Forecasting and forecast narratives: the Bank of England inflation reports

Article

Accepted Version

Clements, M. P. ORCID: https://orcid.org/0000-0001-6329-1341 and Reade, J. J. ORCID: https://orcid.org/0000-0002-8610-530X (2020) Forecasting and forecast narratives: the Bank of England inflation reports. International Journal of Forecasting, 36 (4). pp. 1488-1500. ISSN 0169-2070 doi: https://doi.org/10.1016/j.ijforecast.2019.08.013 Available at https://centaur.reading.ac.uk/85698/

To link to this article DOI: http://dx.doi.org/10.1016/j.ijforecast.2019.08.013

Publisher: Elsevier

Forecasting and Forecast Narratives: The Bank of England Inflation Reports*

Michael P. Clements
ICMA Centre
Henley Business School
University of Reading
M.P.Clements@reading.ac.uk

J. James Reade
Department of Economics
School of Politics, Economics and International Relations
University of Reading
J.J.Reade@reading.ac.uk

July 25, 2019

Abstract

We analyze the narratives that accompany the numerical forecasts in the Bank of England's Quarterly *Inflation Reports*, 1997 - 2018. We focus on whether the narratives contain useful information about the future course of key macro variables over and above the point predictions, in terms of whether the narratives can be used to enhance the accuracy of the numerical forecasts. We also consider whether the narratives are able to predict future changes in the numerical forecasts. We find that a measure of sentiment derived from the narratives can predict the errors in the numerical forecasts of output growth, but not of inflation. We find no evidence that past changes in sentiment predict subsequent changes in the point forecasts of output growth or of inflation, but the adjustments to the numerical output growth forecasts have a systematic element.

Keywords. Sentiment, inflation forecasting, output growth, forecast encompassing. C55, E37, E66.

*The authors would like to thank seminar participants at the University of Birmingham, De Nederlandsche Bank, BI Norwegian Business School, and conference participants at the 16th OxMetrics User Workshop in Washington DC, and the 36th International Symposium on Forecasting for their helpful comments.

1 Introduction

Forecasts are sometimes accompanied by a narrative account. This narrative is often used to provide additional information regarding the numerical forecasts and the outlook more generally, and might include an assessment of the extent of uncertainty associated with the forecast, as well as a discussion of some of the key features and drivers of the forecast numbers. The narrative provides an accompanying 'story' to the forecast, that sets the scene, and in so doing may enhance the credibility of the forecast, by providing explanations for the future trends and developments implicit in the forecast numbers.

Although the narrative is qualitative in nature, it is nonetheless possible to convert this qualitative information into quantitative data. This then facilitates a number of areas of research. These include the following inter-related issues: whether the narrative contains information not reflected in published forecasts; whether the narratives can be used to improve the numerical forecasts; whether the narratives provide an 'early-warning' system for adverse developments which are not fully reflected in the numerical forecasts; whether various behavioural biases often found in agents' forecasts affect the numerical and narrative forecasts in the same way; whether the implicit 'loss function' for the narratives and numerical forecasts is the same; whether the narratives are informative about forecast uncertainty (whether the numerical forecasts take the form of central projections and/or 'fan charts' or probability distributions describing future outcomes).

In this paper we will use the word narrative to denote the descriptive text published along side the numerical forecasts. We use the term sentiment to denote quantitative indices which capture some aspect of the information content of the narratives, in a sense described below.

In this paper we focus on whether the narratives contain useful information that can be used to enhance the accuracy of the numerical forecasts, in a traditional expected squared-error sense. We also look at whether the narratives 'lead' the numerical forecasts. This second aspect is potentially important. For example, when the structure of the economy changes, forecasts often fail, in the sense of being worse than expected based on the model's past performance, see e.g., Clements and Hendry (2006). Hence, advance warning of a break may partially offset the deterioration in forecast performance: see, e.g., Castle, Fawcett and Hendry (2011). The narratives may hint at the possibility of impending breaks, even though it is felt that such developments are too speculative to be reflected in the numerical forecasts.

Relatedly, the recent literature has suggested that macroeconomic forecasters may be heterogeneous in terms of their loss functions, and particularly in terms of the degree of asymmetry of such loss functions: see, e.g., Capistrán and Timmermann (2009). It may also be reasonable to suppose that over-optimistic and too-pessimistic narrative statements are weighed differently by forecasters compared to the errors they make in their numerical forecasts. In short, forecasters may be more candid in their narratives than their numerical forecasts. Alternatively, the production/reporting of a point forecast - as opposed to a density forecast - may count against providing an 'early warning'.

Finally, the literature on the psychology of judgement under uncertainty (e.g., Kahneman and Tversky's 'heuristics and biases' research, as reviewed in O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley and Rakow (2006, ch. 3)) suggests that the way in which judgments are elicited may matter. An example of this is that there can be systematic differences between forecasts of central tendencies reported as point predictions as opposed to those implicit in histogram forecasts (see Engelberg, Manski and Williams (2009) and Clements (2009, 2010)). Hence there is no reason to necessarily expect the narratives and numerical forecasts to mirror each other.

In this paper we study the Bank of England's quarterly forecasts of the two key macro-variables:

inflation and output growth, and the associated narratives taken from the Bank of England Quarterly *Inflation Reports*. The narratives accompanying the text are translated into quantitative variables. We consider the *Inflation Reports* from February 1997 (when they became machine-readable) until August 2018, and in so doing contribute to a growing literature that converts institutional narratives into quantitative indices: see Gentzkow, Kelly and Taddy (2019) for a review of the use of 'text', both in terms of method and applications.

The plan of the rest of the paper is as follows. In Section 2 we review relevant literature. Section 3 outlines the construction of qualitative indices from the narratives. Section 4 explains the various facets of our forecast-based evaluation of the narratives. Section 5 describes the Inflation Report data, section 6 presents our empirical results, and section 7 offers some concluding remarks.

2 Literature Review

There is a long tradition in economics of converting some types of qualitative data into numerical indices, such as, say, the proportion of survey respondents expecting their sales to increase over the next 6 months, as opposed to staying the same, or falling. See, for example, the review of survey expectations by Pesaran and Weale (2006). There is a rapidly-expanding literature on the use of 'text', such as the nuanced narratives of the sort that appear in the Bank's *Inflation Reports*. We present a highly-selective review of some of the literature in this section: see Gentzkow *et al.* (2019) for further reading.

A key paper is the investigation by Stekler and Symington (2016) of the 'narratives' that constitute the minutes of the Federal Open Market Committee (FOMC) between 2006 and 2010. Their study quantifies the qualitative statements concerning current and future trends in the economy, from the FOMC minutes, and compares the resulting indices to forecasts from the Greenbook and the US Survey of Professional Forecasters. Ericsson (2016) considers the Stekler and Symington (2016) quantitative indices as forecasts, and evaluates their properties, and focuses on the method of calibrating the scores to the historical data to generate the indices. Romer and Romer (2008) had already made informal use of FOMC narratives to consider the link between the FOMC forecasts, the Staff ('Greenbook') forecasts, and the conduct of monetary policy by the FOMC. They looked at three episodes when the Staff and FOMC forecasts were at odds with each other. Hansen and McMahon (2016) also consider the effects of FOMC communications regarding economic conditions and future monetary policy decisions on the macroeconomy (see, also Lucca and Trebbi (2009)), as part of a wider literature that uses the analysis of text and narratives to investigate the effects of Central bank communication on inflation expectations, on the stock market, and on the real economy.

Relative to Stekler and Symington, and Ericsson, our focus is on whether quantitative indices derived from qualitative statements contain predictive power regarding the current and future course of the economy over and above that contained in the forecaster's usual numerical forecasts. In terms of subject matter, a closely-related paper is Jones, Sinclair and Stekler (2019), which also considers the Bank of England forecasts.

Castle, Hendry and Martinez (2017) consider forecasts and their accompanying narratives from the standpoint of policy-makers enacting changes in policy, and term the forecasts and narratives *forediction*. Their interest is in what forecast failure implies about the status of the forecasting model, the forediction, and the validity of the policy. Clements and Hendry (2005) showed that forecast failure need not necessarily invalidate the economic theory underlying the forecasting model. However, it is argued that forecast failure generally does entail forediction failure and the invalidity of the policy. Their policyoriented focus is different from ours, which is squarely on forecasting, but their contribution serves to illustrate the range of questions that can be addressed when narratives are available.

There have been a large number of studies evaluating the Bank of England's numerical forecasts, including Clements (2004), IEO (2015) and Fawcett, Körber, Masolo and Waldron (2015). The Bank's forecasts include measures of uncertainty surrounding the central tendencies or point predictions, continuing the Bank tradition that began with the publication of the 'fan charts'. This allows the possibility in future work of comparing these measures of uncertainty with the narratives.

A fundamental building block of the analysis of text or narrative is a method for mapping to quantitative indices - real numbers - which are amenable to statistical analysis. Given the richness of the English language, Di Fatta, Reade, Jaworska and Nanda (2015) argue that standard sentiment indices need to be adapted to the context in which they are being applied, noting that words often have different connotations and meanings in different contexts, and further focusing only on adjectives and adverbs as the descriptive words in the English language. Indeed, natural language processing attempts to automatically learn 'meaning' from statistical analyses of the distributions and co-occurrences of words: see e.g., Pennington, Socher and Manning (2014).

Methods of varying degrees of complexity, with varying degrees of intervention on the part of the investigator, and different requirements in terms of training the algorithms, have been proposed in the literature. We have chosen automatic methods over methods which require the researcher to read the text and assign a score to individual sections of text. The former are quick, reproducible, and can easily be applied to new *Reports* as time passes. They can also be readily 'tweaked' and allow experimentation. However, they less able to tease out the different gradations of meaning or 'tones' that a human would be capable of. Within the set of automatic methods, varying degrees of sophistication are possible. One may simply count the number of occurrences of a set of pre-selected words in a piece of text, as in Rosa and Verga (2008), for example. Towards the other extreme, Hansen and McMahon (2016) use computational linguistics tools: Latent Dirichlet Allocation (LDA) and dictionary methods. LDA is used for topic modelling, in the sense of discovering the topic of each sentence of a communication, and 'dictionary methods' essentially entails counting positive and negative words in that sentence. Hansen and McMahon (2016) note the widespread use of LDA in some disciplines, and reference the literature applying computational linguistic tools to the analysis of monetary policy.

We use a relatively simple dictionary approach, but with 'incrementers' and 'decrementers', as described in the following section.

3 Constructing the Sentiment Index

We consider two types of forecasts: the standard quantitative forecasts produced by the Bank of England, and sentiment "forecasts" (S) inferred from the *Inflation Report* narratives. Each S_t is intended to capture some of the the information in the narrative in the *Report* at time t, for each report in our sample. Such a measure transforms thousands of words into a single number.

A simple method of creating a sentiment index maps words into positive (+1), negative (-1) or neutral (0) scores, and evaluates the resulting sequences. We calculated an index in this way using a dictionary provided by Nielsen (2011), and took the average sentiment per word in a report. This was used for comparison purposes.

Alba (2012) develop the sentiment index that we use. This attempts to better capture the nuances and subtleties of the language that is used in the narratives. It makes use of the *Natural Language*

Table 1: Example sentences and sentiment scores.

Number	Sentence	Score
1	Excluding the boost to growth from the rebound in activity following the heavy snow in 2010 Q4, however, GDP was broadly flat.	1.5
2	Within that, and abstracting from the effects of snow, manufacturing and services output grew moderately, but there was a sharp fall in construction output.	-1.5
3	The extent of spare capacity within businesses is uncertain: the growth rate of companies' effective supply capacity appears to have slowed during the recession, but it is likely that some margin of spare capacity remains.	0
4	Employment has recovered somewhat but unemployment remains elevated.	1.5

Toolkit Bird, Loper and Klein (2009) in the *Python* programming language. First we manually rank the 3000 most common words in historical *Inflation Reports*, restricting our attention to adverbs (1000), adjectives (1000) and nouns (1000). We score words that express positive sentiment with a 1, and words that express a negative sentiment with a -1, leaving other words with a score of 0. We categorize words into three additional classes: decrementers, incrementers, and inverters. Decrementers decrease the score attached to a subsequent word (e.g. "slightly bad"), while incrementers increase the score (e.g. "really bad"), and inverters invert the sign of the score attached to a word (e.g. "not bad"). Decrementers halve the score attached to a word, while incrementers double it.

We subsequently sum the scores attached to words in a report to provide a sentiment score for that report. In order that sentiment scores are not biased by the size of a report, we divide the sum by the number of sentences in a report.

In Table 1 we provide some examples of sentences from *Inflation Reports* and the sentiment score attached to them.

Sentence 1 scores 1.5 because "boost" and "growth" both score positively, whereas flat is classed to be negative, but "broadly" decrements this to a half. Sentence 2 scores positively because of the term "grew", which is decremented by "moderately", but negatively because of the "fall", which was "sharp". Sentence 3 scores zero and shows the difficulty in evaluating sentences, because "growth" attracts a positive score, as does "effective", even though the latter is only part of the term "effective supply". The words "slowed" and "recession" score negatively and hence balance out the first two words. Finally, sentence 4, like sentence two, has "recovered" scoring positively, although "somewhat" is a decrementer, while "elevated" scores positively. Again this shows the difficulty of scoring text, since elevated when attached to the term unemployment should ideally be inverted as a score.

In addition to quantifying the sentiment of entire *Reports*, we could focus only on sentences containing particular keywords. For example, we might consider only sentences that mention "committee" in order to focus more on the opinions recorded as having been expressed by the Monetary Policy Committee; equally, we might consider only sentences with the word "demand" in, in order to consider the demand side of the economy. We do not pursue this line of inquiry here.

We note that a number of sentences in *Inflation Reports* are included for information purposes only, such as the brief of the Monetary Policy Committee at the start of the report, and the text beneath many of the forecast plots that describes the construction of the plot. Such sentences are factual and hence do not contain adjectives and adverbs, and as such any sentiment score they are attributed will be zero or very close to zero. Furthermore, these text blurbs are common to all *Inflation Reports*, and hence can be thought of only as having a nominal effect, rather than a real effect, on sentiment between reports.

4 Forecast Evaluation

4.1 Forecast Accuracy

We denote the variable of interest at time t as y_t, and a (standard numerical) forecast of the variable at time $t+h$, made at time t, as $\widehat{y}_{t+h|t}$. Each *Report* gives a forecast of the current quarter (the quarter in which the report is issued) and for each of the next 7 quarters. We adopt the convention throughout that $h=1$ refers to a current quarter forecast. We define the forecast error as $\widehat{e}_{t+h|t} = y_{t+h} - \widehat{y}_{t+h|t}$, and use squared error loss to determine forecast accuracy, with empirical counterpart given by the mean squared forecast error (MSFE). For T forecasts of length h, this is simply $T^{-1}\sum_{t=1}^{T}\widehat{e}^2_{t+h|t}$. We will sometimes use the Root Mean Squared Forecast Error (RMSFE), the square root of the MSFE. For the quantitative forecasts at least a natural way of assessing accuracy is in terms of the RMSFE relative to a set of benchmark forecasts. For inflation forecasting 'no-change' predictors often provide stiff competition (see, e.g., Atkeson and Ohanian (2001), Stock and Watson (1999)), whereas for output growth autoregressive models (which do not impose a unit root in the growth rate) are preferred. Formal tests of predictive ability can be made using approaches such as those popularized by Diebold and Mariano (1995) (see, e.g., Clark and McCracken (2011) and Giacomini (2011) for recent reviews) to see whether differences in RMSFE reflect statistically significant differences between the forecasts. We employ a regression-based Diebold and Mariano (1995) test, where we construct the term e_t^{DM} as:

$$e_t^{DM} = L\left(y_t - \widehat{y}_{t|t-h}\right) - L\left(y_t - \widehat{y}^{BM}_{t|t-h}\right),$$

where $\widehat{y}^{BM}_{t|t-h}$ is the benchmark forecast (either a 'no-change' forecast or a forecast from an autoregressive model). Conventionally, the loss function L is squared error loss, i.e., $L(e) = e^2$. Equal forecast performance implies $E(e_t^{DM}) = 0$, and this test can be implemented using the regression model:

$$e_t^{DM} = \alpha + u_t \tag{1}$$

with the null of equal accuracy, $H_0 : \alpha = 0$. The significance of α implies difference in forecast performance. If $\alpha > 0$ this implies that the benchmark forecast is 'better' than the Bank forecast, while $\alpha < 0$ implies the opposite, namely that the Bank forecast is 'better' than the benchmark.

In principle, the sentiment index (S) could be assessed for accuracy in a similar way, and it could be formally compared to the numerical forecasts ($\hat{y}$). Two possible complications arise in assessing the accuracy of S. Firstly, although the forecast origin is well defined, as the survey in which the narrative appeared, the period to which the narrative relates is often less clear. Secondly, in some cases it may only be possible to directly relate S to the macro-variable of interest when it has been appropriately "scaled", for example, to have the same mean and variance, possibly over a relatively short window of data immediately prior to the forecast origin. (Using a window of data also allows for some non-constancy in the relationship between the sentiment index and the variable of interest).

A straightforward way of scaling S to generate a sentiment forecast (denote s) is as follows. To construct a sentiment forecast for the *Report* at time t, we regress:

$$y_i = \alpha + \beta S_i + u_i$$

for $i = t - 10$ to $t - 1$, and use the estimated parameters to calculate

$$s_t = \hat{\alpha} + \hat{\beta} S_t$$

s_t can be calculated in real-time, as it makes use of actual values only up to period $t-1$ (and the first estimate of observation period $t-1$ will be known at the time of the *Report* in quarter t). (The dependence of $\hat{\alpha}$ and $\hat{\beta}$ on t is suppressed in the notation for simplicity). The single subscript on s denotes the forecast origin (i.e., the date of the *Report*) and the target date is deliberately left unspecified given the nature of the forecast. In the following we will consider whether some of the key findings depend on whether we use s or S as the measure of sentiment.

There are a number of other dimensions beyond forecast accuracy in which it might prove useful to compare the forecasts, including the extent to which the two sets of forecasts contain complementary information on future movements in the macro-variable.

4.2 Forecast Efficiency

A popular test of forecast efficiency (or optimality) is that of Mincer and Zarnowitz (1969), with recent extensions due to Patton and Timmermann (2012). The Mincer and Zarnowitz (1969) (MZ) regression tests forecast optimality at a given horizon. The regression is:

$$y_t = \alpha + \beta \widehat{y}_{t|t-h} + u_t, \tag{2}$$

where the observations range over t for a given h, and the null of optimality is that $\alpha = 0$ and $\beta = 1$. By writing:

$$y_t - \widehat{y}_{t|t-h} = \alpha + (\beta - 1)\, \widehat{y}_{t|t-h} + u_t \tag{3}$$

we obtain $Cov\left(y_{t+h} - \widehat{y}_{t+h|t}, \widehat{y}_{t+h|t}\right) = (\beta - 1)\, Var\left(\widehat{y}_{t+h|t}\right)$, so that the covariance is zero if and only if $\beta = 1$. Hence it is a test of whether the forecast efficiently uses all the information available at the forecast origin, such that the resulting forecast error is not systematically related to the forecast origin information, as filtered via the forecast. The condition $\alpha = 0$ and $\beta = 1$ is sufficient for unbiasedness (but not necessary: see Holden and Peel (1990)).

Autocorrelation consistent (AC) standard errors are used for multi-step forecasts ($h > 1$) to account for the overlapping forecasts phenomenon.

We report the results for the regression (3), and so the constant and slope are both zero under the null.

4.3 Forecast Encompassing

The MZ regression (2) can be supplemented with additional variables known at $t - h$, e.g.,

$$y_t = \alpha + \beta \widehat{y}_{t|t-h} + \kappa' z_{t-h} + u_t, \tag{4}$$

where (as shown) z_{t-h} may comprise a vector of such variables. The null is now that $\alpha = 0$, $\beta = 1$ and $\kappa = 0$. Our interest is in when the additional variable(s) include the s (or S) index, so for example:

$$y_t = \alpha + \beta \widehat{y}_{t|t-h} + \gamma s_{t-h} + u_t \tag{5}$$

which takes the form of the test for forecast encompassing between forecasts $\widehat{y}_{t|t-h}$ and s_{t-h} suggested by Fair and Shiller (1990). Tests of forecast encompassing (see Chong and Hendry (1986)) assess whether *ex post* a linear combination of forecasts results in a statistically significant reduction in the mean squared forecast error (MSFE) relative to using a particular forecast. Forecast encompassing is formally equivalent to 'conditional efficiency' due to Nelson (1972) and Granger and Newbold (1973), whereby a forecast is conditionally efficient if the variance of the forecast error from a combination of that forecast and a rival forecast is not significantly less than that of the original forecast alone. The null that $\widehat{y}_{t|t-h}$ encompasses s_{t-h} is given by $\gamma = 0$ against the (usually) one-sided alternative that $\gamma > 0$. Alternative forms of forecast encompassing tests include:

$$y_t - \widehat{y}_{t|t-h} = \alpha + \beta s_{t-h} + u_t \tag{6}$$

where the null is as before, and the interpretation is whether s_{t-h} can help explain the forecast errors associated with $\widehat{y}_{t|t-h}$. Forecast encompassing of one model, say $M2$ by another, say $M1$, is a more stringent requirement than simply that $M1$ is more accurate than $M2$ on RMSE (see, e.g., Ericsson (1992)). $M2$ might still contain useful information not in $M1$, raising the possibility that a simple linear combination of the two might produce superior forecasts in the future.

Our forecast encompassing tests are parameterized so that the dependent variable is the forecast error for the quantitative forecasts.

4.4 Forecast Updating

Patton and Timmermann (2012) show that the actual value of y_t can be replaced by a short-horizon forecast, say, $\widehat{y}_{t|t-h_1}$, to give:

$$\widehat{y}_{t|t-h_1} = \alpha + \beta \widehat{y}_{t|t-h_2} + u_t \tag{7}$$

where $h_2 > h_1$. This may be useful when the vintage of the actual being targeted is unknown. Although CPI/RPI inflation data are not subject to large revisions, national accounts variables such as real output growth are. But replacing the actual by a short-horizon forecast requires that the latter is an efficient forecast of the actual values. Otherwise the nature of the test changes, and may have no power to detect mis-specification, a situation described by Nordhaus (1987), p.673 as 'A baboon could generate a series of weakly efficient forecasts by simply wiring himself to a random-number generator, but such a series of forecasts would be completely useless.'

In addition to the tests of forecast encompassing, we consider whether forecasts are efficiently updated, in the sense that forecast revisions are unpredictable from information available at the forecast origin, such as the level of sentiment, or change in sentiment. In terms of (7), we introduce $(s_{t-h_2} - s_{t-h_3})$ as an additional explanatory variable, and set $\beta = 1$ to give the test regression:

$$\widehat{y}_{t|t-h_1} - \widehat{y}_{t|t-h_2} = \alpha + \gamma \left(s_{t-h_2} - s_{t-h_3} \right) + u_t. \tag{8}$$

This is a test of whether $s_{t-h_2} - s_{t-h_3}$ 'leads' $\widehat{y}_{t|t-h_1}$, in the sense that past values of s predict subsequent values of $\hat{y}$. Formally, the null is that the revision between the forecasts of y_t made at time h_1 and h_2 should be unpredictable at the time the longer horizon (h_2) forecast is made (called a 'strong-efficiency' test by Nordhaus (1987)) against the alternative that the change in the value of s between $t - h_2$ and $t - h_3$ has predictive power.

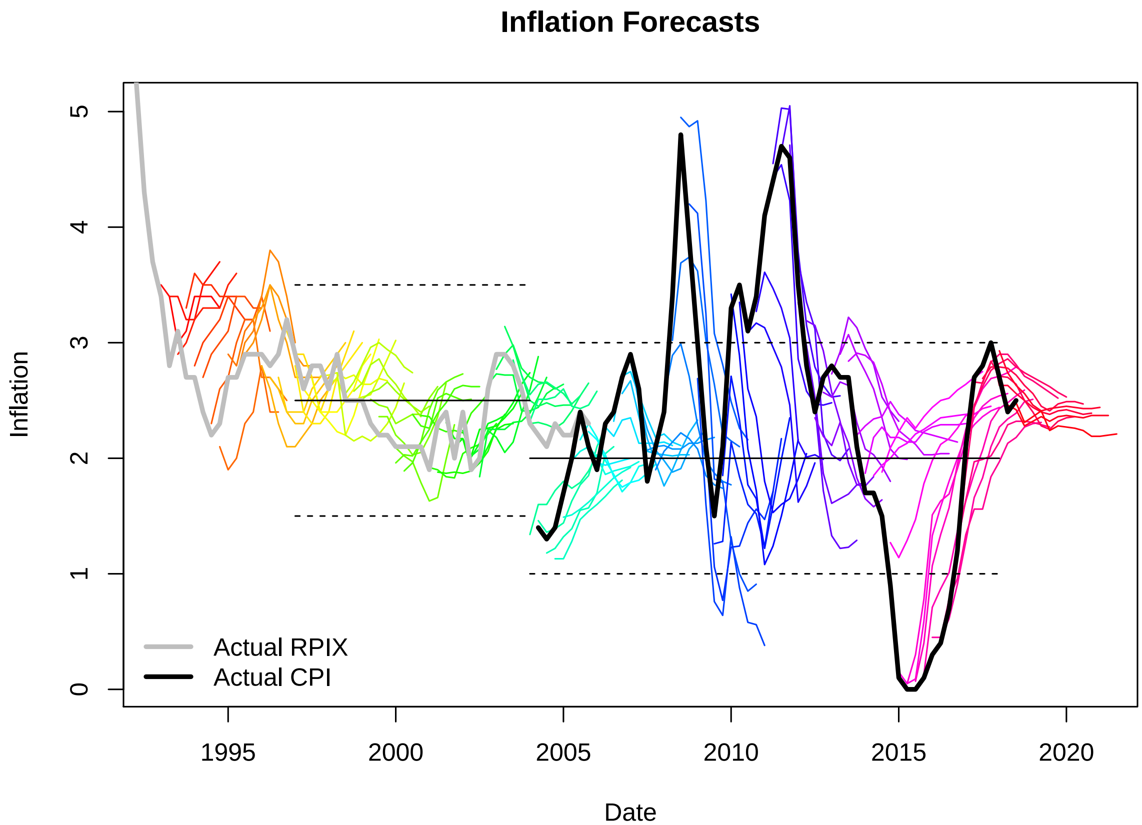

Figure 1: Forecasts and actual inflation since 1993 (forecasts to 2018).

5 Bank Forecast Data

The Bank has produced forecasts of inflation for up to between seven and thirteen quarters ahead since 1993. In August 1997 the Bank began also forecasting GDP growth in each *Report*, for eight quarters ahead. Each *Report* contains a number of additional forecasts that the Bank produces, but the MPC signs off on forecasts of these two variables. It also signs off on the unemployment rate forecasts, but only from 2013, so we do not consider these forecasts. We consider the forecasts and narratives in the *Reports* from 1997 to August 2018.

At the time of writing, since the first report in February 1993, there have been 100 such reports up to 2018. Forecasts are provided of various measures of the central tendencies of the forecast distributions of the macro variables, and we use the mean forecasts. For inflation the forecasts were for RPIX (the retail price index excluding mortgage interest) up to the end of 2003, and subsequently are of the CPI (consumer price index).

Some of the key properties of the forecasts are evident from Figures 1 and 2. Figure 1 shows the forecasts since 1993 alongside outcomes for inflation. Up to 2003 the Bank was instructed to target RPIX inflation at 2.5%, but after that CPI inflation at 2%. These are marked, along with the dotted lines marking the upper and lower bounds. In both periods there is clear evidence that forecasts move

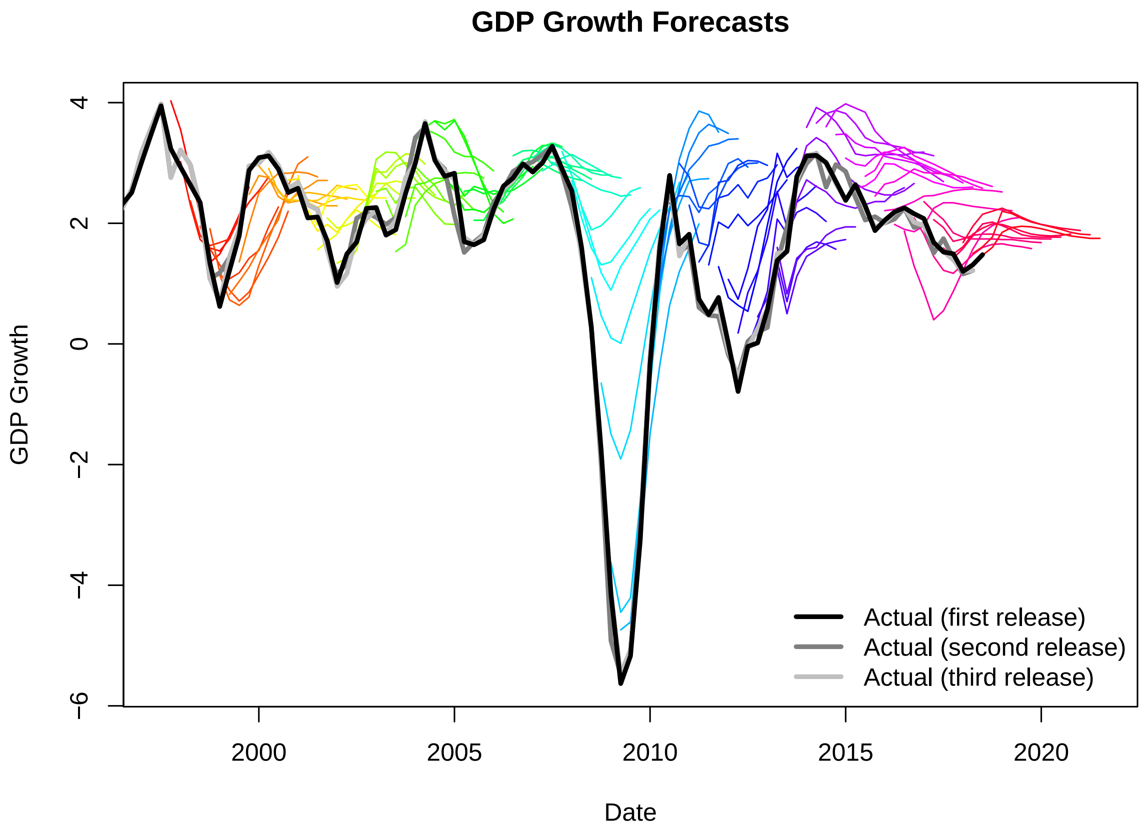

Figure 2: Forecasts and actual GDP growth since 1998 (forecasts to 2018).

towards the target values irrespective of the forecast origin values of inflation. As evident from the figure, the period up to 2003 was relatively quiescent in terms of inflation outcomes, but after this, and in particular after 2006, inflation became much more variable (and presumably less predictable). Figure 2 also suggests forecasts of output growth had an attractor, of around 2%.

The accompanying narratives have varied in size, as evident from the word counts recorded in Figure 3. Aside from the spike around the turn of the century (when MPC minutes were attached to *Inflation Reports*), the word count has generally trended upwards to its current level of around 35,000 words, from nearer to 20,000 words in 1993.

We construct two sentiment indices using the methods outlined in Section 3. The one called the 'Afinn' index is constructed using a dictionary provided by Nielsen (2011), and scores words, whereas the Alba index (Alba (2012)) builds on words used in *Inflation Reports* classified as positive, negative, decrementers, incrementers and inverters. We plot the resulting indices in Figure 4.

Both indices display the same broad patterns, with increasing sentiment throughout the 1990s, with record highs in the period 2000-2008 before a dramatic fall in line with the financial crisis (although slightly ahead of the GDP fall), and subsequent recovery. Our subsequent analysis is based on the Alba index.

As described earlier, it may be important to scale these measures of sentiment prior to their use

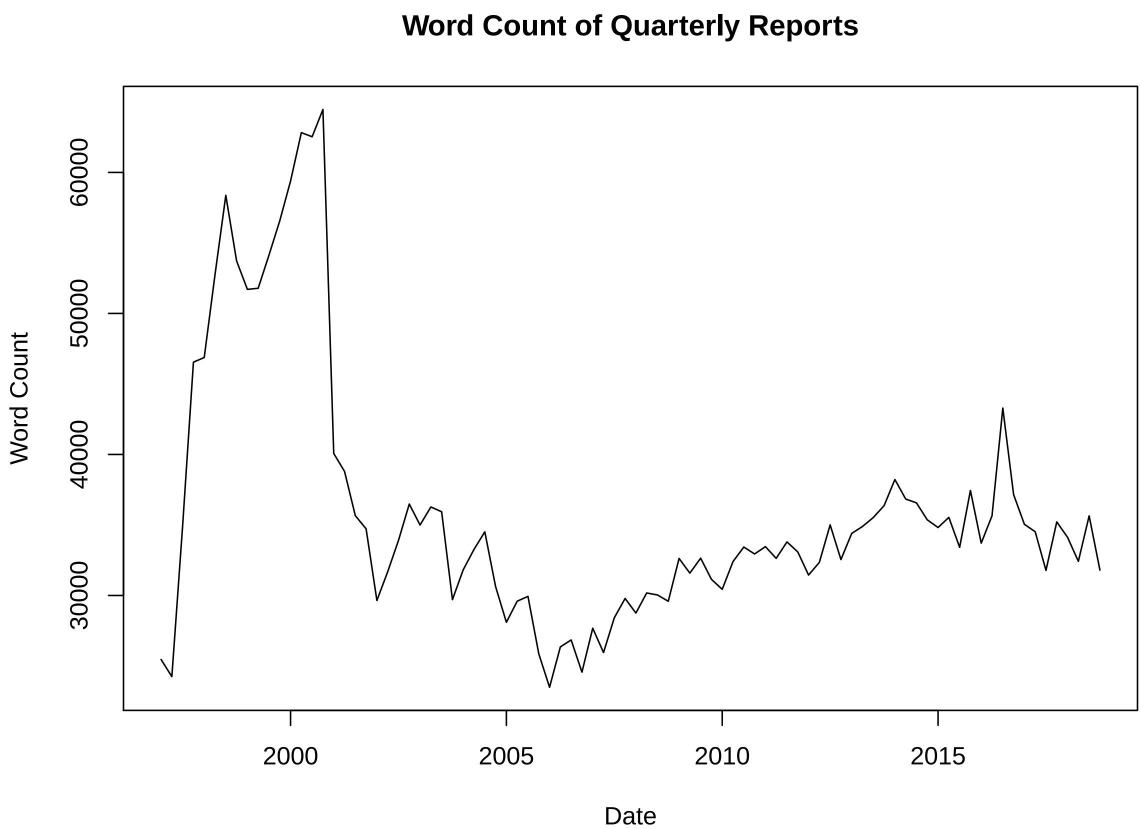

Figure 3: The word count of *Inflation Reports* since 1993.

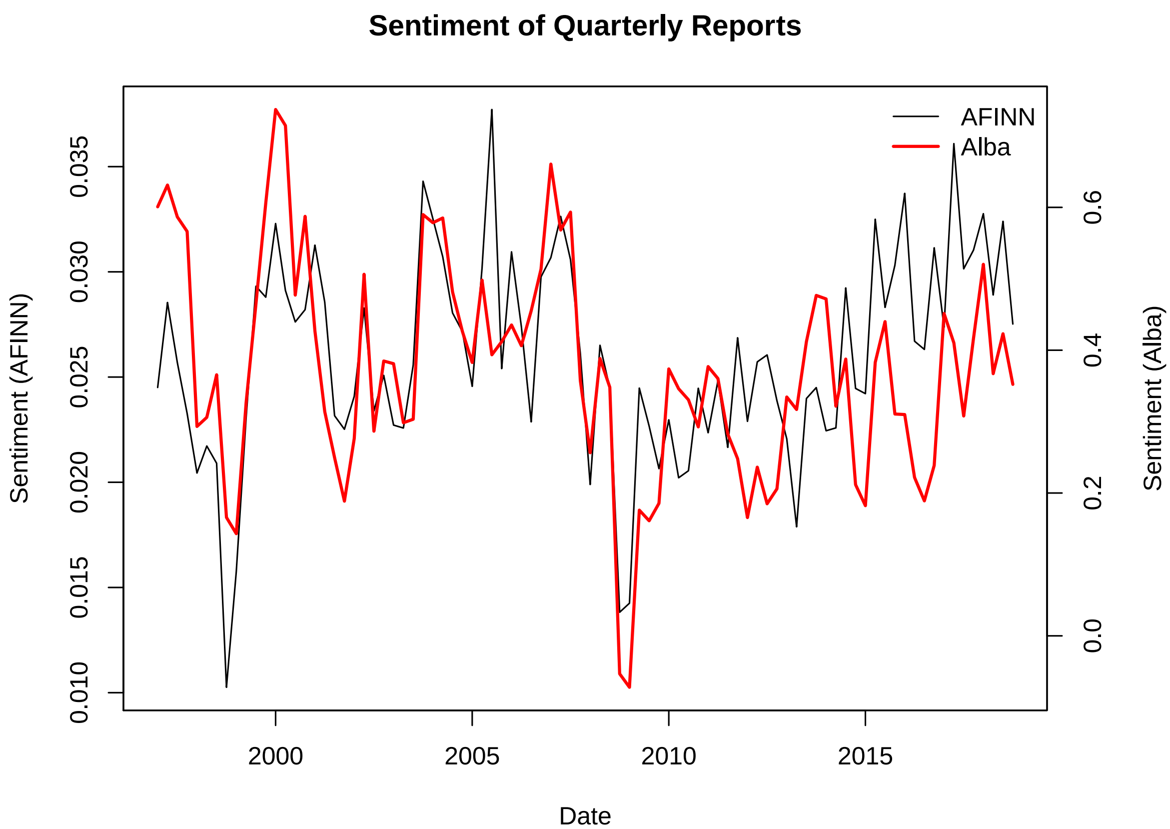

Figure 4: The two sentiment measures, plotted in raw, sentiment-per-sentence, form.

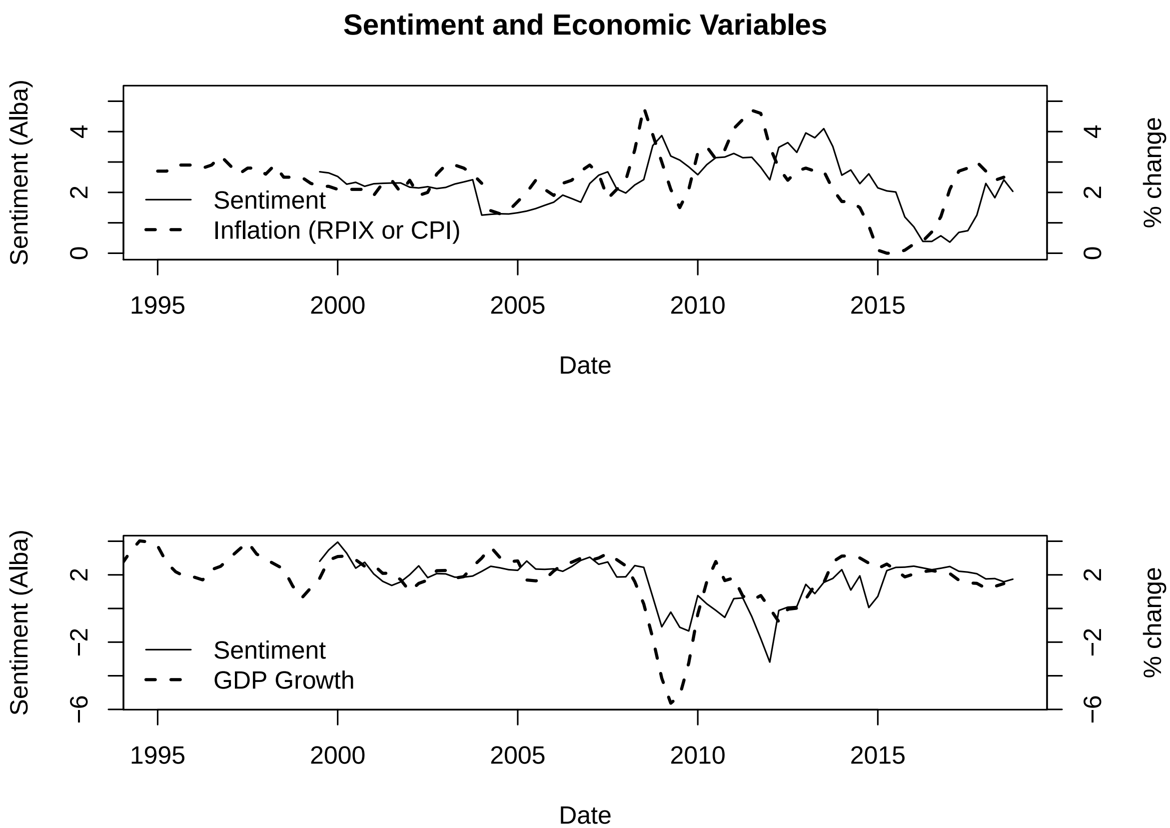

Figure 5: Scaled sentiment for inflation (top panel), and GDP growth (bottom panel).

as forecasts. We use the same indices for inflation and output growth, and scale them separately by recent inflation or output growth depending on the variable being forecast, as described in section 4. The resulting plots are in Figure 5, and the impact of scaling is clear, with the resulting series much more closely following the trajectories of each data series throughout the sample period.

6 Results

6.1 Forecast Accuracy

We report Diebold-Mariano regressions (specified in (1)) of the Bank's forecasts against standard time-series models, and against forecasts derived from the narratives - the sentiment indices.[1] We might expect the Bank's forecasts to outperform the simple time-series forecasts at short horizons, but to struggle as the horizon increases (see Clements (2015) for a comparison of survey expectations and time-series forecasts). For output we use an AR(1) model for the growth rate, and for inflation a 'no-change' in growth rate forecast, because the inflation rate is close to having a unit root. For both output growth and inflation the negative coefficients in table 2 indicates the Bank forecast performs better than the benchmark, but only significantly so for inflation.[2]

We also compare the Bank forecasts against the sentiment index forecasts. Unsurprisingly, the Bank's numerical forecasts are more accurate than those based on the (scaled) sentiment indices in this head-to-head, with large differences, although the statistical significance of these differences only holds at the shorter horizons. More pertinently, we are interested in whether the sentiment indices nevertheless convey useful information about the future evolution of these variables, and we consider this below.

6.2 Forecast Efficiency

Before considering whether the possible incremental value of the sentiment indices, we consider whether the Bank forecasts are efficient in the Mincer-Zarnowitz sense. The Mincer-Zarnowitz regressions are defined in (3), and we report outcomes for inflation in Table 3, and for GDP growth in Table 4. Because the regression equation is parameterized with the forecast error as the dependent variable, both the constant and slope ought to be zero. The null is rejected at the 10% level for inflation at $h = 2$, and at more stringent levels of significance at longer horizons, and for output growth at all horizons.

These results suggest that the numerical forecasts do not make an efficient use of all the forecast origin information, in the sense that the forecast errors are predictable from information available at the time the forecasts were made, namely, from the forecasts themselves. Our primary interest is in whether the numerical forecasts' errors may be predictable from other information available at the forecast origin: namely, the sentiment index. We investigate this using forecast encompassing regressions.

6.3 Forecast Encompassing Tests

Although the head-to-head comparisons of the Bank and sentiment forecasts favour the Bank's numerical forecasts, the sentiment forecasts may nevertheless contain useful additional information not present in the numerical forecasts. We begin by running forecast encompassing tests to see whether the numerical

[1] All the test statistics in this paper are based on heteroscedasticity and autocorrelation-corrected standard errors.

[2] More accurate forecasting models could undoubtedly be found. Models which make use of related series, and higher-frequency data, as in, e.g., Ghysels, Sinko and Valkanov (2007) and Clements and Galvão (2008) might give better forecasts, but this is not our focus.

forecasts encompass the forecasts from the scaled sentiment index: see table 5 for inflation, and tables 6 and 7 for output growth, where the actuals are taken to be the initial release and the second revision, respectively. The tests are based on equation (6), although the dependent variable is the forecast error for the numerical forecasts, and we include the numerical forecast as an additional explanatory variable along with sentiment. We present each forecast horizon in a separate column, with robust standard errors in parentheses. For inflation there is no evidence against the null hypothesis that the sentiment forecast is encompassed: we do not reject the null hypotheses that the parameters on forecast sentiment are zero at all horizons.

Contrasting results are found for output growth. The sentiment forecast is statistically significant at the two shortest horizons. This suggests that for output growth at $h = 1$ and $h = 2$ an optimal combination of the numerical and sentiment forecasts would assign a non-zero to the sentiment forecast. However we are mindful that these regressions are in-sample, and that because we are testing 8 horizons the probability of a "false positive" is larger than the nominal significance level.

The choice of initial release or second revision actuals for output growth makes little difference. For our measures of inflation, revisions are inconsequential. [3]

The findings for both output growth and inflation are largely unchanged in terms of statistical significance if instead we use unscaled sentiment. The coefficients change to reflect the (absence of) scaling. See tables 8 and 9.

6.4 Forecast Updating

Turning to adjustments to forecasts, we consider updates to forecasts between adjacent *Inflation Reports*. A requirement of an efficient forecast - in the sense of making use of all the available information - is that subsequent revisions to the forecasts of the same target (here, y_t) should not be predictable using information available at the time of the original forecast. A key focus is whether sentiment, as expressed in the narratives accompanying the forecasts, 'lead' the numeric forecasts. This gives rise to the regression given in equation (8), with the (lagged) change in the sentiment index as the sole regressor. More general specifications could be allowed, and we could consider the persistence in the numerical forecast updates themselves, but we do not do so here.

Tables 10 and 11 report the results. There is no evidence that changes in the (scaled) sentiment index lead changes in the Bank's numeric forecasts for output growth or for inflation. For output growth, however, the negative (statistically significant) intercepts indicate that forecasts have been revised down as the forecast horizon shortens, suggesting an element of predictability.

7 Conclusions

We have considered the extent to which the narratives surrounding the Bank's published forecasts provide additional information regarding the inflation outlook. Not surprisingly, the Bank's numerical forecasts are more accurate than forecasts derived from the narratives as sentiment indices. There is some evidence that the sentiment indicator provides additional useful information - relative to the Bank's numeric forecast - for output growth at the shorter horizons, but not for inflation. We also considered the possibility that sentiment might 'lead' changes in the short-horizon numeric forecasts. This transpires

[3] Croushore (2011) provides a recent survey of forecasting and evaluating forecasts when data are subject to revision. Note we are use a consumer price index measure rather than a deflator.

not to be the case for either variable. However, we did find that revisions to the output growth forecasts were in part predictable.

These findings suggest that the numeric output growth forecasts do not exhaust all the information in the *Reports*. They were also found not to depend on whether we use the 'raw' sentiment index, or the linear projection of the macro-variable on the index.

There are a number of avenues for future research. We have looked at the relationship between changes in the narratives and (subsequent) changes in the numerical forecasts, although allowing more complicated dynamics more generally might prove fruitful, and requiring the relationship to be between the changes might be restrictive.

There might be better ways to construct sentiment indicators. Sentiment might become more important if we consider disaggregated measures, and this might be a fruitful avenue for future research.

Given the interest in density forecasting, and the Bank's published history of inflation fancharts, a natural extension of our research is to a consideration of the relationship between the narratives and statistics derived from the fancharts, such as estimates of uncertainty and downside and upside risks.

Finally, we have presented an in-sample evaluation of the relationship between the numerical and text forecasts, whereas an out-of-sample analysis would cast further light on the robustness of our findings.

References

Alba, J. (2012). *Basic Sentiment Analysis.* GitHub, GitHub repository, https://github.com/fjavieralba/basic_sentiment_analysis.

Atkeson, A., and Ohanian, L. (2001). Are Phillips Curves useful for forecasting inflation?. *Federal Reserve Bank of Minneapolis Quarterly Review*, **25**, 2–11. (1).

Bird, S., Loper, E., and Klein, E. (2009). *Natural Language Processing with Python*: OÃćÂĂÂŹReilly Media Inc.

Capistrán, C., and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, **41**, 365–396.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting breaks and during breaks. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, pp. 315–353. Oxford: Oxford University Press.

Castle, J. L., Hendry, D. F., and Martinez, A. B. (2017). Evaluating forecasts, narratives and policy using a test of invariance. *Econometrics*, *5*(3).

Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series.* Oxford: Clarendon Press.

Clark, T. E., and McCracken, M. W. (2011). Testing for unconditional predictive ability. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic*, pp. 415–440: Oxford University Press.

Clements, M. P. (2004). Evaluating the Bank of England density forecasts of inflation. *Economic Journal*, **114**, 844 – 866.

Clements, M. P. (2009). Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and*

Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 8, pp. 206–226. Oxford: Oxford University Press.

Clements, M. P. (2010). Explanations of the Inconsistencies in Survey Respondents Forecasts. *European Economic Review*, **54(4)**, 536–549.

Clements, M. P. (2015). Are professional macroeconomic forecasters able to do better than forecasting trends?. *Journal of Money, Credit and Banking*, **47,2-3**, 349–381.

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, **26**, 546–554. No. 4.

Clements, M. P., and Hendry, D. F. (2005). Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, **67**, 931–956.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

Croushore, D. (2011). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.

Di Fatta, G., Reade, J., Jaworska, S., and Nanda, A. (2015). Big Social Data and Political Sentiment: the Tweet Stream during the UK General Election 2015 Campaign. Proceedings, 8th IEEE International Conference on Social Computing and Networking (SocialCom 2015).

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, **27(1)**, 30–41.

Ericsson, N. R. (1992). Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *Journal of Policy Modeling*, **14**, 465–495.

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

Fair, R. C., and Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, **80**, 39–50.

Fawcett, N., Körber, L., Masolo, R., and Waldron, M. (2015). Evaluating UK point and density forecasts from an estimated DSGE model: the role of off-model information over the financial crisis. Staff working paper 538, Bank of England.

Gentzkow, M., Kelly, B. T., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*. Forthcoming.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, **26**, 53–90.

Giacomini, R. (2011). Testing conditional predictive ability. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic*, pp. 441–455: Oxford University Press.

Granger, C. W. J., and Newbold, P. (1973). Some comments on the evaluation of economic forecasts.

Applied Economics, **5**, 35–47. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Hansen, S., and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*(S1), 114–133.

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

IEO (2015). Evaluating forecast performance. Discussion paper, Independent Evaluation Office, The Bank of England.

Jones, J. T., Sinclair, T. M., and Stekler, H. O. (2019). A Textual Analysis of Bank of England Growth Forecasts. *International Journal of Forecasting*. Forthcoming.

Lucca, D. O., and Trebbi, F. (2009). Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. Nber working papers 15367, National Bureau of Economic Research, Inc.

Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pp. 1–46: NBER.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, **62**, 902–917. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Nielsen, F. Å. (2011). *AFINN*. Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, http://www2.imm.dtu.dk/pubdb/p.php?6010.

Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. *Review of Economics and Statistics*, **69**, 667–674.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*: John Wiley and Sons, Ltd.

Patton, A. J., and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, *30*(1), 1–17.

Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Association for Computational Linguistics.

Pesaran, M., and Weale, M. (2006). Survey Expectations. In Elliott, G., and Timmerman, A. (eds.), *Handbook of Economic Forecasting*, Vol. 1, pp. 715–776. Elsevier.

Romer, C. D., and Romer, D. H. (2008). The FOMC versus the Staff: Where can monetary policymakers add value?. *American Economic Review*, **98:2**, 230–235.

Rosa, C., and Verga, G. (2008). The Impact of Central Bank Announcements on Asset Prices in Real Time. *International Journal of Central Banking*, *4*(2), 175–217.

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

Stock, J. H., and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, **44**, 293–335.

Table 2: Diebold-Mariano Test Statistics (DM) of Equal Forecast Accuracy Between the Bank Forecasts, and Benchmark Forecasts or Sentiment Indices.

	Inflation. Bank forecasts versus 'no-change' forecasts.							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DM	−0.146***	−0.304***	−0.356**	−0.247	−0.044	0.048	−0.100	−0.409
	(0.049)	(0.111)	(0.160)	(0.204)	(0.277)	(0.360)	(0.448)	(0.462)
Observations	104	103	102	101	100	99	98	95
	GDP growth. Bank forecasts versus an autoregressive model.							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DM	−0.366	−1.111	−1.606	−1.541	−0.680	0.026	0.544	0.868*
	(0.311)	(0.912)	(1.379)	(1.483)	(0.892)	(0.533)	(0.470)	(0.515)
Observations	84	83	82	81	80	79	78	77
	Inflation. Bank forecasts versus scaled sentiment forecasts.							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DM	−0.916***	−1.021***	−0.947***	−0.708	−0.403	−0.242	−0.366	−0.583
	(0.232)	(0.293)	(0.325)	(0.435)	(0.704)	(0.880)	(0.910)	(0.815)
Observations	78	77	76	75	74	73	72	71
	GDP growth. Bank forecasts versus scaled sentiment forecasts.							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DM	−1.594*	−1.497	−1.537	−1.366	−1.109	−0.884	−0.611	−0.520
	(0.820)	(0.989)	(1.117)	(1.184)	(0.923)	(0.848)	(0.910)	(1.125)
Observations	77	76	75	74	73	72	71	70
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 3: Mincer-Zarnowitz regressions for inflation.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.048	0.218	0.484	1.083	2.040**	2.734***	2.880***	2.840***
	(0.053)	(0.229)	(0.457)	(0.683)	(0.899)	(0.932)	(0.836)	(0.776)
Forecast	−0.028	−0.102	−0.203	−0.439**	−0.839***	−1.141***	−1.203***	−1.187***
	(0.019)	(0.082)	(0.159)	(0.216)	(0.295)	(0.358)	(0.338)	(0.295)
Observations	104	103	102	101	100	99	98	95
R^2	0.020	0.044	0.064	0.124	0.235	0.337	0.351	0.286
F Test of Efficiency	1.727	2.582*	3.434**	7.138***	15.683***	25.97***	27.323***	19.507***

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

Table 4: Mincer-Zarnowitz regressions for GDP growth (first release).

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	−0.307***	−0.432**	−0.547	−0.599	−0.428	0.150	1.227	2.419***
	(0.075)	(0.216)	(0.419)	(0.748)	(1.084)	(1.321)	(1.226)	(0.868)
Forecast	0.045	0.069	0.067	0.017	−0.120	−0.405	−0.856**	−1.321***
	(0.027)	(0.075)	(0.139)	(0.256)	(0.359)	(0.435)	(0.392)	(0.293)
Observations	84	83	82	81	80	79	78	77
R^2	0.03	0.024	0.009	0	0.005	0.027	0.07	0.116
F Test of Efficiency	13.769***	9.561***	7.781***	7.618***	8.207***	10.061***	13.286***	16.7***

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

Table 5: Encompassing regression for each forecast horizon for inflation forecasts, regressing forecast errors on scaled sentiment forecasts. Robust standard errors in parentheses.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.047	0.216	0.405	0.838**	2.003***	3.651***	4.689***	4.892***
	(0.051)	(0.194)	(0.328)	(0.411)	(0.536)	(0.546)	(0.925)	(1.040)
Forecast	−0.008	−0.039	−0.124	−0.341	−0.919**	−1.616***	−1.889***	−1.822***
	(0.020)	(0.097)	(0.175)	(0.278)	(0.380)	(0.338)	(0.351)	(0.331)
Scaled Sentiment	−0.017	−0.063	−0.043	0.016	0.075	−0.009	−0.215	−0.336
	(0.024)	(0.085)	(0.161)	(0.221)	(0.241)	(0.224)	(0.230)	(0.250)
Observations	78	77	76	75	74	73	72	71
R^2	0.014	0.033	0.033	0.063	0.171	0.342	0.395	0.342
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 6: Encompassing regression for each forecast horizon for GDP growth (initial release) forecasts, regressing forecast errors on scaled sentiment forecasts. Robust standard errors in parentheses.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	−0.457***	−0.627***	−0.773**	−0.954*	−0.946	−0.497	0.629	2.279**
	(0.068)	(0.191)	(0.299)	(0.534)	(0.898)	(1.276)	(1.271)	(0.958)
Forecast	−0.026	−0.004	0.048	0.065	0.019	−0.209	−0.660	−1.249***
	(0.034)	(0.105)	(0.193)	(0.333)	(0.432)	(0.495)	(0.453)	(0.363)
Scaled Sentiment	0.169***	0.186**	0.119	0.088	0.047	0.036	0.001	−0.095
	(0.041)	(0.089)	(0.182)	(0.259)	(0.235)	(0.142)	(0.104)	(0.103)
Observations	77	76	75	74	73	72	71	70
R^2	0.235	0.131	0.043	0.016	0.002	0.007	0.04	0.1
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 7: Encompassing regression for each forecast horizon for GDP growth forecasts (second revision), regressing forecast errors on scaled sentiment forecasts. Robust standard errors in parentheses.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.494^{***}	-0.641^{***}	-0.757^{**}	-0.927^{*}	-0.909	-0.547	0.600	2.264^{**}
	(0.116)	(0.198)	(0.328)	(0.546)	(0.932)	(1.255)	(1.278)	(1.066)
Forecast	-0.024	-0.003	0.039	0.058	-0.013	-0.210	-0.653	-1.254^{***}
	(0.047)	(0.115)	(0.205)	(0.341)	(0.437)	(0.483)	(0.453)	(0.393)
Scaled Sentiment	0.183^{***}	0.186^{*}	0.116	0.075	0.062	0.058	-0.003	-0.084
	(0.042)	(0.102)	(0.198)	(0.266)	(0.232)	(0.145)	(0.112)	(0.116)
Observations	76	75	74	73	72	71	70	69
R^2	0.212	0.109	0.033	0.011	0.003	0.008	0.037	0.092
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 8: Encompassing regression for each forecast horizon for inflation forecasts, regressing forecast errors on unscaled sentiment forecasts. Robust standard errors in parentheses.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.026	0.156	0.456	1.067	2.265^{**}	3.459^{***}	3.772^{***}	3.537^{***}
	(0.074)	(0.259)	(0.524)	(0.756)	(1.046)	(1.014)	(0.791)	(0.626)
Forecast	-0.014	-0.063	-0.144	-0.350	-0.889^{***}	-1.535^{***}	-1.768^{***}	-1.590^{***}
	(0.018)	(0.078)	(0.161)	(0.221)	(0.332)	(0.406)	(0.402)	(0.235)
Unscaled Sentiment	-0.007	-0.047	-0.260	-0.481	-0.417	0.088	0.564	0.271
	(0.119)	(0.285)	(0.378)	(0.539)	(0.456)	(0.366)	(0.708)	(0.429)
Observations	88	87	86	85	84	83	82	81
R^2	0.006	0.019	0.037	0.08	0.187	0.347	0.389	0.317
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 9: Encompassing regression for each forecast horizon for GDP growth (initial release) forecasts, regressing forecast errors on unscaled sentiment forecasts. Robust standard errors in parentheses.

	Forecast Error (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	−0.593***	−0.770**	−0.811*	−0.745	−0.414	0.283	1.482	2.924***
	(0.121)	(0.312)	(0.442)	(0.598)	(0.927)	(1.254)	(1.191)	(0.834)
Forecast	−0.013	−0.023	−0.023	−0.047	−0.113	−0.323	−0.740	−1.201***
	(0.034)	(0.076)	(0.164)	(0.344)	(0.458)	(0.538)	(0.497)	(0.373)
Unscaled Sentiment	1.078***	1.415**	1.222*	0.780	−0.087	−0.912	−1.511	−2.262**
	(0.356)	(0.616)	(0.731)	(1.261)	(1.117)	(0.892)	(0.915)	(1.123)
Observations	84	83	82	81	80	79	78	77
R^2	0.132	0.084	0.03	0.005	0.005	0.032	0.085	0.15

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

Table 10: Numerical forecast updates and prior sentiment revisions.

	Forecast Adjustment for Inflation (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	−0.004	0.042	0.088	0.090	0.079*	0.024	−0.041	−0.057
	(0.050)	(0.052)	(0.056)	(0.069)	(0.043)	(0.045)	(0.039)	(0.039)
Change in Sentiment	−0.176	−0.244	−0.199	−0.111	0.036	0.027	−0.006	−0.066
	(0.168)	(0.200)	(0.174)	(0.176)	(0.082)	(0.071)	(0.066)	(0.065)
Observations	75	75	75	75	75	75	75	75
R^2	0.029	0.044	0.024	0.007	0.001	0.001	0.0001	0.006

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

Standard errors in parentheses beneath coefficient estimates.

Table 11: Numerical forecast updates and prior sentiment revisions.

	Forecast Adjustment for GDP growth (quarters ahead)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.102^{*}	-0.136^{**}	-0.176^{**}	-0.135^{*}	-0.102	-0.078	-0.039	-0.025
	(0.058)	(0.062)	(0.074)	(0.074)	(0.072)	(0.054)	(0.048)	(0.043)
Change in Sentiment	0.073	0.121	0.139	0.116	0.119	0.075	0.071	0.045
	(0.102)	(0.119)	(0.136)	(0.110)	(0.094)	(0.076)	(0.074)	(0.070)
Observations	75	75	75	75	75	75	75	75
R^2	0.016	0.032	0.030	0.021	0.034	0.018	0.019	0.009

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

Standard errors in parentheses beneath coefficient estimates.