The Impact of Sentiment on Judgement and Econometric Model Based Forecasts*

Ali Kabiri[a,d], John S. Landon-Lane[b], Kaiwen Qiu[b], Norman R. Swanson[b] Jacob Turton[c], and David Tuckett[d]

[a] *Department of Economics and International Studies, Buckingham University, Hunter Street, Buckingham, MK18 1EG, United Kingdom*

[b] *Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, 08836, U.S.A*

[c] *Department of Computer Science, University College London, London, Gower Street, London, WC1E 6BT, U.K.*

[d] *Center for the Study of Decision-Making Uncertainty, University College London, London, Gower Street, London, WC1E 6BT, U.K.*

October 2023

Abstract

In this paper we introduce an anticipatory sentiment index that is measured using textual data from US economic and financial news. In order to determine the usefulness of the index for forecasting, we carry out a series of experiments in which the marginal predictive content of the index is assessed. In particular, we first assess the predictive content of the index in econometric models that are used for real-time macroeconomic forecasting. We then assess whether our sentiment index contains information that can be utilized to improve the quasi-judgemental forecasts reported in the United States Survey of Professional Forecasts (US-SPF). This is done by testing the rationality of US-SPF forecasts when conditioning on both real-time macroeconomic variables and our sentiment index. We find that the use of sentiment can improve the forecasting performance of econometric models, and that improvements are the greatest for shorter forecast horizons. In addition, we find that US-SPF forecasts incorporate the information contained in our sentiment indexes for short forecast horizons, but less so for longer horizons. We conclude that our anticipatory sentiment index can be used to improve not only econometric model-based forecasts, but also judgement-based forecasts.

Keywords: Judgement, Real-time forecasting, forecasting efficiency, sentiment.

*We are grateful to Valentina Corradi and Kenwin Muang for useful comments and suggestions.

Email addresses: ali.kabiri@buckingham.ac.uk (Ali Kabiri), jlandonl@economics.rutgers.edu (John S. Landon-Lane), kq60@economics.rutgers.edu (Kaiwen Qiu), nswanson@economics.rutgers.edu (Norman R. Swanson), jacob.turton.16@ucl.ac.uk (Jacob Turton), d.tuckett@ucl.ac.uk (David Tuckett)

1 Introduction

The importance of the interplay between the use of judgement and econometric models when constructing predictions has been discussed in many papers in the economics literature. One common view amongst practitioners, as stated by Lawrence, Goodwin, O'Connor, and Onkal (2006), is that judgement is an " ... indispensable component of forecasting ...". Along these lines, Turner (1990) describe how differences in rival model and judgement-based forecasts can be explained by the use of judgemental adjustments. Zellner, Abbas, Budescu, and Galstyan (2021) further explores the relationship between the use of quantitative forecasting models and human judgement, and argues that neither approach uniformly dominates the other and that there is a need to understand the nature of the interaction between econometric models and judgement.

In this paper, we broaden the discussion by exploring the interplay between judgement, model-based forecasts, and sentiment measures extracted using big-data methods. In particular, we examine the marginal predictive content of a novel sentiment index for both model-based and judgemental forecasts.

The potential usefulness of sentiment for forecasting is discussed in a number of papers. for example, Johnson and Tuckett (2022) examine the role of sentiment as a mediating variable in judgement. They show the link between narratives - the interpretation of text by agents - and their subsequent forecast choices in a laboratory setting. Sharpe, Sinha, and Hollrah (2023) apply an emotion-word based lexicon to textual narratives of Federal Reserve 'Greenbook' forecasts, which combine model-based and judgement forecasts, and are able to improve economic forecasting outcomes for the US macroeconomy. Sharpe et al. (2023) argue that narrative sentiment is useful for incorporating new information into forecasts. In their framework, as well as ours, sentiment is closely related to the concept of 'affect' . The notion of 'affect' involves the idea that stimuli create an emotional based evaluation, which may be unconscious (see e.g. Slovic, Finucane, Peters, and MacGregor (2002)) and that the 'emotional core' of an attitude is the main determinant of many judgments and behaviors that can impact judgement through several potential channels (see Kahneman, Diener, and Schwarz (1999); Kahneman and Ritov (1994)). For example, "affect as information" is a scenario where decision makers use their affective state as information in their judgement process, as discussed in Slovic et al. (2002); Slovic and Peters (2006); Peters, Västfjäll, Gärling, and Slovic (2006); Schwarz (2002); Schwarz and Clore (2003) or, as a motivator of information processing styles and behavior, as discussed in Peters et al. (2006); Isen, Nygren, and Ashby (1988); DeSteno, Petty, Rucker, Wegener, and Braverman (2004).

The approach taken in this paper is to use the idea from modern neuroscience literature that all cognition (knowing) is based on a mixture of cognitive skills and affect (see Shapiro (2011)). Moreover, such emotions / feelings are often useful cues to judgement because they track real information about the environment. Similarly, the use of feelings, rather than cognition alone can improve an agents ability to interpret their environment (see Loewenstein, Weber, Hsee, and Welch (2001); Pham (2004); Schwarz (2002)). When constructing our sentiment index, we use methods from the Natural Language Processing (NLP) literature in neuroscience to construct a sentiment index from text contained in a rich dataset of US economic and financial news. We then evaluate the usefulness of our sentiment index for forecasting by carrying out a series of prediction experiments. In our first set of experiments, we construct real-time macroeconomic predictions for a group of variables including industrial production, inflation, interest rates, and unemployment. These predictions are formulated using models that contain sentiment, in order to assess the marginal predictive context of the index, when conditioning on other macroeconomic variables. In our second set of experiments, we evaluate the interplay between judgement and sentiment by examining quasi-judgement forecasts from the United States Survey of Professional Forecasters (US-SPF). In particular, we carry out forecast rationality tests in order to assess whether US-SPF predictions fully incorporate the information contained in our sentiment index.[1]

Our results from our real-time forecasting experiment shows that incorporating sentiment into econometric models leads to lower out-of-sample forecast errors for short-term forecasting horizons. In seven out of ten cases the best econometric forecasting model is a model that incorporates sentiment. For longer-term forecast horizons the evidence is weaker. In only four out of ten cases the best forecasting model includes sentiment. We then test whether the forecasting performance from the quasi-judgemental forecasts of the US-SPF can be improved by including sentiment. We find that the forecasts from the US-SPF incorporates the information contained in our sentiment series for short term forecast horizons but less so for longer-term forecast horizons, notably for the interest rate variables.

The rest of the paper is structured as follows. In Section 2, we explain how textual analysis is used to construct our anticipatory sentiment index. In Section 3, we describe the real-time data used in our experiments, and discuss the importance of carrying out truly real-time forecasting experiments when comparing judgement based and model-based forecasts. Section 4 contains details on our empirical setup and Section 5 summarizes our findings based on our empirical analysis. Finally, Section 6 provides concluding remarks.

[1]Clements, Rich, and Tracy (2023) surveys the literature on the use of Surveys of Professional Forecasters from the United States and from Europe (ECB-SPF) in the context of using these forecasts to elicit information about expectations and discusses the nature of these forecasts with relation to judgement. Studies by the European Central Bank (de Vincent-Humphreys, Dimitrova, Falck, & Henkel, 2019) and by Stark (2013) describe the mix of formal modelling and judgement in the US-SPF and the European ECB-SPF. In particular both Stark (2013) and an ECB report based on a 2008 survey reports that forecasts contain a mix of econometric forecasts and subjective adjustments. In a 2019 report on the third ECB survey of forecasters, the differences between short-horizon and longer-horizon forecasts are described. For shorter-horizon forecasts, professional forecasters reported that they use a mix of reduced-form econometric models and judgement, while for longer-horizon forecasts, professional forecasters use a mix of structural econometric models and judgment. This survey does not distinguish between the use of judgement to adjust a forecast from an econometric model and the use of judgement to select the appropriate structural model for longer forecast horizons.

2 A New Sentiment Index

Using word lexicons to measure sentiment from text has a long history. The Harvard General Inquirer (Harvard GI, subsequently Harvard-IV) project of Stone, Dunphry, Smith, and Ogilvie (1966) constructed a lexicon consisting of 'positive' and 'negative' words. This list was rather large and represents something of a blunt instrument to measure underlying sentiment contained in text. Loughran and McDonald (2011) noted that this list was not useful for measuring sentiment in financial statements as words like 'liability' had a specific and not necessarily negative context when reading financial statements. Loughran and McDonald (2011) pared down the Harvard-IV lexicon into a lexicon that is better applied to financial text. Tuckett, Smith, and Nyman (2014) convened a group of experts and proposed an alternative lexicon based on the concept of 'approach' and 'avoidance'. This lexicon was based on using textual analysis to uncover the underlying narrative sentiment contained in text.

Our measure of sentiment is based on the intuition that news stories not only communicate factual news about the economy, financial markets and politics, but also have sentiment embedded in the language used in these news items. It is therefore plausible that forecasts of the future state of the economy and financial markets use such information. We define our sentiment index as the balance between the normalised frequency of specific sentiment words that are most likely to indicate 'risk or reward' sentiment. Our analysis of sentiment connects more broadly to work in behavioral finance studying how sentiment or biased beliefs impact asset prices, as discussed in Barberis (2018), and to studies of the impact of sentiment on the macro-economy, as discussed in Lagerborg, Pappa, and Ravn (2022).

Our sentiment index is constructed using a refinement of the Relative Sentiment Shift (RSS) lexicon of Tuckett et al. (2014). This refinement is based on the underlying semantic features of the words in the RSS lexicon and is designed to capture anticipatory sentiment as opposed to reactionary sentiment. In particular, we follow Turton, Kabiri, Tuckett, Smith, and Vinson (2021), who separates the 'avoidance' component of the RSS lexicon into two mutually exclusive lexicons, one containing avoidance words that 'load' up on 'Cognition' and 'Drive' and the other containing avoidance words that 'load' up on 'Fearful' and 'Surprised' semantic features.[2] For a technical discussion of the Turton et al. (2021) methodology, see Turton, Vinson, and Smith (2020), who discuss how to use Statistical Natural Language Processing (Stat.NLP) methods to create synthetic semantic features for words that are not analyzed in the set of semantic features assigned to the words in the dataset used by Binder et al. (2016).

These features in turn allow us for the identification of synthetic 'Binder'-features for the words in the original RSS lexicon from Tuckett et al. (2014). Summarizing, our approach uses regression-based neural networks to find the dominant Binder-features for the 'avoidance' (negative sentiment) words and for the 'approach' (positive sentiment) words that make up the RSS lexicon. Following

[2]Note that Turton et al. (2021) builds on both the work of Tuckett et al. (2014) and the work of Binder et al. (2016), who defines 65 human-calibrated semantic features for a limited sample of words which allow each word to be mapped to its neural correlates.

Turton et al. (2020), we find that the dominant Binder-features contained in the RSS avoidance lexicon are 'Fearful', 'Surprised', 'Cognition', and 'Drive'. In addition, we follow Turton et al. (2021), who separate the 'avoidance' component of the RSS lexicon into two mutually exclusive lexicons, one containing avoidance words that 'load' up on 'Cognition' and 'Drive' and the other containing avoidance words that 'load' up on 'Fearful' and 'Surprised' semantic features. In the sequel, these two sub-lexicons are labelled RSS(Alt-1) and RSS(Alt-2), respectively. [3] Kabiri, Landon-Lane, Turton, and Tuckett (2023) show that the sentiment derived from the RSS (Alt-1) sub-lexicon is significantly better at predicting economic uncertainty than both RSS (Alt-2) and other alternative sentiment lexicons used in the literature, including that used in Loughran and McDonald (2011).

The lexicon that we use, RSS(Alt-1), contains 'avoidance' words that load up on 'Cognition' and 'Drive'. 'Cognition' and 'Drive' are the semantic groups which we identify as reflecting agent's perceptions and hence, their decision-making when faced with new information. Binder et al. (2016) defines 'Cognition' to be "A form of mental activity or a function of the mind" and 'Drive' to be " ... someone or something that motivates you to do something ... ". Words that score relatively highly in 'Cognition' and 'Drive' include: 'apprehensive', 'anxious', 'distrust', 'doubts', 'frets', 'hesitating', 'jitters','misgive', 'nervy', and 'uncertain'. Kabiri, Landon-Lane, et al. (2023) argue that the sentiment associated with 'Cognition' and 'Drive' are anticipatory in nature.

We therefore use the RSS (Alt-1) sub-lexicon to construct a sentiment index from the Reuters Historical Archive Database.[4] All news articles tagged 'New York' or 'Washington D.C.' were used for the period of January 1996 to December 2019 to construct an anticipatory sentiment index based on these sentiment groups. All data tagged under the categories of "Arts, Culture & Entertainment","Oddly Enough", "Religion" & "Sports" were filtered from the data-set. The data set comprised 3.3 million news items. Kabiri, Landon-Lane, et al. (2023) show that this index is informative in understanding the dynamics of the US macro-economy and financial markets and that the anticipatory sentiment index has new information over and above that contained in many macroeconomic and financial variables.

The sentiment index used in our analysis, S_t, is constructed as follows. A "bag-of-words" approach is used. For each month we calculate the total number of words (Tot_t), the number of approach/excitement words, (App_t), and the number of avoidance/ anxiety words (Av_t). The sentiment index is then calculated using :

$$S_t = \frac{App_t - Anx_t}{Av_t}. \tag{1}$$

The sentiment index is depicted in Figure 1. The index is plotted as the solid line with its units on the right hand axis. Various macroeoconomic variables are also plotted in this figure, including output growth, inflation, unemployment, and interest rates. Of note is that sentiment during recessions typically fell with a large drop during the period of the Global Financial Crisis. Additionally,

[3] The two sub-lexicons differ in the list of 'avoidance' words. The approach' list is the same for both sub-lexicons.

[4] We thank Reuters/Refinitive for their generous sharing of these data.

the sentiment index is cyclical in nature with a number of spikes. These spikes line up with major events during this period. [5]

3 Data

The data used in the forecasting exercise is described in Table 1. We utilize monthly U.S. real-time macroeconomic data contained in the monthly FRED-MD database, which is available from the Federal Reserve Bank of St. Louis. This dataset contains 130 macroeconomic variables. In our analysis, we forecast five key macroeconomic indicators including CPI inflation (cpi_t), Industrial Production (ip_t), the Civilian Unemployment Rate (UR_t), the 3-Month Treasury Bill Rate ($IR3M_t$), and the 10-year Treasury Bond Rate ($IR10Y_t$). Our historical dataset spans the period 1996:3 - 2019:12.

The full list of all macroeconomic variables in the FRED-MD dataset is available upon request from the authors. Of note is that this dataset is truly real-time, in the sense that a "vintage" of data is available for each calendar dated observation. Consider the value of unemployment for June 2020. In July 2020, the government reported a "first release" value for June. In August of 2020, however, they further updated their "estimate" of unemployment for June. Namely, they reported a "second release" for January. This process of revision continues indefinitely. Namely, as the government changes data collection and processing methodology, collects new data and/or revises definitions of variables, new releases are reported. A "vintage" of data is a date, say June 2020. For unemployment, there is a whole vector of truly real-time data that includes different releases, all available in June 2020. For example, this vintage includes a 1st release value for May, a 2nd release value for April, and a 3rd release value for March, etc. In this sense, there is an entirely unique vintage of industrial production available each month, and the values of the calendar dated observations in each vintage change because the government updates historical values of the variable every month. Using this type of data allows the practitioner to truly simulate a forecasting environment in which models are updated at each point in time using data that were actually available at that time. Were we to simply collect industrial production data from a website today, calendar dated observations in our dataset from 2020 would reflect revisions that occurred after 2020. For variables that are revised regularly, this means that forecasting experiments of the type carried out in this paper would be invalid, in the sense that they would be utilizing "future data" as explanatory variables when estimating forecasting model regressions. For further discussion of the structure of real-time datasets, as well as methods for real-time forecasting, refer to Swanson (1996), Swanson, Ghysels, and Callan (1999), and Kim and Swanson (2018).

For variables that are subject to revision, this means that forecasting experiments of the type carried out in this paper would be invalid, in the sense that they would be utilizing "future data" as explanatory variables when estimating forecasting model regressions, were the correct vintage of data not used at each point in time when estimating models and constructing forecasts. This is

[5] See Kabiri, Landon-Lane, et al. (2023) for the identification of the events associated with the major spikes in the figure.

important, as participants in financial markets in the United States make real-time decisions based not only on first releases of key indicators such as GDP and unemployment, but they also make decisions based on updated (or new releases) of previously available data. In this sense, they are basing their decisions, at any given point of time on the first release of the most recent observation, the second release of the previous months' observation, the third release of the observation from the month before that, and so on. In the above jargon, they use the most recent vintage of data when making decisions. In our forecasting experiments reported below, we use the same strategy, updating the vintage of data used prior to model specification, estimation, and forecasting, at each point in time. This requires the use of notation when denoting our time series variables that is more general than simply denoting a calendar dated time series observation as y_t, say. For further details, refer to the next section.

We also analyze quarterly predictions from the monthly U.S. Survey of Professional Forecasters (SPF), available from the Federal Reserve Bank of St. Louis for all six of the above macroeconomic indicators. The SPF data, however, are available at a quarterly frequency, as is a historical real-time dataset that accompanies the survey predictions. Thus, for our experiments involving analysis of quarterly data, we solely utilize the SPF dataset, even though the variables analyzed remain the same, other than the frequency at which they are reported.[6] These data contain a significant judgmental component (Clements et al., 2023). Our historical quarterly SPF dataset spans the period 2007:3 - 2020:3. Following numerous papers that examine US. SPF data, we utilize median responses across all SPF respondents when carrying out our forecast efficiency tests. For discussion, see Baghestani (2005),Branch and Evans (2006), and Schorfheide and Song (2021).

Based on the results of application of Dickey-Fuller unit root tests, we transform our IP_t and CPI_t data using log differences. Namely, in the sequel we examine $ip_t = ln(IP_t/IP_{t-m})$ and $cpi_t = ln(CPI_t/CPI_{t-m})$, where $m = 1$ for monthly data and $m = 3$ for quarterly data. recall that we do this because our forecasting experiments utilize monthly data, while our efficiency test regressions use quarterly data. For these results, we aggregate our monthly macroeconomic indicators to the quarterly frequency, as detailed in Table 2.

Additionally, we examine the impact of sentiment during NBER business cycle recessionary and expansionary periods, with recessions defined as having occurred during the following periods: 2001:3 - 2001:11 and 2007:12 - 2009:6. These intervals were chosen to document the diverse effects of sentiment shocks on macroeconomic indicators across a spectrum of economic climates and events.

4 Empirical Methods

In this section, we discuss the experimental setup used in our analysis of: (i) the marginal predictive content of S_t relative to that of other observable variables in real-time econometrics forecasting

[6] Note that interest rate data used in our experiments are not real-time, in the sense that they are not revised over time. For this reason, we simplify our data collection by using FRED-MD interest rate data in all of our empirical experiments, regardless of data frequency. We do this by aggregating the monthly FRED-MD interest rate rate data using as detailed in Tables 1 and 2, where all our our target forecasting variables are listed.

equations; and (ii) the efficiency of largely judgement based predictions made by respondents to the Survey of Professional Forecasters, when including S_t and observable macroeconomic variables in real-time forecast efficiency test regressions.

4.1 Experimental Setup

In our forecasting experiments, we utilize either monthly real-time target variables for econometric model-based estimation or quarterly real-time variables for forecast efficiency testing.

When constructing econometric model-based forecasts, the number of lags as well as all model parameters are re-estimated prior to the construction of each new prediction, using rolling 120-month windows of data. Forecasts are constructed for horizons of h= 1, 3, 6, and 12 months-ahead. Our out-of-sample forecasting period is 2007:4 - 2020:1.

For efficiency testing, our quarterly data is for the period 2007:2 - 2020:1. For these regressions, we use SPF forecast errors as well as realizations on sentiment and various macroeconomic variables. Models are not re-estimated at each point in time, as our SPF forecasts are obtained directly from the Federal Reserve Board of Philadelphia website. Forecasts in these regressions are for horizons of h= 1, 2, 3, and 4 quarters-ahead. See below for further details.

Recall from Section 2 that sentiment, S_t, is constructed using a sub-lexicon based on the RSS lexicon of Tuckett et al. (2014). using methods proposed by Turton et al. (2021). As discussed in Kabiri, James, Landon-Lane, Tuckett, and Nyman (2023), the avoidance part of this sub-lexicon is made up of words that load heavily on 'Cognition' and 'Drive', two semantic features that reflect uncertainty about information agents are currently receiving and reflect on actions that agents might take. This sentiment index is used to investigate whether (i) the information in sentiment could be used to improved forecast accuracy, and (ii) whether quasi-judgemental forecasts are efficient relative to the information contained in our sentiment variable.

4.2 Econometric Model-based Forecasting

Since we utilize real-time data in our experiments, we require notation that accounts for the vintage of a calendar dated observation. For real-time data that are subject to revision, let ${}_{t+k}Y_t$ denote the release k value of variable Y for calendar date t. In this sense, setting $k = 1$ defines preliminary or first release data, while $k = 2$ corresponds to data that have been revised once, and so on, for $k = 1, ..., f$, where f denotes data that are fully revised. Note that this nomenclature allows for data that are not subject to revision (such as $IR3M_t$, $IR10Y_t$ and SPR_t), since any value of k yields the same data. Our econometric forecasting models are variants of the following equation:

Model 1: AR(SIC)+Macro+Sentiment:

$$ {}_{t+h+k}y_{t+h} = \alpha + \sum_{i=1}^{p1} \beta_i' \cdot {}_tY_{t-i} + \sum_{i=1}^{p2} \gamma_i' \cdot S_{t-i} + \epsilon_{t+h} \qquad (2) $$

where $Y_t = (ip_t, cpi_t, IR3M_t, IR10Y_t, UR_t)'$, and our target forecasting variable y_t is one of the six variables contained in Y_t. Here, $p1$ and $p2$ are selected using the Scharwz Information Criterion (SIC). Additionally, define $\beta_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3}, \beta_{i,4}, \beta_{i,5})'$. In the sequel, we set $k = 1$, which implies that interest lies in predicting first releases of h-step ahead values of y_t in the above model.

Model 2: AR(SIC)+Macro: In Model 1 (for variable $l = 1, \ldots, 5$), set $\gamma_i = 0$ for $i = 1, ..., p2$.

Model 3: AR(SIC)+ Sentiment: In Model 1 (for variable $l = 1, \ldots, 5$), set $\beta_{i,j} = 0$, for $i = 1, ..., p1$, and $j \neq l$.

Model 4: AR(SIC): In Model 1 (for variable $l = 1, \ldots, 5$), set $\beta_{i,j} = 0$, for $i = 1, ..., p1$, and $j \neq l$. Also, set $\gamma_i = 0$ for $i = 1, ..., p2$.

Models 1 to 4 are univariate regression models in which the parameters are assumed to be constant across the estimation sample. An alternative set of models are also estimated where the model coefficients are allowed to differ across the business cycle. In particular, we estimate a "switching regression" where model coefficients differ between business cycle recessions and business cycle recoveries. We use the NBER recession dates of 2001:3 - 2001:11 and 2007:12 - 2009:6. For these periods an indicator variable is defined and a switching regression model is defined with all variables interacted with this indicator variables. The switching regression versions of Models 1 through 4 are as follows:

Model 5: AR(SIC)+Macro+Sentiment Switching Regression:

$$\begin{aligned}
{}_{t+h+k}y_{t+h} = &\, I_0\{t\}(\alpha + \sum_{i=1}^{p1} \beta_i' \cdot {}_tY_{t-i} + \sum_{i=1}^{p2} \gamma_i' \cdot S_{t-i}) \\
&+ \sum_{k=1}^{K} I_k\{t_1 < t < t_2\}(\alpha_k + \sum_{i=1}^{p1} \beta_{i,k}' \cdot {}_tY_{t-i} + \sum_{i=1}^{p2} \gamma_{i,k}' \cdot S_{t-i}) + \epsilon_{t+h},
\end{aligned}$$

where $\beta_{i,k}$ and $\gamma_{i,k}$ are defined conformably and $I_k\{\cdot\}$ denotes the indicator function, which takes the value one when dates correspond to those stated in the argument to the indicator function, and equals zero otherwise. Additionally, $I_0\{\cdot\}$ takes the value one for all time periods other than those associated with non-zero values of $I_k\{\cdot\}$, for $k = 1, ..., K$, and equals zero otherwise. Note here that $K = 1$ so the $k = 0$ refers to business cycle expansions and $k = 1$ refers to business cycle recessions.

Model 6: AR(SIC)+Macro Switching Regression: In Model 5 (for variable $l = 1, \ldots, 5$), set $\gamma_i = 0$ for $i = 1, ..., p2$, and $\gamma_{i,k} = 0$, for $i = 1, ..., p2$ and $k = 1, ..., K$.

Model 7: AR(SIC)+ Sentiment Switching Regression: In Model 5 (for variable $l = 1, \ldots, 5$), set $\beta_{i,j} = 0$, for $i = 1, ..., p1$, and $j = 2, ..., 5$ and $\beta_{i,j,k} = 0$, for $i = 1, ..., p1,$, $j \neq l$, and $k = 1, ..., K$.

Model 8: AR(SIC) Switching Regression: In Model 5 (for variable $l = 1, \ldots, 5$), set $\gamma_i = 0$ for $i = 1, ..., p2$, and $\gamma_{i,k} = 0$, for $i = 1, ..., p2$ and $k = 1, ..., K$ and $\beta_{i,j,k} = 0$, for $i = 1, ..., p1,$, $j \neq l$, and $k = 1, ..., K$.

A summary of the econometric models used in this forecasting exercise is contained in Table 3. The benchmark model is **Model 4**, which is an autoregressive model with the number of lags chosen by the Schwarz Information Criterion (SIC).

Forecasting performance is evaluated using point mean squared forecast errors (MSFEs), where

$$MSFE = \frac{1}{P}\sum_{t=1}^{T}(y_{j,t} - \hat{y}_{j,t})^2,$$

and $\hat{y}_{j,t}$ denotes the prediction for target variable y_j that is made using data that are truly available in real-time at period t. However, in order to compare our findings with those of our benchmark AR(SIC) model (i.e., Model 4), in our tabulated results, MSFEs, relative to that of the benchmark AR(SIC) model are reported.

We also report the results of Diebold- and Mariano (DM) predictive accuracy tests (see Diebold and Mariano (1995)). Recall that the null hypothesis of the DM test is:

$$H_0 : \mathrm{E}[L(\hat{\epsilon}^{(1)}_{t+h})] - \mathrm{E}[L(\hat{\epsilon}^{(2)}_{t+h})] = 0,$$

where the $\hat{\epsilon}^{(i)}_{t+h}$ are prediction errors associated with model i, for $i = 1, 2$. Here, $L(\cdot)$ is a quadratic loss function, and the test statistic is

$$\mathrm{DM}_P = P^{-1}\sum_{t=1}^{P}\frac{d_{t+h}}{\hat{\sigma}_{\bar{d}}},$$

where $d_{t+h} = [\hat{\epsilon}^{(1)}_{t+h}]^2 - [\hat{\epsilon}^{(2)}_{t+h}]^2$, $\bar{d}$ denotes the mean of d_{t+h}, $\hat{\sigma}_{\bar{d}}$ is a heteroskedasticity and autocorrelation consistent estimate of the standard deviation of $\bar{d}$, and P denotes the number of ex-ante predictions used to construct the test statistic. Our tests utilize the asymptotic normality result discussed in Diebold and Mariano (1995) when constructing test critical values. For further discussion of critical values used when implementing this test under model nestedness conditions, and when accounting for parameter estimation error, refer to Corradi and Swanson (2006).

4.3 Efficiency Testing Using Judgement Based Forecasts

As discussed in Clements et al. (2023), professional forecasters including those surveyed in the U.S. Survey of Professional Forecasters (SPF) maintained by the Federal Reserve Bank of Philadelphia provide both model and judgement based point predictions for specific core macroeconomic variables at various forecast horizons. The same can be said of the ECB-SPF. For example, Clements et al. (2023) point out that surveys conducted by the ECB in 2008, 2013, and 2018 indicated a prevalence of judgement based forecasters among survey participants, although the number of model-based forecasters appeared to be increasing over time (see de Vincent-Humphreys et al. (2019) for further discussion). For these reasons, a natural vehicle for assessing the predictive linkage between judgment and sentiment is the SPF, with the caveat that forecasts in this survey are not solely judgement-based. In order to do this, we draw on the literature on both rationality and on the role of judgement in macroeconomic forecasting (see e.g. Clements (1995), Swanson and Van Dijk (2006) and the references cited therein).

To test for the rationality of SPF forecasts, and by proxy the rationality of judgment conditional on sentiment, early papers followed (Nordhaus, 1987), and fitted regressions of the form:

$$y_{t+h} = \alpha + \beta \hat{y}_{t+h} + \epsilon_{t+h}, \tag{3}$$

where $\hat{y}_{t+h}$ is a horizon h forecast and y_{t+h} denotes the actual observation on some target variable, y. When evaluating whether forecasts were unbiased and efficient (i.e., whether $\alpha = 0$ and $\beta = 1$, respectively) heteroscedasticity and autocorrelation consistent(HAC) standard errors were generally used for multi-step ahead forecasts, $h > 1$. In this case, "rationality" was deemed "true" is there was no bias and no inefficiency. This literature has seen many contributions over the last few decades. A key paper, for example, is Romer and Romer (2000) who propose testing forecast efficiency using real-time variables. In Swanson and Van Dijk (2006), the authors propose a forecast efficiency test that allows the investigator to determine the "timing" of data rationality, in the sense that one can assess how many months it takes before an early release of a variable is efficient, in the sense that all information available at the time of the first release of the data has been properly conditioned on. Faust and Wright (2008) construct two versions of OLS-based tests called "naive" and "transparency", and use conventional Wald statistics to test for forecast efficiency. In Ellingsen, Larsen, and Thorsrud (2022), the authors show that SPF forecasts under-perform traditional (real-time) econometric based forecasts for several key macroeconomic variables, while Corradi and Swanson (2012) introduce distribution based tests based on stochastic dominance principles for assessing the accuracy of SPF predictions.

Summarizing, in our experiment, we assess bias and efficiency of SPF forecasts. However, in our assessment, we condition not only on macroeconomic variables available at the time forecasts were made, but also on sentiment measures that were available in real-time. This approach allows us not only to re-assess the efficiency of SPF forecasts using the standard macroeconomic dataset, but also using our new sentiment dataset. In this manner, for example, we are able to assess whether efficiency with respect to macroeconomic variables carries over to sentiment. Is a variable that is efficient when accounting for real-time macroeconomic information still efficient when sentiment is added to the conditioning information set used when specifying our test regressions? Our test regression in this context is a variant of that discussed above, and follows most closely that used in Swanson and Van Dijk (2006). Namely, consider:

$$ {}_{t+h+k}y_{t+h} - \hat{y}^{SPF}_{t+h} = \alpha + \sum_{i=1}^{p1} \beta_i' \cdot {}_tY_{t-i} + \sum_{i=1}^{p2} \gamma_i' \cdot S_{t-i} + \epsilon_{t+h}, \tag{4}$$

where $\hat{y}^{SPF}_{t+h}$ denotes the h-step ahead SPF forecast made at time period t for ${}_{t+h+k}y_{t+h}$. Since we set $k = 1$, we are assuming that forecasters are interested in predicting first release data. Recall also that while our above forecasting experiments utilize monthly data, our efficiency test analysis is based on the use of quarterly data, given that SPF forecast surveys are carried out on a quarterly basis.

The forecast efficiency tests that we carry out involve testing the following hypotheses:

$$H_0 : \alpha = \beta = \gamma = 0$$
$$H_1 : \text{There exists at least one parameter which is not equal to zero.}$$

Note that if $\alpha \neq 0$ then forecasts are biased, while rejections due to β or $\gamma \neq 0$ imply inefficiency, in the sense that variables available at the time forecasts were made contained information that could have been used by forecasters to improve the accuracy of their predictions. Finally, note that in order to assess the impact of multicollinearity between Y_t and S_t on our test results, we carry out efficiency testing including both variables, following the above formulation, as well as variants thereof where we alternately exclude either Y_t or S_t.

5 Empirical Findings

5.1 Results from the forecasting experiment

A forecasting experiment was performed using the models described in Section 4.2. The forecasting experiment is based on real-time data using a rolling window of 120 months. Starting in April of 2007 we use data from the preceding 120 months to estimate eight econometric models reported in Table 3. The base of all models is the autoregressive (AR) model. The number of lags is chosen using the Schwarz Bayesian Informatation Criterion (SIC) for each estimation window separately. Thus we allow for the model to adjust as the sample period dates change. Once estimated, we use the model's estimated parameters to forecast 1, 3, 6, and 12 months ahead. Forecast errors are calculated and the mean squared forecast error (MSFE) is calculated for each model over the forecast window.

Table 4 reports the results of this forecast experiment. In this table, relative mean squared forecast errors are reported with the base model being Model 4 - the standard AR model with no sentiment included. A relative MSFE less than 1 for a given model is evidence that the model has a better overall forecast over the period from April 2007 to January 2020. The relative MSFE is reported for all five variables, for all four forecast horizons, for all models. Also reported in Table 4 are Diebold - Mariano (DM) test results for the hypothesis that given model's MSFE is significantly different from the base model's MSFE.

The results for the one month forecast horizon ($h = 1$) provide evidence that our sentiment variable has predictive content. For the growth rate of industrial production, the 3 month and ten-year interest rates, and for unemployment, models including sentiment have the lowest relative MSFE. For CPI-inflation, the lowest relative MSFE is also a model that includes sentiment but in this case the base model is better. For the forecast horizon of three months ($h = 3$), we get similar results. For the interest rate variables and unemployment, models with sentiment are the best models while for CPI-Inflation a model with sentiment is better than all others except for the base model. For the growth rate of industrial production the best model is a model without sentiment. It should be noted, however, that the best model with sentiment (Model 5) is only slightly worse.

The next part of Table 4 contains the results for longer forecast horizons of six months and a year. For the six month horizon ($h = 6$) the model with the best forecasting performance is the base model. For the interest rate variables (3-month and 10-year) models with sentiment are the best while for unemployment we see the switching regression model without sentiment (Model 8) has the lowest relative MSFE. The sentiment version of the switching regression model (Model 7) is only slightly worse, however. For the one year horizon ($h = 12$) we get the following results. For the growth rate of industrial production, the best models are the switching regression models. The switching regression models without the macro variables (Models 7 and 8) are the best with the switching model without sentiment slightly better. For CPI-inflation, the best model is a model without sentiment included. For the two interest rate variables, the switching models that include the macro variables are the best. For the three month interest rate the best model does not include sentiment while for the ten-year interest rate the best model does include sentiment. Finally for unemployment, the best model is a switching regression without sentiment.

To summarize, there is evidence for the short horizons that including sentiment in the econometric model leads to better forecast performance. For the short forecast horizons (of one and three months), models that include sentiment are the best for seven out of ten variable/ horizon pairs. This is evidence that the information contained in our sentiment variable is informative with respect to forecasting over short horizons. For the longer forecast horizons (of six to twelve months) there is less evidence that including sentiment improves econometric forecasts. In only four out of ten variable/ horizon pairs is a model with sentiment included the best performing model. Thus the evidence is weak for including sentiment in econometric forecasting models over longer horizons.

5.2 Evaluating Judgement Based Forecasts

We next move to evaluating the efficiency of the quasi-judgemental forecasts from the Survey of Profession Forecasters. We first test whether the SPF forecasts are efficient, and then test whether the inefficiency is due to the exclusion of sentiment from the forecast. The summary results of the efficiency tests are reported in Table 5 with more granular results reported in Table 6. It should be stressed that the SPF forecasts are quarterly. The forecast horizons are one, two, three, and four quarters respectively.

Looking first at the overall (Wald joint test) of efficiency we see that for the shortest horizon ($h = 1$) the null hypothesis of efficiency is not rejected for all variables except CPI-inflation. For that case, the evidence from the individual t-tests shows that the null hypothesis that the coefficient on CPI-inflation is zero is strongly rejected (p-value much less than 0.01) and the null hypothesis that the coefficient on sentiment (S_t) is also rejected (p-value of 0.0553).

For the two quarter horizon ($h = 2$), we see that the overall test of efficiency is rejected for two of the five variables (the short-term and long-term interest rates) but in these cases the rejections are due of the SPF forecasts not efficiently including information from one of the macro variables. For the case of the short-term interest rate the source of the inefficiency come from the short-term interest rate itself. The individual t-test for the null that the SPF forecast efficiently incorporates information from the short-term interest rate has a p-value of 0.0020. For the long-term interest rate the variable that is causing inefficiency is CPI-inflation. The individual t-test for the null that the SPF forecast for the long-term interest rate efficiently incorporates information from CPI-inflation has a p-value of 0.0264. Notably, we cannot reject the hypothesis that the SPF forecasts for all the variables at the horizon of two quarters efficiently incorporates information contained in our sentiment series, S_t.

As we move to longer forecasts horizons (three and four quarter horizons) there is evidence that the information from our sentiment series is not efficiently incorporated in the SPF forecasts for three of the five variables. For the growth rate of industrial production and for CPI-inflation there is no evidence that the longer horizon SPF forecasts are inefficient. For the two interest rates and for unemployment there is evidence that the SPF forecasts are inefficient. Based on individual t-tests, the null hypotheses that information from sentiment is efficiently incorporated in the SPF forecasts can be rejected for the short-term interest rate at both the three-quarter and four-quarter horizons. The p-values for these tests are 0.0184 and 0.0638 respectively. For the long-term interest rate, the null hypothesis that the SPF forecast efficiently incorporates information from our sentiment variable can be rejected for the three-quarter forecast horizon with a p-value of 0.0938. Similarly for unemployment, the null null hypothesis that the SPF forecast efficiently incorporates information from our sentiment variable can be rejected for the three-quarter forecast horizon with a p-value of 0.0076.

The results from the efficiency tests for the SPF forecasts roughly agree with our earlier results from the econometric models. The quasi-judgemental forecasts from the Survey of Professional Forecasts incorporate information from our sentiment variable over shorter horizons while there is evidence for three out of the five variables that the SPF forecasts do not efficiently incorporate information from our sentiment variable for the longer forecast horizons. Earlier we showed that econometric models could be improved by using our sentiment variable for shorter horizons but less so for longer horizons.

Our results are in accord with information obtained from a survey of the professional forecasters that make up the ECB-SPF, as discussed in de Vincent-Humphreys et al. (2019). In the survey, conducted in 2018, it was reported that short-horizon forecasts typically use reduced-form econometric models together with judgement while longer-horizon forecasts are often constructed using structural econometric models. It was also reported that longer-horizon forecasts additionally use judgement as an input to producing the forecast, not only in the sense that judgement informs the final prediction, but also in the sense that judgement is used to inform the decision concerning which structural model (if any) are used as an aid to judgement formation.

The above findings are consistent with one possible explanation for our different findings based on short versus long-horizon forecasts. In particular, assume that an SPF forecaster constructs their predictions based on both an econometric model component and a judgement component. Moreover, assume that one part of the judgement component involves choosing between competing econometric models. Furthermore, assume that the distribution of model-based predictions based on short-term models is more concentrated than that based on long-term models.[7] In this case, one might expect that forecasts become less efficient over longer horizons while sentiment becomes more relevant, since there are large differences between the models and judgements used in the formation of forecasts.

Tying into the above story, note that longer-horizon forecast errors may be more influenced by changes in trend or turning points in business cycles than short-horizon forecasts. Thus, there are likely to be more models to choose from at longer horizons. It is not surprising that the longer-run forecasts are inefficient in this context if one assumes that as the forecast horizon increases, a growing subset of the large set of models that SPFers choose from is inefficient.

6 Discussion and Conclusion

In this paper we introduce a sentiment index that loads heavily on anticipatory sentiment. We use this sentiment index to investigate whether the information contained in sentiment can improve econometric forecasts and whether quasi-judgemental forecasts are efficient relative to the information contained in our anticipatory sentiment series. Our results suggest that sentiment contains marginal predictive content that can be used to improve both econometric model-based forecasts and survey based judgement forecasts.

More specifically, incorporating sentiment into econometric models improves out-of-sample forecasting performance, particularly at shorter forecast horizons. This result is not surprising given the well known stylized fact that model based predictions are usually much more precise at short horizons.

Interestingly, though, the story changes when judgement type SPF forecasts are evaluated using efficiency tests. In this case, our evidence suggests that sentiment matters more for longer horizon forecasts than for shorter horizon forecasts. One reason for this may simply be that some participants of the SPF utilize both pure judgement and model based predictions when forming their forecasts, as discussed in de Vincent-Humphreys et al. (2019). Moreover, one might expect that the set of models from which judgement type forecasters extract information when making their forecasts grows as the horizon increases, and that the models incorporate increasingly disparate information sets. This in turn opens up the possibility that an increasing number of survey participants use inefficient model based forecasts, hence leading to the increased relevance

[7] This is likely to be the case for two reasons. First, it is well known that short-term forecasts are substantially more precise than long term forecasts, in general. Second, many SPF forecasters change from using simple reduced form models to using more complex structural models when increasing the horizon of their forecasts, as noted in the ECB survey discussed above.

of sentiment at longer horizons. A simpler explanation might be that our sentiment index is constructed using information that is heavily drawn from the financial markets. Thus, our index may offer more information for forecasting financial variables. However, it is very difficult to "beat" simple AR models at short horizons when predicting financial variables such as interest rates. It is not surprising, then, that most of the inefficiencies that we uncover correspond to cases where financial variables are forecast at multiple step-ahead horizons.

References

Baghestani, H. (2005). Improving the accuracy of recent survey forecasts of the t-bill rate: other survey information can help. *Business Economics*, *40*, 36–40.

Barberis, N. C. (2018). Psychology-based models of asset prices and trading volume. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Foundations and Applications 1* (pp. 79 – 176). Amsterdam: North-Holland.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*, 130–174.

Branch, W. A., & Evans, G. W. (2006). A simple recursive forecasting model. *Economics Letters*, *91*(2), 158–166.

Clements, M. P. (1995). Rationality and the role of judgment in macroeconomic forecasting. *Economic Journal*, *105*, 410–420.

Clements, M. P., Rich, R. W., & Tracy, J. S. (2023). Surveys of professionals. In R. Bachmann, G. Topa, & W. van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 71–106). Elsevier.

Corradi, V., & Swanson, N. R. (2006). Predictive density evaluation. In G. Elliott & A. Timmermann (Eds.), *Handbook of Forecasting* (pp. 197–284). Elsevier.

Corradi, V., & Swanson, N. R. (2012). A survey of recent advances in forecast accuracy comparison testing, with an extension to stochastic dominance. In X. Chen & N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White jr* (pp. 121–144). Springer.

DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., & Braverman, J. (2004). Discrete emotions and persuasion: The role of emotion-induced expectancies. *Journal of Personality and Social Psychology*, *86*(1), 43–56.

de Vincent-Humphreys, R., Dimitrova, I., Falck, E., & Henkel, L. (2019). Twenty years of the ECB survey of professional forecasters. *ECB Economic Bulletin*, *1*, 34–58.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Ellingsen, J., Larsen, V. H., & Thorsrud, L. A. (2022). News media versus fred-md for macroeconomic forecasting. *Journal of Applied Econometrics*, *37*(1), 63–81.

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

Isen, A. M., Nygren, T. E., & Ashby, G. F. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of Personality and Social Psychology*, *55*(5), 710–717.

Johnson, S. G. B., & Tuckett, D. (2022). Narrative expectations in financial forecasting. *Journal of Behavioral Decision Making*, *35*(1), 22–45.

Kabiri, A., James, H., Landon-Lane, J., Tuckett, D., & Nyman, R. (2023). The role of sentiment in the US economy: 1920 to 1934. *The Economic History Review*, *76*(1), 3-30.

Kabiri, A., Landon-Lane, J., Turton, J., & Tuckett, D. (2023). The role of emotion in credit markets. *Working Paper, Rutgers University.*

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: Foundations of hedonic psychology.* Russell Sage Foundation, New York.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty*, *9*(1), 5–37.

Kim, H.-H., & Swanson, N. R. (2018). Methods for backcasting, nowcasting and forecasting using factor-MIDAS: With an application to Korean GDP. *The Journal of Forecasting*, *37*, 281–302.

Lagerborg, A., Pappa, E., & Ravn, M. O. (2022). Sentimental Business Cycles. *The Review of Economic Studies*, *90*, 1358–1393.

Lawrence, M., Goodwin, P., O'Connor, M., & Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493-518.

Loewenstein, G., Weber, E., Hsee, C., & Welch, N. (2001, March). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65.

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

Peters, E., Västfjäll, D., Gärling, T., & Slovic, P. (2006). Affect and decision making: a "hot" topic. *Journal of Behavioral Decision Making*, *19*(2), 79-85.

Pham, M. T. (2004). The logic of feeling. *Journal of Consumer Psychology*, *14*(4), 360-369.

Romer, C. D., & Romer, D. H. (2000). Federal Reserve information and the behavior of interest rates. *American Economic Review*, *90*(3), 429–457.

Schorfheide, F., & Song, D. (2021). *Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic* (Tech. Rep.). Cambridge, MA: National Bureau of Economic Research.

Schwarz, N. (2002). Situated cognition and the wisdom in feelings: Cognitive tuning. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 144–166). The Guilford Press, London.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*(3-4), 296–303.

Shapiro, L. (2011). *Embodied cognition.* New York, NY, US: Routledge/Taylor & Francis Group.

Sharpe, S. A., Sinha, N. R., & Hollrah, C. A. (2023). The power of narrative sentiment in economic forecasts. *International Journal of Forecasting*, *39*(3), 1097-1121.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*, *31*(4), 329-342.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science*, *15*(6), 322-325.

Stark, T. (2013). *SPF panelists' forecasting methods: A note om the aggregate reslts of a november 2009 special survey* (Tech. Rep.). Federal Reserve Bank of Philadelphia.

Stone, P. J., Dunphry, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis.* Cambridge, MA: MIT Press.

Swanson, N. R. (1996). Forecasting using first-available versus fully revised economic time-series data. *Studies in Nonlinear Dynamics and Econometrics*, *1*, 47–64.

Swanson, N. R., Ghysels, E., & Callan, M. (1999). A multivariate time series analysis of the data revision process for industrial production and the composite leading indicator. In R. F. Engle & H. White (Eds.), *Cointegration, Causality, and Forecasting: Festschrift in honour of Clive W.J. Granger* (pp. 47 –75). Oxford University Press.

Swanson, N. R., & Van Dijk, D. (2006). Are statistical reporting agencies getting it right? Data rationality and business cycle asymmetry. *Journal of Business & Economic Statistics*, *24*(1), 24–42.

Tuckett, D., Smith, R., & Nyman, R. (2014, 07). Tracking phantastic objects: A computer algorithmic investigation of narrative evolution in unstructured data sources. *Social Networks*, *38*, 121–133.

Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, *9*(4), 315-345.

Turton, J., Kabiri, A., Tuckett, D., Smith, R. E., & Vinson, D. P. (2021). *Differentiating approach and avoidance from traditional notions of sentiment in economic contexts.* (Working Paper)

Turton, J., Vinson, D., & Smith, R. (2020, May). Extrapolating binder style word embeddings to new words. In *Proceedings of the Second Workshop on Linguistic and Neurocognitive Resources* (pp. 1–8). Marseille, France: European Language Resources Association.

Zellner, M., Abbas, A. E., Budescu, D. V., & Galstyan, A. (2021). A survey of human judgement and quantitative forecasting methods. *Royal Society Open Science*, *8*(2), 201187.

Table 1: Monthly Real-Time Variables Used in Econometric Model-Based Forecasting Experiments*

Target Variable	Abbrev.	Description	Data Transformation
Industrial Production	IP_t	Monthly vintages and monthly observations, levels, seasonally adjusted. Base period varies across vintages.	$ip_t = log(\frac{IP_t}{IP_{t-1}})$
CPI Inflation	CPI_t	Monthly vintages and monthly observations, levels, seasonally adjusted. Base period varies across vintages.	$cpi_t = log(\frac{CPI_t}{CPI_{t-1}})$
3-month Treasury Bill Rate	$IR3M_t$	Monthly observations. FRED-MD Treasury data, not subject to revision.	no transformation
10-Year Treasury Bond Rate	$IR10Y_t$	Monthly observations. FRED-MD Treasury data, not subject to revision.	no transformation
Unemployment Rate	UR_t	Monthly vintages and monthly observations, levels, seasonally adjusted. Base period varies across vintages.	no transformation
Sentiment	S_t	Monthly frequency textual sentiment index.	no transformation

*Notes: This table lists the forecast variables employed in our real-time prediction experiments, along with the data transformation used for each variable. All variables are obtained from the monthly real-time FRED-MD dataset, with sample period from 1996:03 - 2019:12. See Section 3 for complete details.

Table 2: Quarterly Real-Time Variables Used in Forecast Efficiency Experiments*

Target Variable	Abbrev.	Description	Data Transformation
Industrial Production	IP_t^q	Quarterly vintages and monthly observations, levels, seasonally adjusted. Base period varies across vintages.	$ip_t^q = log(\frac{IP_t}{IP_{t-3}})$
CPI Inflation	CPI_t^q	Quarterly vintages and monthly observations, levels, seasonally adjusted. Base period varies across vintages.	$cpi_t^q = log(\frac{CPI_t}{CPI_{t-3}})$
3-Month Treasury Bill Rate	$IR3Mt^q$	Third monthly observation in each quarter. FRED-MD Treasury data, not subject to revision.	no transformation
10-Year Treasury Bond Rate	$IR10Y_t^q$	Third monthly observation in each quarter. FRED-MD Treasury data, not subject to revision.	no transformation
Unemployment Rate	UR_t^q	Quarterly vintages and monthly observations, levels, seasonally adjusted. Third monthly observation in each quarter. Base period varies across vintages.	no transformation
Sentiment	S_t^q	Monthly frequency textual sentiment index. Third monthly observation in each quarter.	no transformation

*Notes: This table lists the forecast variables employed in our forecast efficiency analysis, along with the data transformation used for each variable. All data, other than the interest rate data, which are collected from our FRED-MD database, are obtained from the Survey of Professional Forecasters website at the Federal Reserve Bank Philadelphia. For these quarterly frequency variables, the SPF includes both professional forecasts and actual real-time data. The transformations listed are conveyed to survey participants, so that the variables they are forecasting are the same as the actual variables reported on. See Section 3 for complete details.

Table 3: Forecasting Models Used Econometric Model-Based Forecasting Experiments*

Model	Description
Model 1: AR(SIC)+Macro+Sentiment	AR(SIC) model augmented with lagged sentiment and macroeconomic variables.
Model 2: AR(SIC)+Macro	AR(SIC) model augmented with lagged macroeconomic variables.
Model 3: AR(SIC)+Sentiment	AR(SIC) model augmented with lagged sentiment variables.
Model 4: AR(SIC)	Autoregressive model with lags selected using the SIC.
Model 5: AR(SIC)+Macro+Sentiment Switching Regression	Switching regression variant of Model 1, with dates detailed in Section 3.
Model 6: AR(SIC)+Macro Switching Regression	Switching regression variant of Model 2, with dates detailed in Section 3.
Model 7: AR(SIC)+ Sentiment Switching Regression	Switching regression variant of Model 3, with dates detailed in Section 3.
Model 8: AR(SIC) Switching Regression	Switching regression variant of Model 4, with dates detailed in Section 3.

*Notes: This table includes brief descriptions of the 8 different forecasting models used in monthly predictions of ip_t, cpi_t,$IR3M_t$, $IR10Y_t$, UR_t. For complete details refer to Section 4.2.

Table 4: Summary of Econometric Model-Based Forecasting Experiment Results*

h=1	Sentiment Included	ip_t	cpi_t	$IR3M_t$	$IR10Y_t$	UR_t
Model 1	**YES**	**0.9334**	1.5518***	1.0077	0.9803	0.9926*
Model 2	NO	1.0201	1.4649***	1.0089	1.0068	0.9888***
Model 3	**YES**	1.0066	**1.0792*** **	0.9998	**0.9764*** **	1.0042**
Model 5	**YES**	3.8685***	2.9191***	**0.6464*** **	2.6767***	0.5930***
Model 6	NO	3.8789***	2.8639***	0.6471***	2.6849***	0.5916***
Model 7	**YES**	3.5877***	2.8396***	0.6483***	2.9212***	**0.5898*** **
Model 8	NO	2.6141***	2.8031***	0.6498***	2.9473***	0.5910***
h=3						
Model 1	**YES**	1.2652	3.2084***	1.0298	**0.9495* **	0.9607***
Model 2	NO	1.4835**	3.0652***	1.0319	1.0213	0.9417***
Model 3	**YES**	1.0249	1.2765***	1.0051	0.9605***	1.0167***
Model 5	**YES**	0.8521***	3.3121***	0.5725***	1.6254***	**0.4942*** **
Model 6	NO	**0.8105*** **	1.8408***	0.5770***	1.6334***	0.4976***
Model 7	**YES**	1.3325***	**1.2044*** **	**0.5682*** **	1.8116***	0.5137***
Model 8	NO	1.7145***	1.3407***	0.5732	2.3740***	0.5263***
h=6						
Model 1	**YES**	1.4947**	4.9601***	1.1741*	1.0299	0.8881***
Model 2	NO	1.5953**	4.8639***	1.1484*	1.1028**	0.8598***
Model 3	**YES**	1.0826*	**1.3445*** **	1.0363	**0.9542*** **	1.0356***
Model 5	**YES**	2.0823***	2.3378***	0.5130***	1.3330***	0.4636***
Model 6	NO	2.1378***	2.3818***	0.5171***	1.3297***	0.4636***
Model 7	**YES**	1.3925***	1.8116***	**0.4988*** **	1.4516***	0.4567***
Model 8	NO	**1.0167*** **	1.7329***	0.5054***	1.4428***	**0.4565*** **
h=12						
Model 1	**YES**	4.4940***	1.7014***	2.4023***	1.0184	0.7904***
Model 2	NO	4.7299***	1.5142***	2.3123***	1.1564**	0.7371***
Model 3	**YES**	1.1137**	1.1094**	1.0587	0.8955***	1.1154***
Model 5	**YES**	0.8009***	2.5713***	0.4508***	**0.8027** **	0.4992***
Model 6	NO	0.7641***	2.6797***	**0.4394*** **	0.8127**	0.4925***
Model 7	**YES**	**0.6661*** **	2.8378***	0.4829***	0.9343	0.4891***
Model 8	NO	0.6691***	**0.9399*** **	0.5015***	1.0066	**0.4807*** **

*Notes: Numerical entries are MSFEs, relative to the benchmark model (Model 4). Minimum relative MSFEs for each variable and horizon are highlighted. Switching Regression Results are based on models with switching periods reflecting NBER business cycle dates. Superscripts *,**,or *** denote rejection of the DM predictive accuracy null hypothesis,in favor of the alternative that the listed model has significantly lower MSFE than the AR(SIC) benchmark, at a 10 %,5%,or 1% significance level,respectively, for MSFEs that are less than one. Starred entries associated with MSFEs greater than unity indicate that the AR(SIC) benchmark "wins".

Table 5: Summary of Efficiency Test Results*

									Efficiency Test Results			
										Rejection due to		
	α	β	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	Overall	*Bias*	S_t	Y_t
					Industrial Production							
h=1	5.2856	-0.0714	-0.4947	0.2164	1.1468	-1.8388	-0.5001	0.1982	NO	NO	NO	NO
h=2	-1.9406	-0.0714	-0.1161	0.9113	-1.0004	1.0638	-0.5964	0.0950	NO	NO	NO	NO
h=3	-0.6144	0.1103	0.0325	-0.7410	0.0705	-0.2977	-0.1134	0.0971	NO	NO	NO	NO
h=4	-0.1876	-0.0256	0.1692	-0.2458	-0.2279	-0.3481	-0.2526	0.1133	NO	NO	NO	NO
					CPI Inflation							
h=1	-0.4665	-0.1249	-0.0041	-1.2598	-0.7165	0.5687	-0.2971	0.3927	**YES**	NO	**YES**	**YES**
h=2	-0.2934	-0.0316	-0.0719	0.2961	-0.3923	0.1185	-0.1997	0.1108	NO	NO	NO	NO
h=3	-2.1140	-0.0084	-0.0417	-0.1010	-0.0038	0.0599	0.0645	0.1001	NO	**YES**	NO	NO
h=4	-2.3780	-0.0233	0.0317	-0.1036	-0.0661	0.1879	0.0366	0.0766	NO	**YES**	NO	NO
					3-Month Treasury Bill Rate							
h=1	0.0775	0.0017	-0.0047	0.0129	-0.0095	-0.0033	-0.0072	0.0598	NO	NO	NO	NO
h=2	-0.1655	0.0038	-0.0082	-0.0513	0.2389	-0.1143	0.0561	0.3314	**YES**	NO	NO	**YES**
h=3	-0.5286	-0.0249	0.0112	-0.0035	0.2098	-0.0554	0.0543	0.4922	**YES**	**YES**	**YES**	**YES**
h=4	-0.7392	-0.0234	0.0274	-0.0678	0.2886	-0.1191	0.0954	0.4982	**YES**	**YES**	**YES**	**YES**
					10-Year Treasury Bond Rate							
h=1	0.0621	0.0001	-0.0038	-0.0126	-0.0059	0.0071	-0.0114	0.0842	NO	NO	NO	NO
h=2	-0.1218	0.0062	0.0247	0.1239	0.0362	0.0361	-0.0048	0.2700	**YES**	NO	NO	**YES**
h=3	-0.2858	-0.0171	0.0173	-0.0440	0.0950	-0.0047	0.0044	0.2547	**YES**	**YES**	**YES**	**YES**
h=4	-0.3376	-0.0019	-0.0009	0.0216	0.1094	-0.0233	0.0057	0.2452	NO	**YES**	NO	**YES**
					Unemployment							
h=1	-0.0252	0.0072	-0.0351	-0.0103	0.0512	0.0227	-0.0051	0.3546	**YES**	NO	NO	**YES**
h=2	-1.7592	-0.0155	0.0936	0.0847	-0.8124	1.1079	-0.1455	0.1404	NO	**YES**	NO	**YES**
h=3	-0.5196	0.1200	0.0024	0.0514	-0.1697	0.1445	0.1083	0.2405	NO	NO	**YES**	NO
h=4	-0.3853	0.0497	0.0051	-0.0741	-0.2572	0.1070	0.0814	0.2196	NO	NO	NO	NO

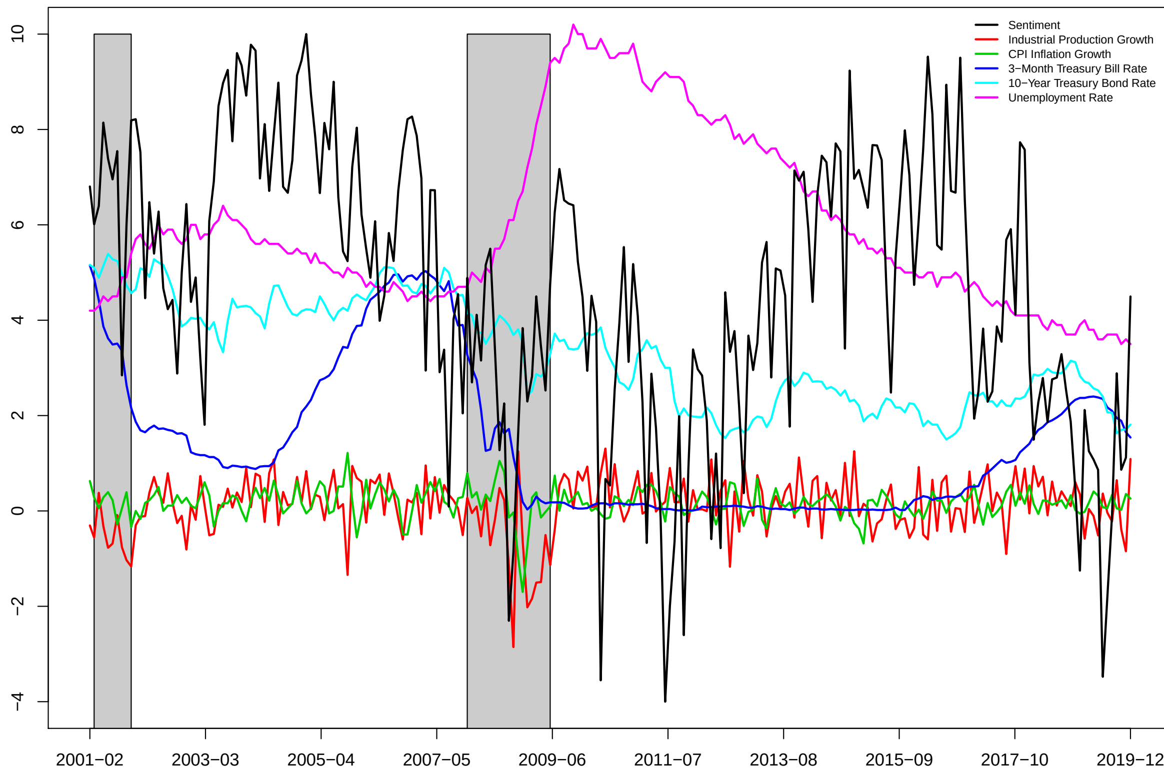
*Notes: Entries under column headers α, β, and γ_i, for $i = 1, \ldots, 5$ contain the coefficients associated with our real-time efficiency test regressions. The γs correspond to the five included macroeconomic variables including Industrial Production, CPI Inflation, the 3-month Treasury Bill Rate, the 10-year Treasury Bond Rate, and the Unemployment Rate, the β coefficient corresponds to our sentiment variable, S_t, and α is the test regression intercept. The rest of the columns (under headers overall, bias, S_t and Y_t) report whether the efficiency hypothesis is rejected (or not) based on examination of t-statistics associated with the individual regressors or the Wald statistic for the overall test. A 'YES' implies failure of efficiency (a p-value less than 0.1) due to missing information contained in the respective variable. All test regressions are carried out for $h = 1, ..., 4$ quarter-ahead (forecast) test regressions. The p-values of the overall Wald test and the individual t-tests can be found in Table 6

Table 6: Individual Efficiency Test Results (p-values)*

	Overall (Wald-Stat)	α	β	γ_1	γ_2	γ_3	γ_4	γ_5
Industrial Production								
h=1	0.1752	0.1088	0.7307	0.1795	0.8336	0.4292	0.2336	0.4302
h=2	0.4692	0.3314	0.5644	0.6010	0.1487	0.2228	0.2602	0.1128
h=3	0.4606	0.7460	0.2912	0.8732	0.1907	0.9181	0.7387	0.7260
h=4	0.3996	0.9084	0.7835	0.3304	0.6036	0.6859	0.6497	0.3539
CPI Inflation								
h=1	0.0136	0.6409	0.0553	0.9707	0.0002	0.1123	0.2311	0.1313
h=2	0.4083	0.6774	0.4730	0.3624	0.1848	0.1782	0.7217	0.1337
h=3	0.4487	0.0000	0.7397	0.4019	0.4613	0.9817	0.7832	0.4152
h=4	0.5486	0.0000	0.4192	0.5529	0.4783	0.7032	0.4274	0.6617
3-Month Treasury Bill Rate								
h=1	0.6301	0.1364	0.6004	0.4185	0.4309	0.6778	0.8905	0.4700
h=2	0.0346	0.3573	0.7366	0.6815	0.3635	0.0020	0.1806	0.0982
h=3	0.0021	0.0066	0.0184	0.5749	0.9495	0.0031	0.5287	0.0926
h=4	0.0019	0.0013	0.0638	0.2361	0.2836	0.0003	0.2454	0.0108
10-Year Treasury Bond Rate								
h=1	0.5148	0.2524	0.9780	0.5277	0.4611	0.8068	0.7814	0.2791
h=2	0.0775	0.4824	0.5634	0.2051	0.0264	0.6094	0.6583	0.8822
h=3	0.0931	0.1234	0.0938	0.3794	0.4176	0.1562	0.9569	0.8882
h=4	0.1042	0.0791	0.8574	0.9649	0.6935	0.0985	0.7925	0.8565
Unemployment								
h=1	0.0247	0.7912	0.2381	0.0020	0.7345	0.2315	0.6134	0.7814
h=2	0.3115	0.1077	0.8169	0.4376	0.8023	0.0707	0.0338	0.4705
h=3	0.1099	0.5107	0.0076	0.9777	0.8256	0.5528	0.6970	0.4221
h=4	0.1391	0.5642	0.1974	0.9423	0.7016	0.2672	0.7328	0.4640

*Notes: See notes to Table 5. This table reports the individual p-values associated with rejection (or not) due to the constant (α), sentiment (β), and each of the 5 variables ($\gamma_1, \ldots, \gamma_5$).

Figure 1: Sentiment and Real-time Target Forecast Variables*



*Notes: This figure plots monthly data for sentiment and five real-time macroeconomic variables for the sample period 2001:2 to 2019:12, corresponding to the full sample period for which our real-time macroeconomic variables are available. See Section 3 for further details.