements and Reade (2016) use a computational procedure to analyze sentiment in the text of the Inflation Reports, but we are focused here on the elicited forecast from the text (see Ericsson, 2016). These elicited forecasts are different from, but likely related to, sentiment.

correlated for both of our out-of-sample periods. On the whole, this suggests that the text of the Inflation Report accurately reflected the current state of the United Kingdom economy. With regard to turning points, we find that the onset and severity of the Great Recession was not predicted ahead of time, but that the Bank did recognize underlying weakness in the economy beforehand. This is evidenced by a large drop in the score in the first quarter of 2008.

Also of interest is how well the qualitative assessments made by the Bank of England reflect the quantitative projections, and how the information content of the two sets of predictions differ. [4] In the Inflation Report, the Bank provides a much more thorough evaluation of economic conditions than its numerical nowcasts and forecasts can provide. Furthermore, it may be that the Bank chooses to convey information in the text of the Reports that it does not incorporate in its numerical predictions, as these predictions can have significant effects on financial market activity. In particular, we test if the forecast errors of the quantitative output growth projections published in the Inflation Report are associated with the qualitative assessments in the same Report. We find that the information conveyed in the text differs from the information in the quantitative forecasts and could potentially be used to reduce the absolute magnitude of quantitative forecast errors. Although there might be some timing differences in when the quantitative forecasts were produced and when the text was written, in general they should have been based on approximately the same information sets. This suggests that there may be different loss functions for the quantitative forecasts as compared to the text forecasts. A visual comparison of the series showed that the quantitative projections better reflected the actual

---

[4] We do not report evaluations of the quantitative forecasts here as they have been evaluated thoroughly in other work. The Bank of England's Independent Evaluation Office (IEO, 2015) and Groen et al. (2009) evaluated the quantitative forecasts in the Inflation Report. For evaluations of the inflation fan charts, see Clements (2004) and Wallis (2004).

path of output growth overall, but that the qualitative assessments performed particularly well during and around the Great Recession.

Our in-sample analysis is produced by scoring ten years of Reports based on the words in the Overview section of each Inflation Report (2005-2014). We then use an automatic scoring procedure to produce out-of-sample scores for the Reports from 1998-2004 and 2015-2018. Although the correlation is lower out-of-sample than in-sample, the text still appears to generally reflect economic conditions with interesting deviations around the US downturn in 2001 as well as recently after the Brexit referendum. It is somewhat surprising how well the later sample fits because the format of the Inflation Report changed in 2015.[5] We also show that the text appears to have information that could be used to reduce the Bank of England's quantitative nowcast and forecast errors.

The next section in this paper describes the methodology used to analyze and score the text of the Bank of England's Inflation Reports. The third section presents the qualitative projections derived from the textual analysis, both in-sample and out-of-sample, and the quantitative projections obtained from the Bank of England, as well as the benchmark real-time output growth series. The fourth section discusses the evaluation procedure and the models that are used to compare the three data sets both in-sample and out-of-sample. The fifth section details the results of the analysis and the last section provides conclusions.

2. Scoring Methodology

The approach of this paper builds on existing literature on qualitative forecasting, starting with Goldfarb et al. (2005). Their goal was to construct an index based on qualitative forecasts in

---

[5] In August 2015 the Overview section at the beginning of the Inflation Report was replaced with a less comprehensive Monetary Policy Summary.

news articles during the Great Depression. Each article was assigned a score for its nowcast – that is, the assessment of current economic conditions – and a score for its forecast – that is, the outlook for future conditions. These scores were then averaged over the articles for each period, which allowed for the construction of a time series that could be evaluated using standard techniques.

Early applications of scoring procedures created indices that were centered on zero, and which took values between –1 and +1 inclusive. In more recent forecasting literature, however, it has been suggested that a wider range of negative values may be needed to effectively quantify pessimistic outlooks during times of severe contraction (Ericsson, 2016). To address this issue, Mathy and Stekler (2018) extended the lower bound of their index to –2 in their analysis of qualitative statements made during the Great Depression. Likewise, Catalfamo (2018) extended the lower bound of her index to –1.5 in her analysis of qualitative statements made during the Great Recession. The difference in the minima of these two indices arises from the difference in the severities of the downturns that they were designed to represent.

We use a similar procedure to convert the qualitative statements made by the Bank of England into quantitative values. Since this paper examines Bank of England Inflation Reports between 1998 and 2018, a period that includes the Great Recession, our index was defined using the same extrema as Catalfamo (2018).

We construct our scoring index as a discrete scale, ranging between –1.5 and +1 inclusive, with regular increments of 0.25. Higher index values correspond to a more positive outlook for output growth, while lower index values correspond to a more negative outlook. For instance, a score of +1 is defined as indicating "exceptional growth", while a score of –1.5 is

defined as indicating a "trough." [6] In order to apply this index to the selected text in a consistent manner, we follow Stekler and Symington (2016) and identify key words and phrases that corresponded to each index value.

2.1 In-Sample Scoring

For the ten year period 2005-2014 we identified recurring expressions in the Overview section of the Inflation Report that were used to describe the outlook for output growth in order to properly characterize the specific language used by the Bank of England to communicate its assessment of economic conditions. Table 1 presents the scoring index and an abbreviated example set of the key words and phrases. The Appendix provides a complete table of the recurring expressions and their corresponding index values.

To illustrate the recurring expression identification procedure, consider the first Inflation Report in our dataset, published in February of 2005. In describing the current state of the UK economy, the Bank of England noted that "the pace of expansion has slackened", but that "output growth was around trend" and "manufacturing output returned to modest growth". In these statements, the key words and phrases identified for the index were "slackened" (+0.25), "around trend" (+0.5), and "modest growth" (+0.5). Providing an outlook moving into the future, the Bank predicted that "four-quarter GDP growth remains robust and close to trend" with "solid growth in investment and public consumption", but that "household spending grows at a slightly weaker pace". In these statements, the key words and phrases identified for the index were "robust" (+0.5), "close to trend" (+0.5), "solid growth" (+0.5), and "slightly weaker" (+0.25).

---

[6] Note that no Reports were scored as 0, which would convey a neutral Report.

Each Report was analyzed individually to convert the qualitative assessments into numerical values using our scoring procedure.

2.2 Out-of-Sample Scoring

To produce the scores for the Reports from 1998-2004 and 2015-2018, we used a fully automatic approach to collect the counts of the number of distinct expressions of the pre-selected terms for each score category in each Inflation Report. We allow for the impact of the number of terms for different score categories to be different by producing regression weights based on the in-sample scores.[7] For the period 2005-2014, we regress the overall manual score (S_t) that was assigned to each Report on counts of each category weighted by their expression scores as follows:

$$S_t = \beta_0 + \beta_1 * 1 * (count_{1t}) + \beta_2 * 0.75 * (count_{0.75t})$$
$$+ \cdots + \beta_{10} * (-1.5) * (count_{-1.5t}) + e_t$$

We then produced the out-of-sample scores by taking the automatically collected counts of the different expressions and the estimated regression weights based on the in-sample analysis.

3. Data

To construct our time series of scores of the text, we apply the scoring methodology to the text of the Overview section (or Monetary Policy Summary section from August of 2015) of the Inflation Reports released by the Bank of England from 1998-2018.[8] By aggregating these

---

[7] For example, a small number of negative words might bring the score of an entire report down substantially. By using counts, we focus on the presence and frequency of these words. An alternative, often used in sentiment analysis to control for variations in length of text, is to look at share of words, but we used raw counts here.

[8] The Inflation Reports are available from: https://www.bankofengland.co.uk/inflation-report/inflation-reports . We start in 1998 because this is the first full year of available quantitative output growth forecasts in the inflation report.

scores, we construct a time series to represent the evaluation by the Bank of the near term UK economic conditions as conveyed in the text. The in-sample series is displayed as the solid line in Figure 1 and the out-of-sample series is the dashed line. Notably, the series never reaches a value of +1 at any point over this period, as no Report published between 1998 and 2018 was judged to contain an assessment of "exceptional growth." Neither does the series ever take a value of precisely 0, as no Report published over the period of interest was associated with a completely neutral outlook.

To assess how well the text conveys the current economic conditions, we compare our text scores to real GDP growth data. In our analysis we are careful to use real-time data. Recent literature has pointed out that forecasters should not be expected to incorporate unavailable information into their predictions, and therefore that "evaluations of forecasts should usually focus on early releases of the data" (Croushore, 2011). Taking this argument into consideration, we construct the benchmark series for our evaluation using the first release of data on the level of output in the United Kingdom. Obtained from the Bank of England through September of 2016 and then from the ONS, this data set contains observations at a quarterly frequency, and includes the contemporaneously available estimates for each observation published on a monthly basis.[9] Although the original data points were expressed as levels, in terms of British pounds, they were converted to year-over-year growth rates for the purposes of this analysis. In effect, the final benchmark series consists of the first available estimates of year-over-year output growth published for each quarter in the data set.

---

[9] Real-time data available from: https://www.bankofengland.co.uk/statistics/gdp-real-time-database, through September of 2016, and then for October of 2016 through January of 2019 from https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/realtimedatabaseforukgdpabmi.

We also compare our text score series with the quantitative nowcasts and one-quarter-ahead forecasts that are also reported in the Inflation Report. In terms of quantitative projections by the Bank of England, there are two sets of predictions. One is conditioned on market interest rate expectations and the other on a constant interest rate equivalent to the current Bank Rate.[10] According to Groen et al. (2009), the Bank now emphasizes the forecasts conditioned on the market path. The analysis here focuses on the median forecasts conditioned on market-implied interest rates for the publication quarter (nowcasts) and the one-quarter-ahead forecasts.

The evaluation procedure presented in the next section describes how we investigate the relationship between these time series.

4. Evaluation Procedure

The evaluation of the qualitative assessments made by the Bank of England takes place in two parts. In the first, we focus on correlations and graphical comparisons. In the second, we use efficiency regressions to test if the text provides information that could improve the Bank of England's quantitative nowcasts and/or one-quarter-ahead forecasts. Throughout our evaluation, we consider four different sample periods: the full sample (1998-2018), the 2005-2014 in-sample period, the 1998-2004 pre-sample period, and the 2015-2018 post-sample period.

4.1 Correlation and Graphical Analysis

Recent literature has proposed the use of correlation coefficients as a means of comparing projections derived through textual analysis procedures to the corresponding actual values (Catalfamo, 2018; Mathy & Stekler, 2018). We follow this literature and use the correlation, r:

---

[10] Faust and Wright (2008) explore the impact of conditioning on a constant interest rate and argue that conditional forecasts are more difficult to evaluate.

$$r = \frac{\Sigma\left(\widehat{Y_{t,k}} - \bar{\bar{Y}}\right)(X_{t+k} - \bar{X})}{\sqrt{\Sigma\left(\widehat{Y_{t,k}} - \bar{\bar{Y}}\right)^2}\sqrt{\Sigma(X_{t+k} - \bar{X})^2}}$$

where $\widehat{Y_{t,k}}$ is the score associated with the qualitative assessment made at time t for k periods

ahead, $\bar{\bar{Y}}$ is the mean of the set of scores, X_{t+k} is the actual output growth value at time $t + k$,

and $\bar{X}$ is the mean of the set of actual values. We also estimate correlations between the text

scores and the near term quantitative projections also included in the Inflation Report.

A high correlation between the text scores and real GDP growth indicates that there is a

strong positive association. In practice, this means that directional changes in the benchmark

output growth series are mirrored by directional changes in the series that is being evaluated.

Intuitively, such a result implies that the qualitative assessments in question accurately represent

the true state of the economy. On the other hand, a low coefficient indicates that movements in

the benchmark series do not align with movements in the series being evaluated. Naturally, this

leads to the opposite conclusion that the corresponding assessments poorly reflect the true state

of the economy.

In our graphical analysis we focus on three key events: the US downturn in 2001, the

Great Recession, and the economy after the referendum on Brexit. Most of our discussion will

focus specifically on the performance of the scores around the start and end dates of the Great

Recession in the UK. This allows for a determination of whether or not the Bank of England

foresaw the onset of the contraction ahead of time, and whether or not it recognized the

downturn quickly once it had begun. Although for the overall evaluation of the qualitative

nowcasts and forecasts we used real-time output growth data, here the analysis occurs using the

official business cycle peak and trough dates as determined ex post. This is standard practice in

turning point evaluation, as the important question is whether or not a forecaster predicted the onset of a recession, regardless of data revisions and other confounding factors.

4.2 Comparing Qualitative Predictions and Quantitative Predictions

One interesting feature of the Bank of England Inflation Reports is that they include both qualitative assessments and quantitative projections. This leads to two closely related questions – that is, whether or not the information contents of the two sets of predictions differ, and whether or not one set of predictions can be improved by incorporating information from the other set. To investigate these issues we test the quantitative forecasts for efficiency, i.e. whether or not the projections take into account all of the information available at the time that they are published. In particular, the evaluation presented in the next section focuses on whether or not the quantitative output growth projections produced by the Bank of England are efficient with respect to the corresponding qualitative output growth assessments.

Clearly, the text of each Inflation Report contains a set of information regarding the current and future paths of output growth in the form of qualitative assessments. It is assumed that this information set is effectively captured through the textual analysis scoring procedure applied in this paper. Since the quantitative projections made by the Bank of England are published alongside the Inflation Report, they should theoretically contain all of the information represented by the qualitative assessments, as this information was available ex ante. If the quantitative projections are to be called efficient, there should therefore be no value in incorporating any additional information conveyed in the qualitative assessments. This hypothesis can be tested by estimating the general model proposed by Faust and Wright (2013), which is specified here as such:

$$X_{t+k} - \widehat{X_{t,k}} = \alpha + \beta \widehat{Y_{t,k}} + \varepsilon_{t+k},$$

where the variables are defined in the same way as above with $(X_{t+k} - \widehat{X_{t,k}})$ representing the

forecast (or nowcast) error of the quantitative forecasts (or quantitative nowcasts) and $Y_{t,k}$ being

the score of the qualitative analysis from the text. This is similar to the efficiency regression

specified by Hollrah et al. (2017) in their analysis of tonality in the text of the Federal Reserve

Greenbook. By regressing the quantitative forecast errors on the scores associated with the

qualitative assessments, it is possible to determine whether or not the errors could be reduced by

taking into account said assessments. If the null hypothesis that the coefficient β is equivalent to

zero is rejected, there is some value in incorporating information from the qualitative

assessments, and the quantitative projections are therefore inefficient.

5. Results

5.1 Correlations

The correlations between the text score and the benchmark output growth series (as well

the correlations between the text score and the quantitative nowcasts and forecasts) are shown in

Table 2. The full sample correlation between the text scores and output growth in the publication

quarter t is 0.84, which indicates that the former series is highly positively correlated with the

latter. The text scores exhibit nearly as strong a relationship with output growth one-quarter-

ahead $(t + 1)$, as the two series have a correlation of 0.81. These correlations are notably

stronger than those in similar text studies in other contexts such as Mathy and Stekler (2018) and

Catalfamo (2018). Thus, on the whole, the outlooks published by the Bank accurately reflected

the near term conditions of the United Kingdom economy over the period 1998 to 2018.

As we would expect, the in-sample correlation between the text scores and output growth in the publication quarter t is even higher at 0.94. The out-of-sample correlations are lower, although they are still relatively strong.[11] One interesting pattern is that, overall, the in-sample and the full sample text scores are slightly more correlated with current quarter real GDP growth as compared to next quarter. But the pattern reverses out-of-sample: both pre-sample and post-sample text scores are more highly correlated with next quarter's real GDP growth than current quarter growth. In any case the words selected in-sample appear to have relevance both pre-sample and post-sample as well.

In terms of the correlations between the text score and the quantitative nowcasts and forecasts, it is notable that overall the text score is highly related to both. In general the text score is slightly more tightly related to the one-quarter-ahead quantitative forecasts in the same Report rather than the current quarter nowcasts (in-sample it is slightly the other direction). It is also interesting that the post-sample is more tightly related to the forecasts than the pre-sample, whereas the opposite was true when comparing to realized data. Overall, however, it appears that the text is consistent with the quantitative forecasts.

5.2 Graphical Analysis

Figure 2 presents the text score over time along with real GDP growth in the publication quarter t. Although they are on different scales, it is clear that they follow a similar pattern overall, as was reflected in the high correlations discussed above. Three key time periods are interesting to explore in the graphical analysis: the 2001 period around the recession and 9/11

---

[11] The correlation is lowest for the post-sample period, which may be due to the change in the Inflation Report format.

terrorist attacks in the US, the Great Recession of 2008-2009, and the period after the Brexit referendum in June of 2016.[12]

Both the 2001 and the post-Brexit referendum periods appear in Figure 2 as deviations between the text score series and the real GDP growth series, but part of this has to do with scaling. For reference, it is interesting to note that a below zero output growth rate is consistent with a score of -3/4, which can be seen accurately reflected in the second quarter of 2012 when there was a period of weakness in the UK. In 2001 as the US went into recession there was concern reflected in the text that the UK faced downside risks, but the quantitative nowcasts and forecasts shown in Figures 3 and 4 show that the Bank of England had an accurate expectation of the likely outcome for the UK. For the period since the Brexit referendum (2016Q2 to the end of the sample) we can again see that the quantitative nowcasts and forecasts have been consistent with output growth outcomes. Here, the text scores have not yet signaled a recession, but rather a range between slower growth deterioration, which is consistent with the realizations to date. Capturing both of these events in the text out-of-sample also provides verification that the words that were selected from the 2005-2014 period had relevance for these other periods as well.

To evaluate the performance of the Bank of England in regard to turning points, we focus on the Great Recession period from 2008Q2-2009Q3. There was already a notable decline in the text score in the first quarter of 2008. In the Inflation Report for this quarter, the Bank noted a "slowdown concentrated in the financial and retail sectors." They expected "output growth to slow markedly," and they also noted that "the risks to growth are weighted to the downside." Correspondingly, the series decreased sharply, from +0.75 to +0.25. This indicates that the Bank successfully recognized underlying weakness in the economy prior to the onset of the Great

---

[12] We thank the anonymous referee who suggested that we extend our analysis back to the earlier sample to investigate the 2001 period.

Recession. The current outlook worsened even further in the third quarter of 2008, as the Bank recognized that "growth eased" and "residential and business investment prospects deteriorated". As such, the series decreased once again, from +0.25 to –0.25.

Despite these early drops in the text scores series, it does not appear that the Bank of England succeeded in recognizing the onset of the Great Recession as it occurred. Only scores of –0.75 or lower are defined to be consistent with recession conditions, so the score first signals a recession in the final quarter of 2008, midway through the contraction. Only then did the text of the Inflation Report acknowledge that "the economy probably entered a recession in the second half of 2008," and that "output is likely to contract further." In terms of identifying the end of the recession, however, the text of the Inflation Report did well, with the text score series reaching its trough one-quarter-ahead of the trough of the real-time output growth series.

Comparing the text scores to the associated quantitative nowcasts and forecasts in Figures 3 and 4 respectively, neither the text nor the quantitative forecasts identified recession conditions until the final quarter of 2008, and the text identified the trough more quickly. Both the qualitative and the quantitative evaluations of the economic conditions signaled some concern as early as the first quarter of 2008, but the drop in the text score may have been particularly informative of deteriorating conditions.

5.3 Efficiency Regressions

In order to compare the information contents of the quantitative projections and the qualitative assessments discussed above, we regressed the quantitative nowcast errors on the corresponding text scores (Table 3), and the quantitative one-quarter-ahead forecast errors on the text scores from the same Report as the forecasts (Table 4) as an extended version of a Mincer

and Zarnowitz (1969) regression. If the quantitative forecasts (or nowcasts) contain all of the relevant information included in the text, then the quantitative forecast errors should not be predictable by the text scores.

By testing the null hypothesis that the coefficient on the text scores is equal to zero, we can determine whether or not the quantitative predictions could have been improved by incorporating information from the text. In fact we find robust results for all except the pre-sample period that the text scores contain relevant information that could improve both the nowcasts and the forecasts.[13] The strongest results are notably for the recent post-sample period where it appears that the forecast errors for both the current quarter and the one-quarter-ahead quantitative forecasts could have been reduced by including the score of the accompanying text. With forecast errors defined as realized minus predicted, this means that the quantitative forecasts over this period have been overly optimistic (since the constant is also negative for this sample) and that the more negative tone of the text is a more accurate assessment.[14]

Our results are notably different from those of Clements and Reade (2016) who also analyzed the text of the Bank of England Inflation Reports. They used a sentiment index and found little evidence for a relationship between the overall sentiment of current text and forecasts. They did find, however, that more narrow sentiment measures around specific economic conditions could improve point forecasts. Combining our results with theirs suggests that using specific central bank vocabulary helps capture the relevant information in the Inflation Report text.

---

[13] The insignificance for the pre-sample period is likely due to the focus of the text on the potential risks coming from the US downturn in 2001.

[14] See Batchelor (2007) for a discussion of bias patterns in macroeconomic forecasts.

One interpretation of these results is that it is valuable to read the qualitative assessments along with the quantitative forecasts provided by Bank to get a fuller picture of current and future economic conditions. It could also be that the loss function under which the qualitative forecasts are produced differs from the loss function under which the quantitative forecasts are produced.

6. Conclusion

The analysis presented in this paper is valuable from both a methodological standpoint and an empirical standpoint. By adapting and expanding textual analysis techniques established in the recent forecasting literature, we convert text assessments made by the Bank of England into quantitative values both in-sample and out-of-sample periods. We find that the qualitative assessments made by the Bank of England between 1998 and 2018 perform well against an output growth series constructed using real-time data. The in-sample text scores are very highly correlated with output growth in the publication quarter, as well as one-quarter-ahead, and the out-of-sample text scores also remain strongly correlated with realized output growth. This implies that the text assessments published by the Bank of England accurately reflected the near term economic conditions of the United Kingdom economy.

In contrast, the Bank did not succeed at the forecasting of turning points, as it failed to predict the onset of the Great Recession ahead of time in either the text or its quantitative forecasts. Neither the text nor the quantitative forecasts identified recession conditions until the final quarter of 2008, almost midway through the cyclical contraction. Nonetheless, the Bank did recognize underlying weakness in the economy prior to the downturn, as evidenced by drops in the text scores in early 2008. Furthermore, efficiency regressions show that the qualitative assessments contained valuable information not incorporated into the numerical forecasts. Thus

reading the text assessments along with the quantitative forecasts in the Bank of England's Inflation Report provides a more accurate picture of current and future output growth in the UK.

The scoring index and the set of recurring expressions identified in the Bank of England's Inflation Report may also provide value for other researchers interested in examining the Reports. A high level of consistency was observed in the word choice and terminology utilized by the Bank, and thus the methodology could be applied to Reports existing outside of the data set evaluated in this paper. Furthermore, the methodology could potentially be applied to quantify assessments of other macroeconomic variables discussed in the Reports.

References

Bank of England Independent Evaluation Office (2015). Evaluating Forecast Performance. November 2015. Available at https://www.bankofengland.co.uk/independent-evaluation-office/forecasting-evaluation-november-2015

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, *23*(2), 189-203.

Catalfamo, E. (2018). French Nowcasts of the US Economy during the Great Recession: A Textual Analysis. Research Program on Forecasting Working Paper No. 2018-001.

Clements, M. P., & Reade, J. J. (2016). Forecasting and Forecast Narratives: The Bank of England Inflation Reports. ICM-2016-10.

Clements, M (2004), 'Evaluating the Bank of England density forecasts of inflation', The Economic Journal, Vol. 114 (498), pages 844–66.

Croushore, D. (2011). Frontiers of Real-Time Data Analysis. *Journal of Economic Literature*, *49*(1), 72-100.

Ericsson, N. R. (2016). Eliciting GDP forecasts from the FOMC's minutes around the financial crisis. *International Journal of Forecasting*, *32*(2), 571-583.

Faust, J., & Wright, J. H. (2008). Efficient forecast tests for conditional policy forecasts. *Journal of Econometrics*, *146*(2), 293-303.

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In *Handbook of Economic Forecasting*, *vol. 2*. North Holland.

Goldfarb, R., Stekler, H. O., & David, J. (2005). Methodological issues in forecasting: Insights from the egregious business forecast errors of late 1930. *Journal of Economic Methodology*, *12*(4), 517-542.

Groen, J. J., Kapetanios, G., & Price, S. (2009). A real time evaluation of Bank of England forecasts of inflation and growth. *International Journal of Forecasting*, *25*(1), 74-80.

Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School*, *58*(2), 120-127.

Hollrah, C. A., Sharpe, S. A., & Sinha, N. R. (2017). What's the Story? A New Perspective on the Value of Economic Forecasts. Finance and Economics Discussion Series 2017-107, Board of Governors of the Federal Reserve System.

Lundquist, K., & Stekler, H. O. (2012). Interpreting the forecasting performance of business economists during the Great Recession. *Business Economics*, *47*(2), 148-154.

Mathy, Gabriel. & Stekler, H. O. (2018) Was the deflation of the depression anticipated? An inference using real-time data. *Journal of Economic Methodology*, 25(2), 117-125.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER.

Stekler, H. O., & Symington, H. (2016). Evaluating qualitative forecasts: The FOMC minutes, 2006-2010. *International Journal of Forecasting*, *32*(2), 559-570.

Wallis, K F (2004), 'An assessment of Bank of England and National Institute inflation forecast uncertainties', National Institute Economic Review, Vol. 189, pages 64–71

Table 1: Scoring the Text of the Bank of England Inflation Report

Score	Assessment	Examples of Recurring Words and Phrases (Full list in the appendix)
+1	Exceptional Growth	Around its Peak, Markedly Higher, Substantially Higher
+¾	Strong Growth	Brisk, Buoyant, Rapid, Strong, Upturn
+½	Normal Growth	Close to Trend, Firm, Moderate, Modest, Robust, Solid, Steady
+¼	Slow Growth	Below Trend, Decelerate, Slow, Soft, Slackened, Subdued
-¼	Deterioration	Deteriorated, Fall, Muted, Slow Markedly, Sluggish, Weak
-½	Downturn	Fall Rapidly, Marked Deterioration, Weakened Sharply
-¾	Contraction	Contracted, Depressed Markedly
-1	Recession	Contracted Sharply, Recession, Severe Instability
-1¼	Deep Recession	Deep Recession
-1½	Trough	Sharp and Synchronized Downturn, Trough in Output

Figure 1: Scores of Bank of England's Qualitative Assessment of GDP Growth over Time
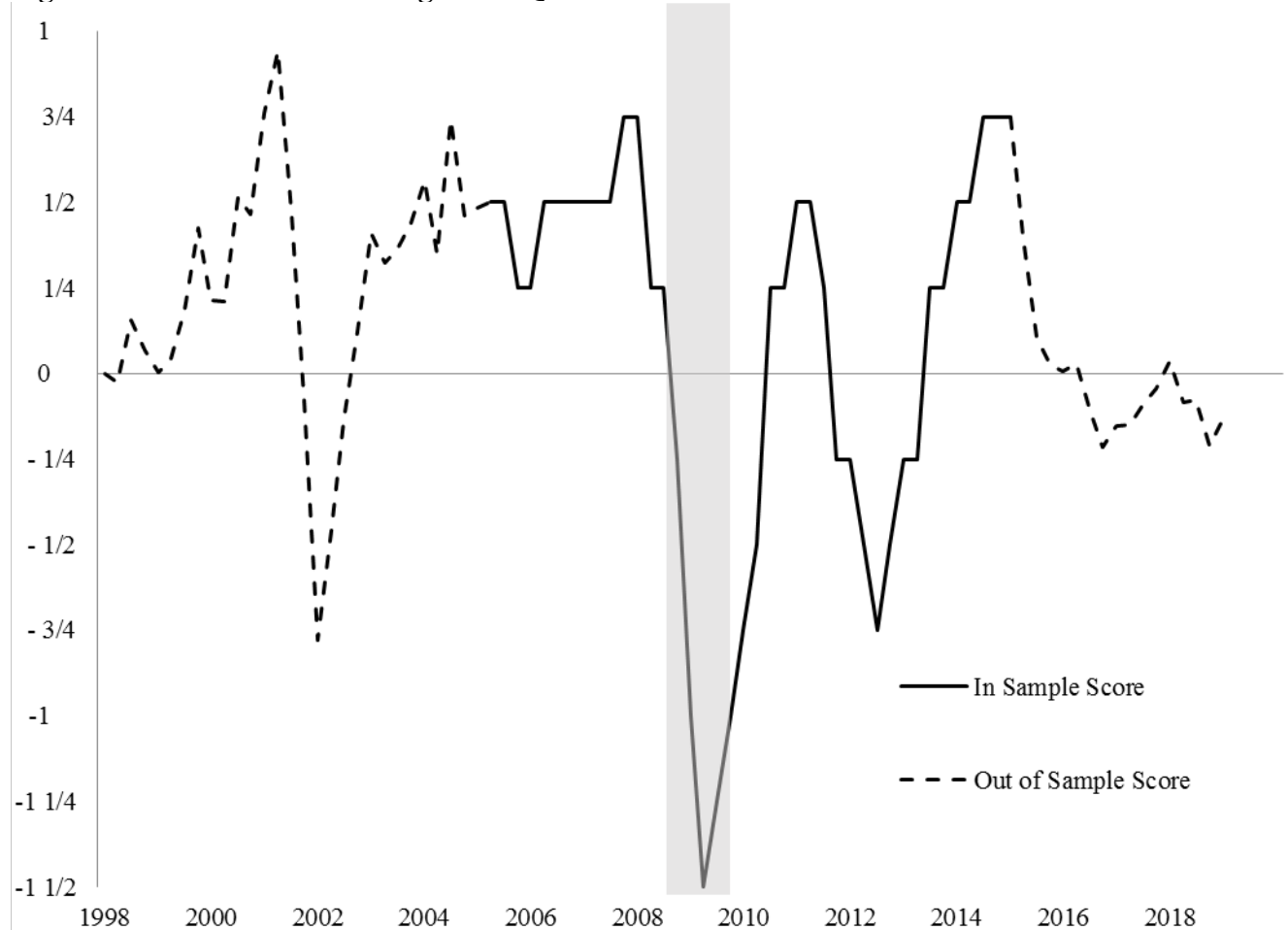
Note: Gray shading in each figure indicates the Great Recession from 2008Q2-2009Q3 according to the dating of the UK parliament.[15]

---

[15] https://www.parliament.uk/documents/commons/lib/research/key_issues/Key-Issues-Recession-and-recovery.pdf

Table 2: Correlations between Text Score and Quantitative Forecast or Realized Data

	Correlation of Text Score with			
	Real GDP Growth current quarter (t)	Real GDP Growth $t+1$	BoE Nowcast current quarter (t)	BoE Quantitative Forecast $t+1$
Full sample (1998-2018)	0.84	0.81	0.83	0.84
In-sample (2005-2014)	0.94	0.89	0.93	0.92
Pre-sample (1998-2004)	0.53	0.59	0.48	0.51
Post-sample (2015-2018)	0.38	0.48	0.65	0.70

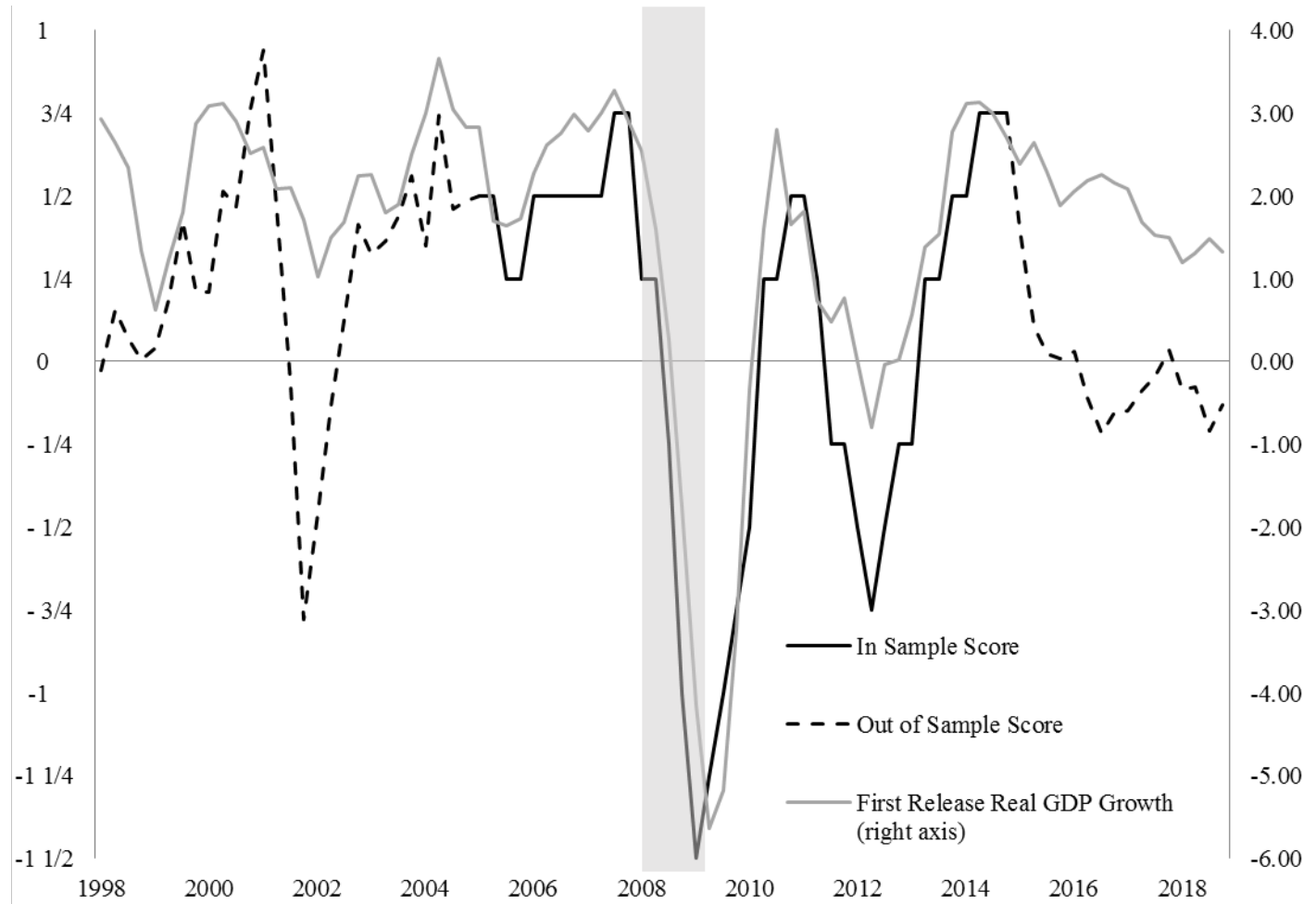**Figure 2: Comparing Text Scores with Real Time Output Growth**

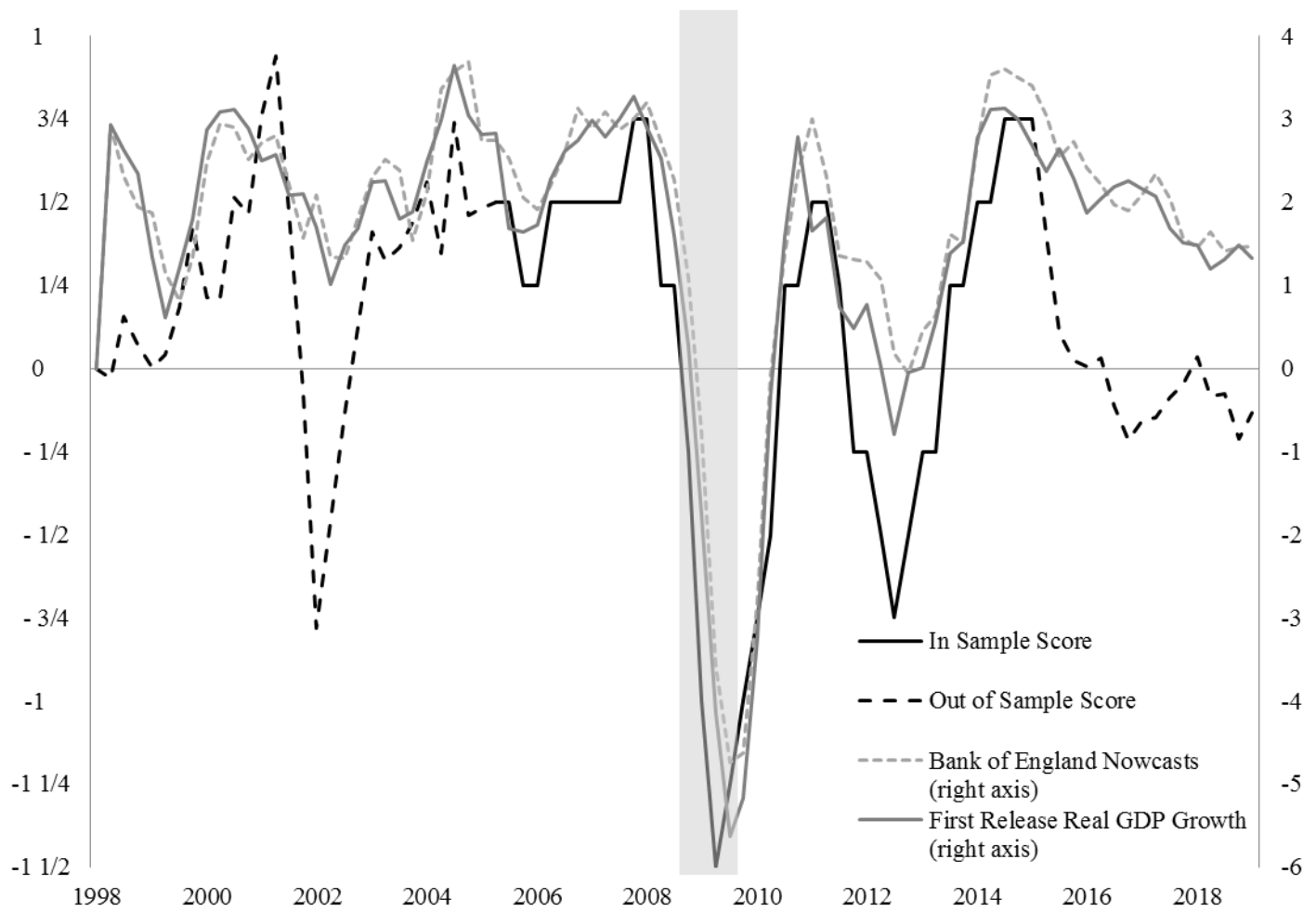**Figure 3: Comparing Text Scores with Bank of England Nowcasts**
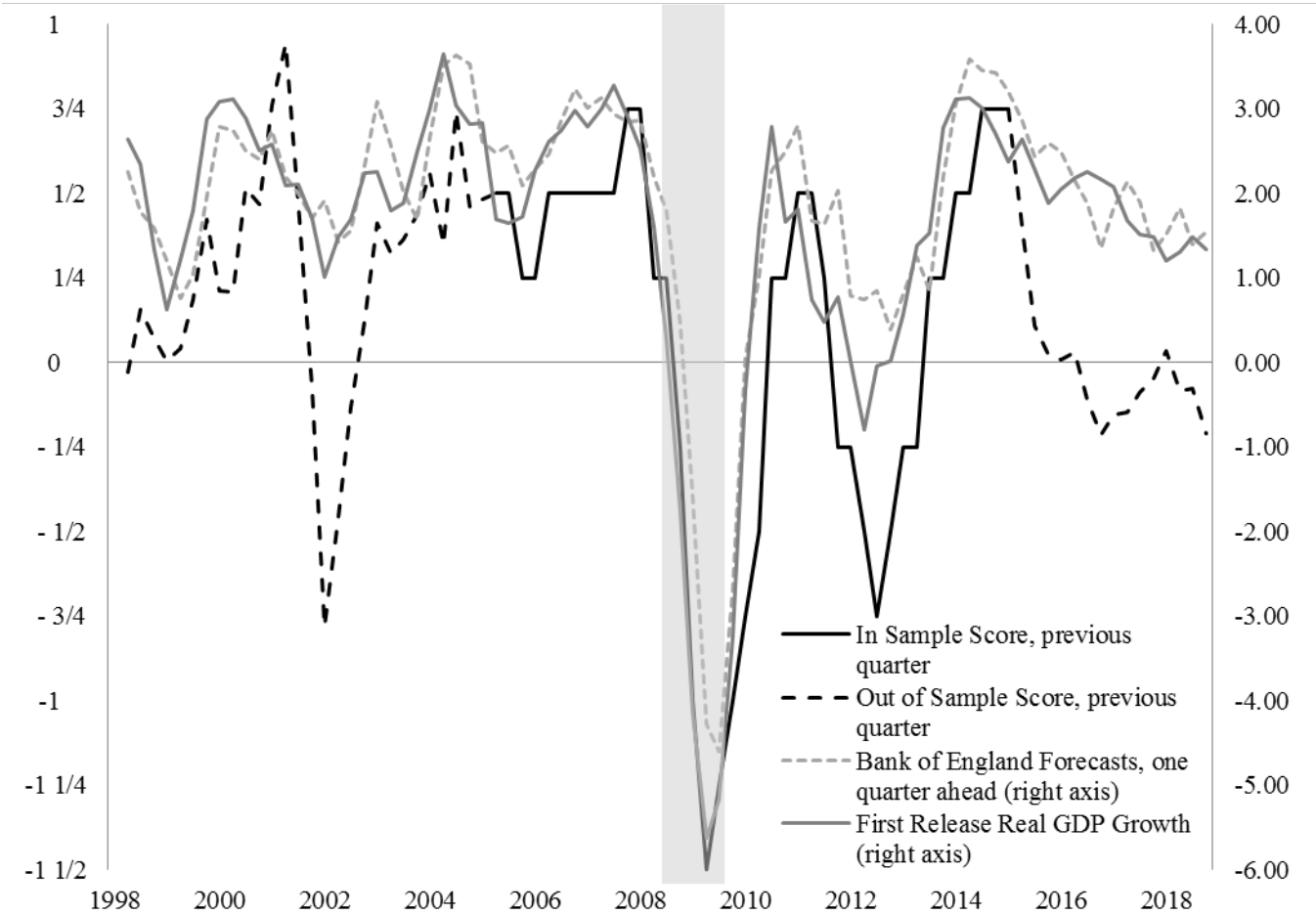
Figure 4: Comparing Text Scores with Bank of England One-quarter-ahead Forecasts

Note the dates on the x-axis are the target date, so scores and forecasts are from the Report the quarter before the date on the axis.

Table 3: Results of Efficiency Regressions for Quantitative Nowcasts

Dependent Variable: Current Quarter (*t*) Forecast Error	Full sample (1998-2018)	In-sample (2005-2014)	Pre-sample (1998-2004)	Post-sample (2015-2018)
Coefficient on current quarter text score	0.24**	0.25***	0.11	-1.06***
	(0.10)	(0.08)	(0.14)	(0.30)
Constant	−0.24***	-0.39***	-0.04	-0.19***
	(0.05)	(0.07)	(0.06)	(0.05)
R^2	0.08	0.14	0.01	0.30
f-statistic	10.64	23.85	0.34	23.84
p-value	0.0001	<0.0001	0.71	<0.0001

Newey-West standard errors in parentheses.
$* \rightarrow p < 0.10,$ $** \rightarrow p < 0.05,$ $*** \rightarrow p < 0.01$

Table 4: Results of Efficiency Regressions for Quantitative Forecasts (one-quarter-ahead)

Dependent Variable: One-quarter-ahead (*t + 1*) Forecast Error	Full sample (1998-2018)	In-sample (2005-2014)	Pre-sample (1998-2004)	Post-sample (2015-2018)
Coefficient on publication quarter text score	0.38*	0.44**	0.13	-0.89**
	(0.20)	(0.20)	(0.28)	(0.31)
Constant	-0.31***	-0.53***	-0.01	-0.14
	(0.11)	(0.17)	(0.13)	(0.10)
R^2	0.08	0.13	0.01	0.26
f-statistic	3.80	5.20	0.12	23.16
p-value	0.03	0.01	0.89	<0.0001

Newey-West standard errors in parentheses.
$* \rightarrow p < 0.10,$ $** \rightarrow p < 0.05,$ $*** \rightarrow p < 0.01$

Appendix: Comprehensive list of recurring expressions and their corresponding scores.
(We considered all variants of these terms and used a recursive search in Python to accurately count all terms and phrases.)

Expressions Scored as +1

Around its Peak
Markedly Higher, Substantially Higher
Sharp Increase

Expressions Scored as +¾

Above its Long-Run Average

Brisk, Brisk Pace, Expand Briskly, Grow Briskly, Rose Briskly

Buoyant

Healthy Growth

Quickened

Rapid, Rapid Pace

Relatively Upbeat

Strong, Strengthen

Upturn

Expressions Scored as +½

Around its Historical Average/Long-Term Average, Around Trend

Boost, Small Boost

Close to its Average/Average Rate/Historical Average/Long-Term Average, Close to Trend

Firm, Firm Growth

Gathered Pace

Gentle Acceleration

Grew Steadily

Moderate, Moderate Contribution, Moderate Expansion, Moderate Growth

Modest, Modest Acceleration, Modest Growth, Modest Pickup, Modest Recovery

Momentum was Maintained

Near its Average, Near Trend

Picked Up

Regains Momentum

Resilient

Rising Steadily

Robust, Robust Growth, Robust Pace

Solid, Solid Growth, Solid Pace

Steady, Steady Expansion, Steady Growth, Steadily Increasing

Expressions Scored as +¼
Below its Long-Term Average, Below Trend
Dampen
Decelerate
Drag Slightly
Dropped Back
Eased, Eased a Touch
Fell Back, Fell Back Modestly
Improving
Modest Deceleration
Rebounded
Recovers
Revived
Signs of Improvement
Slackened
Slightly Weaker
Slow, Slowdown, Slower Growth
Soft
Stalled
Subdued

Expressions Scored as –¼
Appeared Lackluster
Decline Gradually
Deteriorated
Fall
Faltered
Muted
Set to Decline
Slow Markedly, Slowed More Sharply
Sluggish
Sustained Weakness
Weak, Weaker Growth, Weaker Pace

Expressions Scored as –½
Fall Rapidly, Fall Sharply
Marked Deterioration
Persistent Weakness
Sharp Fall
Sharp Tightening
Sizable Drag
Weakened Sharply

Expressions Scored as –¾
Contracted
Depressed Markedly
Plummeted

Expressions Scored as –1
Contracted Sharply
Pronounced Contraction
Recession
Severe Instability

Expressions Scored as –1¼
Deep Recession

Expressions Scored as –1½
Sharp and Synchronized Downturn
Trough in Output