Media and Business Cycle Predictability[1]

Salim Baz[2] Lara Cathcart[3] Alexander Michaelides[4]

April 21, 2026

[1]An earlier version of this paper was circulated under the title "What is the Value of Financial News?". We thank Pedro Bordalo, Martin Ellison, Alexandros Kostakis, Savitar Sundaresan and seminar participants at the International Association of Applied Econometrics conference, the Stockholm conference on non-traditional data, machine learning and natural language processing in macroeconomics and Imperial College London for useful comments.

[2]Department of Finance, Lebanese American University. E-mail: salim.baz@lau.edu.lb.
[3]Department of Finance, Imperial College London, SW7 2AZ. E-mail: l.cathcart@imperial.ac.uk.
[4]Department of Finance, Imperial College London, SW7 2AZ. E-mail: a.michaelides@imperial.ac.uk.

Abstract

We construct empirical measures of U.S. business-cycle activity based on media mentions of the word "recession" in financial newspapers. The MRIs (media recession indicators) are useful predictors of U.S. economic activity, both in-sample and out-of-sample. Moreover, they compare favorably with existing business-cycle predictors (term premium, purchasing managers' index, consumer sentiment index and real stock market returns). Furthermore, we show that the MRIs are useful predictors of the probability of a U.S. recession six months in advance. Our findings also suggest that constructing and using a more sophisticated sentiment-weighted index does not lead to economically significant improvements in business cycle predictability.

Keywords: Media, business cycles, coincident and leading indicators, recessions, out-of-sample predictability.

JEL Classification: E32, E37, G17.

I. Introduction

One of the main criticisms the economics profession has faced since 2008 has been its inability to forecast the arrival and the depth of the Great Recession. The question was famously raised by Queen Elizabeth during a visit to the London School of Economics in 2008, prompting a response letter signed by many academic economists.[1] The National Bureau of Economic Research (NBER) has historically been the most credible arbiter of U.S. recessions (see Berge and Jorda (2011) for further discussion), but credibility comes at the cost of making announcements typically a few quarters after a recession starts or ends, partly because the aggregate data over many sectors are produced with a lag.[2]

Our contribution is based on a simple idea: we measure the frequency with which the word "recession" appears in financial newspapers.[3] Figure 1, Panel A, presents the raw number of articles mentioning the word "recession" every month and Panel B the same number scaled by the total number of financial news articles in that month. The Figure shows that both the raw and scaled count of the articles containing the word "recession" substantially rise before the NBER recessions. Moreover, the measures are in accordance with intuition in the sense that in the 2007–2009 Great Recession period, both measures rise more strongly than in other recession episodes in our sample. We demonstrate that these measures have forecasting ability over proxies for real economic activity and compare favourably with existing predictors that rely on other commonly used forecasting variables. Finally, we show that these media-based indicators can predict recessions in a non-linear framework, where the goal is to predict the probability of entering a recession a few months after the media indicators rise above a certain value.

We build our argument in a series of steps. We first search Factiva, a database that provides

[1]The visit is described in `https://www.ft.com/content/50007754-ca35-11dd-93e5-000077b07658` and the response letter in `https://www.theguardian.com/uk/2009/jul/26/monarchy-credit-crunch`.

[2]Popular predictors include interest rates (Sims, 1980; Bernanke and Blinder, 1992), the term premium (Estrella and Hardouvelis, 1991) and the default spread (Friedman and Kuttner, 1992), though new evidence suggests that the latter is not a robust predictor of real economic activity (Chatterjee and Bazzana, 2024). More recently, novel predictors have included the TED spread, defined as the spread between the 3-month U.S. LIBOR and the 3-month Treasury-bill rate, as a proxy for funding liquidity; the small-minus-big and high-minus-low Fama and French (1995) factors (Liew and Vassalou, 2000); the Baltic dry index (Bakshi, Panayotov and Skoulakis, 2011), and the S&P500 option-implied risk aversion (Faccini, Konstantinidi, Skiadopoulos and Sarantopoulou-Chiourea, 2019).

[3]Analysis with similar implications for the U.S. has occasionally appeared in *The Economist* (https://www.economist.com/britain/1998/12/10/the-recession-index) but we are not aware of such a measure being systematically analyzed and used to predict economic activity. Doms and Morin (2004) also contains an earlier analysis with similar implications.

access to the archives of thousands of newspapers and other publications (e.g., magazines, television outlets, etc.) and obtain the number of articles mentioning a targeted list of phrases (for instance, whether an article contains the word "recession") for every month in our sample. Since our goal is to study the predictability of the business cycle in the U.S., we focus on the two most prominent specialist newspapers with a significant readership there (the *Financial Times* and *Wall Street Journal*) and exclude generalist newspapers and tabloids (like the *The Washington Post* and *LA Times*, for example).[4] We concentrate on U.S. economic news and the word "recession," because it is often used as a proxy by journalists when discussing economic developments and because using a targeted word can avoid ambiguities in interpretation ("recession" unambiguously refers to economic downturns). We propose two different variants of the measure: our first and simplest measure counts the number of articles mentioning the term "recession" (the raw media recession indicator, or raw MRI), while the second divides this number by the total number of articles discussing economic news in that period (the scaled MRI).

To measure real economic activity (REA), we use five variables: the industrial production index (IP), non-farm payrolls (NFP) and three widely used variables by business economists in the private sector and available from The Conference Board: real personal income less transfer payments (RPILTP), real manufacturing and trade sales (RMTS), and the coincident economic indicator (CEI).[5] We first validate the constructed MRIs by comparing them with known REA predictors. Our measure is constructed in a completely different way and based on quantifying the information available in specialist newspapers. Our findings are therefore complementary to other real-time measures of economic activity, namely, now-casting techniques that use economic data at different frequencies to produce more stable real-time measures of economic activity (see Evans (2005); Giannone, Reichlin and Small (2008); Bańbura, Giannone, Modugno and Reichlin (2013)).

We then test and illustrate the usefulness of the MRIs as predictors of REA growth using both in-sample regressions and out-of-sample analysis. Importantly, we show that the relative increase in the adjusted R-squared compared with more traditional recession indicators. Specif-

[4]The *Wall Street Journal* is the most widely read financial news specialist newspaper in the top 100 newspapers published in the U.S.; the *Financial Times* (FT), though published in the UK, has significant U.S. readership. As indicated in the FT toolkit (found under `https://fttoolkit.co.uk/d/audience/ADGA_Certificate_January_to_December_2015.pdf`), nearly 30% of the global audience is in the United States and the Americas.

[5]All five variables are transformed by computing their growth rates, measured as changes in their natural logarithms.

ically, we control for the term premium (TERM), the Purchasing Managers' Index (PMI), the Consumer Sentiment Index (CSI) and the real S&P 500 returns (S&P500). We empirically find that the MRIs have a substantial additional forecasting ability relative to these predictors, and this conclusion remains robust in out-of-sample predictability comparisons. Our out-of-sample predictability conclusions are robust to the estimation cutoff date, and a rolling-window approach indicates relative stability in the estimated coefficients of the media-based predictors.[6]

An interesting question is the context in which the word "recession" is discussed. We show that in these articles the most frequent words are "economy, growth, FED, risk, inflation, interest rates". This analysis illustrates the importance of the central bank, inflation, and interest rates when discussing the risks facing an economy. Moreover, these words do not feature in the same frequency over time; "inflation" was more pervasive in the 1980s and 1990s than from 2000 onwards, while "interest rate" is strongly discussed in connection to recessions in all decades except in the 2000s.

We caution that the context in which a "recession" is discussed can also be positive for the economy if the phrase is "getting out of a recession". Moreover, recent work proposes using media data to construct economic sentiment as a way to predict the economy (for example, Shapiro, Sudhof and Wilson (2022), Barbaglia, Consoli and Manzan (2023), Ellingsen, Larsen and Thorsrud (2022), Gardner, Scotti and Vega (2022)). To determine whether our simple approach can be improved by quantifying the sentiment in each article, we construct a negative sentiment indicator based on the Harvard IV-4 Dictionary for each article following Soo (2018). An interesting question is then to compare the information in this sentiment-based indicator relative to the raw MRI measure. When we do this comparison, we find that most of the business cycle predictability arises from the raw MRI variable. We also use the sentiment-based indicator to change the weights in our raw MRI variable. Specifically, rather than equally-weighting each article mentioning "recessions", we use the sentiment information in each article to construct a sentiment-weighted MRI variable. We find that the gains from taking sentiment into account in this way yield marginal improvements in predicting the business cycle. Furthermore, the raw MRI can be thought of as the scaled MRI multiplied by the total number of articles (N). We find it useful to disentangle the relative contribution of these two channels and conclude that most

[6]Parameter instability and model uncertainty are common features in the data: although some parameters and/or variables improve forecasting accuracy in some time periods, they may completely fail in other time periods (Stock and Watson, 2003; Ng and Wright, 2013).

of the predictability comes from the scaled MRI.

We next use the MRIs in a non-linear way to predict the likelihood of a recession within a pre-specified time horizon. We first use a methodology that evaluates the ability of the models to signal a recession out-of-sample: we use a probit model to estimate the probability of being in a recession within a number of months given available (current) information. Second, we use the generated signal to forecast the lead time to recessions (in months). We find strong evidence that MRIs predict NBER recessions 6 months ahead of time and they perform well relative to the other variables. Furthermore, we perform additional tests to alleviate the concern of data imbalance (Hand (2009)), due to the fact that recessions occur far less frequently than expansions. We apply the minority oversampling technique (Chawla, Bowyer, Hall and Kegelmeyer (2011)) to show that the results are robust to this criticism.

These findings illustrate the importance of complementing existing factors predicting recessions with measures derived from textual analysis of specialized news outlets. This can be especially advantageous in periods in which the zero lower bound (ZLB) on interest rates might be binding for a long period and the central bank might be focusing on quantitative easing and forward guidance in communicating the future course of monetary policy. This is also consistent with recent research of Fendel, Mai and Mohr (2021) showing that the term premium might not be as good of a recession predictor for the euro area as in previous periods because of the ZLB. Our findings are complementary to other real-time measures of economic activity, namely, nowcasting techniques that use economic data at different frequencies to produce more stable real-time measures of economic activity (see Evans (2005); Giannone et al. (2008); Ban̉bura et al. (2013); Ferrara and Simoni (2023); Algaba, Borms, Boudt and Verbeken (2023)).

Recent literature has also found that different measures of newsflow may be useful coincidental and leading indicators of business cycles. Larsen and Thorsrud (2019) use a Latent Dirichlet Allocation (LDA) approach to show that many topics in a Norwegian financial newspaper have predictive ability for economic fluctuations in Norway, and Eggers, Ellison and Lee (2021) show the importance of media recession announcements in affecting economic activity. Manela and Moreira (2017) also construct a measure of stock market uncertainty based on newspaper articles that can predict stock returns and economic recessions, while Kelly, Manela and Moreira (2021) develop a machine learning text selection model that can be applied to forecasting macroeconomic time series. Baker, Bloom and Davis (2016) provide another prominent example using "target phrases" to measure economic policy uncertainty, while Caldara and Iacoviello (2022) use a similar approach to measure geopolitical risk. Relative to these approaches, our contribution lies in showing how a simple measure that relies on an observable trigger-word like "recession" can predict economic activity.[7]

A more closely related study is by Bybee, Kelly, Manela and Xiu (2024) who use a LDA approach based on Wall Street Journal articles to extract broad themes to predict the business cycle. Our contribution follows a simple and focussed approach that is easy to implement by relying on counts, or sentiment associated with, the word "recession". This method is more transparent and easily replicable, allowing us to track media attention over the business cycle without the complexities of topic modelling.

The rest of this paper is organized as follows. Section II constructs and discusses the media-based measures. Section III discusses the data used to proxy for the growth of U.S. REA, as well as other variables used in the literature as predictors of REA. Section IV presents the main empirical results and linear predictive regressions, both in-sample and out-of-sample. Section V discusses a sentiment-weighted MRI and Section VI explores the predictive ability of the media-based indices in a nonlinear context when forecasting the probability of a recession. Section VII concludes.

II. News-Based Measures of Business Cycles

A. Construction of the News-Based Indices

We start by implementing a simple idea; we collect all the articles containing the word "recession" in two specialized financial newspapers; *The Financial Times* and *The Wall Street Journal*. Using smaller targeted word lists avoids producing weaker results from ambiguities in interpretation (Loughran and Mcdonald (2016)). In our case, the term "recession" unambiguously refers to economic downturns. All else being equal, a sharp increase in the number of articles discussing recession could be linked to general pessimism in the media regarding the economy.

We search the archives of these two newspapers in Factiva and obtain the total number of

[7] In unreported results, we include the economic policy uncertainty measure of Baker et al. (2016) and the geopolitical risk measure of Caldara and Iacoviello (2022) as additional control variables. We find that our out-of-sample forecasting results are unchanged.

articles mentioning the targeted word "recession" for every month in our sample, which covers the period from January 1981 to December 2020.[8] Our interest is in finding predictive signals for U.S. REA and financial variables. A number of problems might arise from following our simple search. The first problem is that the search might—and often does—yield articles that discuss the economic situation in other countries. To minimize the noise caused by articles discussing recessions in countries other than the United States, we apply the region filter provided by Factiva to select only articles discussing the United States. Second, to eliminate articles mentioning "recession" in a non-economic context, we use Factiva's subject filter to select articles classified as "economic news" only.

Finally, we note that the number of articles published every month which discuss the U.S. economy significantly is not constant over time. There is no *a priori* justification for using the raw count of articles rather than a measure that is scaled by the total number of articles in the newspapers. They might both give the same information but it is ultimately an empirical question which measure might contain more information about financial variables. We therefore construct two media-based measures: the first is based on a simple count, while the second scales that count by the total number of articles that discuss the economy in the United States.[9]

We construct the monthly raw and scaled media recession indices (MRIs) as follows:

$$MRIRaw_t = \sum_i \mathbb{I}_{\{Recession=1\}_{i,t}} \times \mathbb{I}_{\{Source=FT+WSJ \,\cap\, Economy \,\cap\, US\}_{i,t}} \tag{1}$$

$$MRIScaled_t = \frac{\sum_i \mathbb{I}_{\{Recession=1\}_{i,t}} \times \mathbb{I}_{\{Source=FT+WSJ \,\cap\, Economy \,\cap\, US\}_{i,t}}}{\sum_i \mathbb{I}_{\{Source=FT+WSJ \,\cap\, Economy \,\cap\, US\}_{i,t}}.} \tag{2}$$

In Equations (1) and (2), $\mathbb{I}_{\{Recession=1\}_{i,t}}$ is an indicator variable equal to 1 if article i in month t mentions the term "recession" at least once and 0 otherwise; $\mathbb{I}_{\{Source=FT+WSJ \,\cap\, Economy \,\cap\, US\}_{i,t}}$

[8] Using Factiva, we collect 85,971 articles discussing the U.S. economy, of which 10,210 specifically mention the term "recession".

[9] We experiment with an alternative normalization, under which the denominator is the total number of articles published in the selected newspapers in month t, regardless of the topic and the subject of the article. We find that the results hold in general for this alternative normalization, but that they are generally statistically weaker than the results obtained in the baseline specification. Under this alternative normalization, a sudden increase (decrease) in the number of noneconomic news and/or news about countries other than the United States, such as a natural disaster in Japan or a terrorist attack in Europe, would cause our index to sharply decrease (increase), even in the absence of changes in the news discussing the U.S. economy.

is an indicator variable equal to 1 if article i in month t discusses the U.S. economy; and "Source=FT+WSJ" refers to the publications being used. The scaled MRI in Equation (2) scales the raw MRI in equation (1) by the number of articles mentioning the word "recession" and are related to U.S. economic news. It is an empirical question which of the two measures will be more suited for detecting weakening economic activity.

B. MRI Descriptive Statistics

Table I presents descriptive statistics for the raw and scaled MRIs constructed using Equations (1) and (2), respectively. The number of articles mentioning the term "recession" increases significantly during NBER-dated recessions, from an average of 18 during expansions to 46 during recessions; the scaled MRI experiences a similar increase in recessions (mean of 0.22) when compared with expansions (mean of 0.10). Both differences-in-means (28 for the raw measure and 0.12 for the scaled one) are statistically significant at the 1% level. Interestingly, even though the volatility of the raw MRI more than doubles during recessions (when compared with expansions), the same cannot be said of its scaled counterpart. Specifically, the raw (scaled) MRI's standard deviation increases from 20 (0.07) during expansions to 48 (0.10) in recessions.

The two indices are also positively correlated. The unconditional correlation is 0.69, and this correlation does not vary significantly between expansions and recessions, i.e. the raw and scaled MRIs do not become more (or less) correlated during recessions when compared with expansions. Importantly, the fact that the correlation is not equal to one indicates that the two measures can provide different signals about the state of the economy. Moreover, the two series are positively serially correlated with the first order autocorrelation ranging between 0.74 and 0.82, respectively, indicating that these variables are not strongly persistent.

Figure 1 plots the time series for the raw and scaled MRI in Panels A and B, respectively, and shows how they behave over the business cycle. Both indices rise ahead of a recession, and they remain elevated over most of the recession periods, relative to the expansion periods. Both have their highest values before the 2008 Great Recession. Nevertheless, differences between the two variables can be detected; the scaled MRI sometimes peaks without being followed by a recession (creating potential false alarms), a behavior that appears to be less frequent for the raw MRI.

Figure 1 clearly shows that the MRIs rise *during* NBER recessions but we are interested in their *predictive* ability over the business cycle, so we next focus on the period around the four

NBER recessions in our sample.[10] Zooming in allows us to better visualize the behaviour of the news-based indices shortly before, during, and after recessions and to observe differences in behaviour *between* recessions. We standardize both MRIs to have a unit standard deviation so that they can be more easily compared. Figure 2 shows that all recessions in the United States have been preceded by an increase in either index in excess of 2 standard deviations and that, importantly, this increase *precedes* the NBER announcement date. Therefore, the announcement of the NBER official date is not driving the spike in these indices. When comparing the raw and scaled MRIs, we note that, except for the 1990-1991 recession, the raw MRI experiences a greater spike. The difference can be quite dramatic: this is especially evident in the case of the Great Recession of 2007-2009, where the 17-standard deviation spike in the raw index dwarfs the 5-standard deviation increase in the scaled MRI.

Figure 2 also plots a widely followed coincident indicator created by The Conference Board (CEI). In three of the four recession episodes in our sample (1990–1991, 2001, and 2007–2009), both the raw and the scaled MRI measures exhibit a noticeable increase prior to any downward movement in the CEI indicator. This pattern suggests that the MRI responds to evolving macroeconomic conditions earlier than CEI rather than simply reflecting information already incorporated in CEI. The only exception is the 2020 recession. The dynamics during this episode are atypical because the COVID-19 shock unfolded extremely rapidly, with an abrupt and unprecedented collapse in economic activity. We note that the MRIs in 2020 also rose but CEI in this episode had a large drop when the pandemic was recognized.[11] Our later results are robust whether the COVID recession is included or not in the analysis.

C. Context and MRI

We also want to better understand the context in which a recession is mentioned in the press. To do so, every time "recession" is mentioned, we extract the fragment of the text 10 words before and after the word "recession". We find 15,508 unique words that appear in these fragments (excluding "recession"), and show the frequency with which each of these words appear using a

[10] The 1981–1982 recession is technically in our sample. However, the first article we obtain is dated October 1981, more than 2 months into the 1981–1982 recession. Accordingly, both indices are zero in the entire period leading up to that recession. For this reason, we omit this recession in the zoomed-in analysis.

[11] The same conclusions arise when using another coincident indicator consructed by the Federal Reserve Bank of Atlanta using now-casting methods (GDPNow). The correlation between CEI and GDPNow is 0.94.

word cloud, where the size of each word is indicative of its frequency.

Figure 3 shows that there are some terms that occur very frequently when recessions are discussed in all four decades in our sample. The most frequent words are "economy, growth, FED, risk, inflation, interest rates, recovery, market". These mentions illustrate the importance of the central bank, inflation, and interest rates when discussing the risks facing an economy. These words (in particular, the Federal Reserve, inflation and interest rates) feature prominently in the historical description of Federal Reserve interest rate tightenings described by Blinder (2023).

Interestingly, these words do not feature in the same frequency over time and therefore there is variation over time in the context in which a "recession" is being discussed. We generate a word cloud by decade in Figure 4, allowing us to observe a change in the narratives in which recessions are discussed in the news media. We find that the words "inflation", "fear" and "risk" were more pervasive in the 1980s than from 1990 onwards.

We conclude that articles mentioning "recession" tend to also discuss economic variables like inflation and interest rates, but the frequency can change over time, presumably reflecting events about the probable cause of a recession. The central bank is prominently discussed in all these articles over all four decades.

III. Other Data

A. Variable Definitions

We consider five monthly economic variables as proxies for real economic activity (REA) in the United States. The first variable is the industrial production index (IP) growth, which captures changes in the level of output in the industrial sector. The second variable is the growth rate of total non-farm payrolls (NFP), representing the total number of U.S. workers in the economy, excluding farm employees, the unincorporated self-employed, and unpaid volunteers. The third variable is the growth rate of real personal income less transfer payments (RPILTP), which reflects income earned from market-based economic activity after excluding government transfer payments. The fourth variable is the growth rate of real manufacturing and trade sales (RMTS), capturing fluctuations in inflation-adjusted sales across the manufacturing, wholesale, and retail sectors. The fifth variable is The Conference Board's coincident economic indicator (CEI), a composite index designed to track current economic conditions by aggregating information from IP, NFP, RPILTP, and RMTS.

The proposed predictors of business-cycle activity are benchmarked against several variables that have been shown to have predictive power for real economic activity (REA) growth. The first variable is the term premium (TERM), defined as the spread between the 10-year Treasury bond yield and the 3-month Treasury bill rate (Estrella and Hardouvelis, 1991). The second is the Purchasing Managers' Index (PMI) obtained from the Institute for Supply Management, which is a survey-based diffusion index commonly used to forecast changes in manufacturing activity. Third, we include the Consumer Sentiment Index (CSI) produced by the University of Michigan, which captures households' expectations and has been widely used as a leading indicator of economic conditions. Finally, we incorporate returns on the S&P 500 index (S&P500), reflecting forward-looking information embedded in equity prices.

Table II summarizes data sources, sample start dates, frequencies, and variable transformations.

B. *Descriptive Statistics*

Table III reports descriptive statistics for all variables; Panel A presents the unconditional statistics, while Panel B presents the statistics for recessions and Panel C for expansions. For all five REA proxies, the mean growth is lower during a recession than during an expansion, and the difference-in-means is economically significant as well as statistically significant at the 1% level (last row in Table). Furthermore, we note that U.S. recessions are characterized with higher uncertainty: the standard deviation for the five REA proxies is significantly larger during recessions than during expansions. The findings confirm that, in our sample, the proxies are very tightly correlated with a recession indicator, which is consistent with the literature.

Table III also reports descriptive statistics for the four predictor variables. All of these variables have pronounced and statistically significant differences between recessions and expansions. TERM rises by around 34 basis points during recessions (from 1.70 to 2.04 percent). PMI, CSI and S&P500 all exhibit a negative (positive) mean during the NBER recessions (expansions); as is the case for the other predictive variables, the difference-in-means is also statistically and economically significant for these variables as well.

Figure 5 compares the raw MRI with TERM (Panel A), PMI (Panel B), CSI (Panel C) and S&P500 (Panel D).[12] We normalize all variables to have a unit standard deviation to make the comparison between them easier. This Figure confirms that the MRIs sharply rise immediately before a recession and seem like good candidates for further exploring their forecasting ability for economic and financial variables. For instance, the raw MRI rises by over 17 standard deviations right at the start of the Great Recession in 2008, whereas all other variables rise much later in that recession and by a much lower magnitude (in terms of standard deviations). A negative term premium typically signals a recession. Panel A shows that this is indeed the case for 3 of the last 4 recessions, but the premium did not turn negative before the 1990s recession, even though the term premium did significantly decrease in the period leading up to that recession. The raw MRI is more volatile than the PMI (Panel B), the CSI (Panel C) and even the stock market (Panel D) and rises substantially before each recession after 2000.

[12]To maintain visual clarity and avoid overcrowding, Figure 5 displays the raw MRI exclusively, omitting its scaled counterpart.

IV. MRI as a Predictor

In line with the recession and long horizon predictability literature, we use both in-sample and out-of-sample predictability regressions to evaluate each predictor's empirical performance.

A. In-Sample Evaluation

We first evaluate the in-sample properties of predictability regressions using our news-based indices. The regressions are variants of

$$Y_{t,t+\tau} = \beta_0 + \beta_1 MRI_t + \beta_2 Y_t + \boldsymbol{\beta_3' X_t} + \epsilon_{t,t+\tau}. \tag{3}$$

In Equation (3), $Y_{t,t+\tau}$ is the variable being predicted; it represents the cumulative log growth from month t to month $t+\tau$, for $\tau \geq 1$; MRI_t represents the MRI in month t; and $\boldsymbol{X_t}$ is a control vector for macroeconomic variables. Specifically, we control for the term premium (TERM), the Purchasing Managers' Index (PMI), the Consumer Sentiment Index (CSI) and the S&P 500 returns (S&P500). We empirically test whether the MRIs have any forecasting ability by estimating model (3) for τ taking values from 1 to 12 months. All variables are standardized to have a unit standard deviation so that they can be more readily interpreted. We report Newey-West p-values (in parentheses), estimated with a number of lags equal to $T^{0.25}$ (where T is the sample size), as well as Hansen and Hodrick (1980) p-values (in brackets).

Our sample includes the short COVID-induced recession in the spring of 2020. This recession was largely unexpected and very sudden and it is therefore hard to expect linear models to capture the sudden drop in economic activity and asset prices. Even though results are stronger without the COVID-induced recession, we perform the empirical analysis using data up to December 2020 that include this highly non-linear event. Moreover, we introduce a non-linear model in a later Section that can be used to predict the likelihood of a recession at some point in the future.

Table IV shows evidence of in-sample predictability at different horizons for both measures of REA using either the raw (Panel A) or scaled (Panel B) MRIs. In Panel A, the raw MRI is statistically significant at all predictive horizons for all five REA growth proxies and for all the horizons considered. The coefficients carry the sign we expect (negative), and are economically significant. For instance, the magnitude of the coefficient implies that a 1-standard-deviation increase in the raw MRI predicts a statistically significant drop of 32.1% of a standard deviation in NFP at the 12-month horizon. Similar quantitative conclusions arise from Panel B, which reports results for the scaled MRI. Overall, the results indicate that both the scaled MRI and the raw MRI exhibit strong economic predictability at all horizons.

We also find it informative to report results graphically (up to 24 months). For each of the five dependent variables (IP, NFP, RPILTP, RMTS and CEI), Figure 6 Panels A to E plot the Hansen and Hodrick (1980) p-values based on equation (3) for the independent variables MRI, TERM, PMI, CSI and S&P500, respectively, and at different forecasting horizons. These graphs show that the raw MRI is in general significant and below the 10% p-value for most horizons. The same cannot be said for any of the other predictors in the model: the Raw MRI does the best job in in-sample predictions.[13]

The predictability improvement can be seen more clearly when making adjusted R-squared comparisons across models. Figure 7 plots the adjusted R-squared for the full model (i.e., model (3)), for the restricted model (i.e., model (3) excluding the raw MRI) and the difference between the two for different forecasting horizons. This is done for each of the five dependent variables IP, NFP, RPILTP, RMTS and CEI in Panels A to E, respectively. The adjusted R-squared substantially increases for the specifications with the raw MRI variable with the increase reaching up to 23% for some horizons. The change in adjusted R-squared illustrates that each news-based variable can be a very useful predictor relative to the interest rate predictors at all horizons up to 2 years.

One might worry that a linear model should not be capturing the rapid deterioration in economic activity witnessed in March 2020. We have repeated the analysis limiting our sample to December 2019. We find that the results are largely unchanged (unreported for brevity). Our explanation is that even though the forecasting model is linear, the MRIs are highly non-linear and move very quickly and in a non-linear way, as Figure 1 shows. We consider this an advantage of the MRIs. Nevertheless, in Section VI we explicitly introduce a non-linear model to capture the potential non-linearities that arise from rapid changes in economic activity and asset prices.

[13]For brevity, we only show the results for the raw MRI; the results for the scaled MRI are similar and available upon request.

B. Out-of-Sample Evaluation

Good in-sample performance does not necessarily guarantee that the news-based indices will perform well out-of-sample. For instance, Stock and Watson (1989) construct a leading index with encouraging in-sample results that failed to predict the recession in the early 1990s. The out-of-sample performance is assessed for the period from January 2000 to December 2020.[14] The first step is therefore to estimate models (4) and (5) set out below, respectively, called the restricted and full models; the only difference between the two models is that the full model (5) includes the variable of interest (i.e., MRI_t), whereas the restricted model (4) does not:

$$Y_{t,t+\tau} = \beta_0 + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \epsilon_{t,t+\tau} \tag{4}$$

$$Y_{t,t+\tau} = \beta_0 + \beta_1 MRI_t + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \epsilon_{t,t+\tau}. \tag{5}$$

The equations are estimated recursively using an expanding window, with the first estimation being performed for the period from January 1986 to December 1999 (this window ensures that the estimation period contains at least 10 years of data and includes at least one recession). Then, for each of these two models, using the parameters obtained in the estimation, we construct a forecast of the dependent variable Y for a horizon τ equal to 1, 3, 6, 9 and 12 months ahead. We denote by $\mathbb{E}_t\left[Y_{t,t+\tau}^{Full}\right]$ and $\mathbb{E}_t\left[Y_{t,t+\tau}^{Restricted}\right]$ the τ-month-ahead forecasts estimated by the full and restricted models, respectively, based on information at time t. Similarly, we denote by $u_{t+\tau}^{Restricted}$ and $u_{t+\tau}^{Full}$ the forecasting errors of the restricted and full models, respectively. The forecasting errors are defined as:

$$u_{t+\tau}^{Restricted} = \mathbb{E}_t\left[Y_{t,t+\tau}^{Restricted}\right] - Y_{t,t+\tau} \tag{6}$$

$$u_{t+\tau}^{Full} = \mathbb{E}_t\left[Y_{t,t+\tau}^{Full}\right] - Y_{t,t+\tau}. \tag{7}$$

We then compute the out-of-sample R-squared as:

$$OOSR_\tau^2 = 1 - \frac{Var(u_{t+\tau}^{Full})}{Var(u_{t+\tau}^{Restricted})}. \tag{8}$$

[14]Since the cutoff date is chosen arbitrarily, we reproduce our out-of-sample results for different cutoff dates. Finally, we show that our results are robust to using a moving window (as opposed to the expanding window approach in the baseline results). Our conclusions remain unchanged and for brevity we report the relevant results in the Online Appendix, Section OA.III.

A positive (negative) out-of-sample R-squared at horizon τ, denoted by $OOSR^2_\tau$, means that adding the news-based index MRI_t improves (worsens) the model by reducing (increasing) the variance of the forecasting errors. We follow Clark and Mccracken (2005) to estimate the statistical significance of the out-of-sample R-squared.[15]

Table V presents the results for the raw and scaled MRIs. Our conclusions are as follows. First, the results echo the in-sample findings: the inclusion of either the raw (Panel A) or scaled (Panel B) MRI improves the accuracy of out-of-sample forecasts. Second, we find that the inclusion of the MRIs improves the out-of-sample forecast accuracy, and that the accuracy gained by augmenting the restricted model with the MRIs is economically significant. The out-of-sample forecasts can improve by at least 2.29% (when predicting for instance RMTS a month ahead) up to 30.66% (when predicting RPILTP 12 months ahead).

We conclude that the MRIs are useful indicators of future economic activity. This conclusion holds both in-sample and out-of-sample and is robust to different ways of computing these statistics.[16]

V. Sentiment and MRI

There are different ways of summarizing the information in the MRI measures. Our implicit assumption is that the word "recession" conveys a negative sentiment that predicts a slowdown in economic activity. However, an article might include sentences like "the economy avoided a recession" or "the economy is unlikely to fall into recession". Either of these would not represent bad news about the economy. Another concern might be that the "sentiment" associated with an article might provide an even better forecast for future economic activity.

To address these issues, we download all the articles published in the Financial Times in our sample period that discuss the U.S. economy, thus obtaining 85,971 articles.[17] To capture article sentiment, we use one of the most common techniques in the literature: the dictionary-based

[15]The Online Appendix (Section OA.I) provides further details on how we compute the statistical significance of the out-of-sample R-squared.

[16]In both in-sample and out-of-sample evaluations, we use the same Raw MRI and Scaled MRI series to test the same hypothesis across different variables and horizons; this could be perceived as an instance of multiple testing. We follow the suggestions of Hastie, Tibshirani and Friedman (2009) and apply the Bonferroni method as an adequate correction for multiple testing. We do not report these results for brevity; the results remain statistically significant.

[17]Article content is needed to analyze how the word "recession" is discussed. We focus on the FT for tractability reasons.

approach. For each article, tone is measured by the raw frequency of negative words in that article. We first select two of the Harvard IV-4 Dictionary lists, namely "decrease" and "fall", expanded with inflections and tenses as in Soo (2018); conservatively, we also delete the word "recession" from that dictionary. The words included in the dictionary are listed in the Online Appendix, Section OA.II. For each article, we transform all words to lowercase and remove all figures, punctuation, and stop words, except for negation terms (i.e., no, not, never, none, neither, never, and nobody). We then define sentiment as the number of words in the article that appear in the expanded negative dictionary, divided by the total number of words. We consider all negative words that are preceded within five words by a negation to be positive in meaning. Thus, they are not counted when computing the sentiment for each article. Next, we run the following univariate linear OLS regression:

$$Sentiment_i = \alpha + \beta \times Recession_i + \epsilon_i, \tag{9}$$

where $Sentiment_i$ is the negative sentiment as described above and $Recession_i$ is an indicator variable equal to 1 if article i mentions the word "recession" at least once and 0 otherwise. The results of estimating this model are set out in Table VI: α and β are estimated at 0.0190 and 0.0107, respectively, and are both statistically significant at the 1% level. An article mentioning the term "recession" is associated with a more negative sentiment. To put things in perspective, the mean sentiment is 0.0203 with a standard deviation of 0.0182. This means that the negative sentiment of articles that mention the term "recession" is 59% of a standard deviation higher than their counterparts which do not mention this term; this impact is economically and statistically significant. Therefore, the overall sentiment of articles that contain the word "recession" is negative.

Next, we compare the sentiment for articles that mention the term "recession" (solid line) and the those that do not (dotted line) in Figure 8. Clearly, we can see that articles that mention "recession" have a more negative sentiment than those that do not mention this term, especially during periods approaching recessions (1991, 2001, 2007-2008, 2020).

In order to disentangle the relative contribution of raw MRI and sentiment, we compare the out-of-sample R-squared based on the following models that differ from each other through the treatment of the sentiment variable. Specifically, we estimate:

$$Y_{t,t+\tau} = \beta_0 + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \beta_4 MRIRaw_t + \epsilon_{t,t+\tau} \tag{10}$$

$$Y_{t,t+\tau} = \beta_0 + \beta_1 Sentiment_t + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \beta_4 MRIRaw_t + \epsilon_{t,t+\tau}. \tag{11}$$

The variables Y_t and $\boldsymbol{X_t}$ are defined as in Section IV. The results are set out in Table VII. These results imply that, including the sentiment in a model that already includes the raw MRI has mixed results. The change is not always statistically significant for all horizons and all variables (RMTS for example), while the increase is not substantial.

We also perform the opposite exercise, and estimate the out-of-sample R-squared based on the following model:

$$Y_{t,t+\tau} = \beta_0 + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \beta_4 Sentiment_t + \epsilon_{t,t+\tau} \tag{12}$$

$$Y_{t,t+\tau} = \beta_0 + \beta_1 MRIRaw_t + \beta_2 Y_t + \boldsymbol{\beta_3'} \boldsymbol{X_t} + \beta_4 Sentiment_t + \epsilon_{t,t+\tau}. \tag{13}$$

The results are set out in Table VIII. Even if the model includes sentiment as an explanatory variable, adding the raw MRI improves forecasting ability in all the cases considered, both in terms of economic and statistical significance across all horizons and all variables. It is important to note here that these results are in line with the baseline results reported in the paper (in which sentiment is not included as an independent variable). This result reinforces the idea that the main predictive ability is coming from the frequency with which the word recession is used in the media rather than the tone of the article mentioning "recession".

A different way of using sentiment is to revisit Equations (1) and (2) where each article mentioning "recession" is weighted equally. We propose a different weighting scheme that one could apply to refine the MRI measures. We can extend these measures to reflect the sentiment of an article, because sentiment also can be an important consideration. Specifically, if we can measure the "pessimism" in an article, then ceteris paribus, a pessimistic article should carry a higher weight than a neutral article. We implement this new measure by weighting each article by its overall negative sentiment or tone, and define the sentiment-weighted media-based index

(SWMRI) as follows:

$$SWMRI_{Raw_t} = \sum_i \mathbb{I}_{\{Recession=1\}_{i,t}} \times \mathbb{I}_{\{Source=FT\ \cap\ Economy\ \cap\ US\}_{i,t}} \times Sentiment_{i,t} \tag{14}$$

$$SWMRI_{Scaled_t} = \frac{\sum_i \mathbb{I}_{\{Recession=1\}_{i,t}} \times \mathbb{I}_{\{Source=FT\ \cap\ Economy\ \cap\ US\}_{i,t}} \times Sentiment_{i,t}}{\sum_i \mathbb{I}_{\{Source=FT\ \cap\ Economy\ \cap\ US\}_{i,t}}.} \tag{15}$$

We evaluate the forecasting ability of the sentiment-weighted measures in out-of-sample predictions; results are set out in Table IX. Comparing these results with the baseline results set out in Table V, we find that, in general, the MRIs weighted by sentiment improve out-of-sample performance in most cases, though this improvement is not economically significant. This shows that most of the predictability is due to the frequency of articles mentioning “recession”.

VI. Predicting Recession Likelihood

A. Baseline

We now introduce a non-linear model to capture the potential non-linearities that arise from rapid changes in economic activity. We do so by predicting the likelihood of a recession, rather than the growth rate of REA. Specifically, we introduce a probit model to investigate whether our news variables can forecast the probability of a recession, where the outcome variable is binary (in an expansion or recession state) and defined using the NBER recession cutoff points (see Estrella and Hardouvelis (1991) or Erdogan, Bennett and Ozyildirim (2015)).

The empirical model becomes:

$$Y_{t,t+\tau} = \Phi(\beta_0 + \boldsymbol{\beta_1'} \boldsymbol{X_t}) + \epsilon_{t,t+\tau}, \tag{16}$$

where $Y_{t,t+\tau}$ is a recession indicator equal to 1, if any of the months from t to $t+\tau$ are defined as a recession by the NBER and 0 otherwise, with $\tau \geq 1$; $\boldsymbol{X_t}$ is a vector of predictors, and $\Phi(\cdot)$ is the standard normal cumulative density function. In the baseline specification, we consider a horizon τ equal to 6 months. Furthermore, we follow the previous analysis and consider three specifications for the vector of predictors $\boldsymbol{X_t}$.

- **Model 1**: the vector of predictors consists of TERM, PMI, CSI and S&P500.
- **Model 2**: the vector of predictors consists of TERM, PMI, CSI, S&P500 augmented by the raw MRI.
- **Model 3**: the vector of predictors consists of TERM, PMI, CSI, S&P500 augmented by the scaled MRI.

We perform an out-of-sample evaluation, and use two different approaches to assess the MRIs' forecasting ability, like in Fendel et al. (2021). First, we employ a methodology evaluating the ability of the models to signal recessions, that is, we use the probit model to estimate the probability of being in a recession in a few months given the information we have now. Second, we use the same model in a different way by testing the ability of each model to forecast the lead time to recessions (in months).

Starting with the signalling methodology, we derive a time series for the probability of a recession based on each predictor, computed as follows:

$$Pr(Recession_{t,t+\tau} = 1) = \Phi(\hat{\beta}_0 + \boldsymbol{\hat{\beta}_1'X_t}). \tag{17}$$

In Equation (17), $Recession_{t,t+\tau}$ is an indicator variable equal to 1 if a recession occurs within τ months from date t, whereas $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated coefficients from Equation (16). We assess the ability of each model to predict the likelihood of a recession out-of-sample separately. The out-of-sample analysis is implemented recursively; we first estimate model (16) from the starting date of our sample up to December 1999. We then store the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ and use them to predict the probability of recession for a horizon of τ equal to 6 months using model (17). We then expand the sample size by 1 month and repeat the same procedure.

We follow Kaminsky and Reinhart (1999) and classify each point (signal) in the time series based on four categories across correct and incorrect predictions as follows:

- **True Positive (TP)**: The signal correctly predicts a recession within τ months: $Pr(Recession_{t,t+\tau})$ exceeds 50% at time t and a recession occurs between $t+1$ and $t+\tau$.
- **False Positives (FP)**: The signal falsely predicts a recession within τ months: $Pr(Recession_{t,t+\tau})$ exceeds 50% at time t but a recession does not occur between $t+1$ and $t+\tau$.

- **False Negatives (FN)**: The signal incorrectly fails to predict a recession within τ months: $Pr(Recession_{t,t+\tau})$ is less than 50% at time t but a recession occurs between $t+1$ and $t+\tau$.
- **True Negatives (TN)**: The signal correctly does not predict a recession within τ months: $Pr(Recession_{t,t+\tau})$ is less than 50% at time t and a recession does not occur between $t+1$ and $t+\tau$.

We report the percentage of signals in each category, which is known as the confusion matrix and based on the above classification, we follow Berge and Jorda (2011) to assess the ability of the media-based predictors to provide accurate signals for a future recession in four different ways. Specifically, we compute the percentage of correct signals (PCS) as $(TP + D)/(TP + FP + FN + TN)$, the percentage of recessions detected (PRD) defined as true positives, as $TP/(TP + FN)$ and the percentage of false alarms/positives (PFA) as $FP/(FP + TN)$. All three measures can be used to assess the ability of the media-based measures to provide accurate signals about forthcoming recessions.

A fourth measure combines the individual information in each of these measures to provide a more accurate picture of a model's ability to make correct predictions. Specifically, we construct a receiver operating curve (ROC) that describes the relationship between the PFA on the x-axis and the PRD on the y-axis for different values of the cutoff threshold signaling a recession. A predictive signal based on a random guess falls on the 45-degree line of this curve (PRD = PFA). As such, the ROC points that lie above the 45-degree line are better than a random guess. If this procedure is followed for different signaling thresholds (ranging from 0 to 1), we can compute the area under the ROC (AUROC) as the integral of the resultant ROC. Typically, a signal with an AUROC of less than 0.6 is not considered to be a good predictor, whereas a perfect predictor would have an AUROC of 1. Therefore, for every predictor, we compute and compare the AUROC as a fourth measure for evaluating the predictor.

Table X reports the results for the AUROC as well as the other signaling indicators for a 50% threshold.[18] We first compare models 1 and 2, to evaluate the contribution of the raw MRI. The confusion matrix first reveals that the inclusion of the raw MRI to the baseline model mainly improves the true positives from 0.00% to 9.80%, whilst decreasing the false negatives from 17.55% to 7.76%. The signaling indicators reflect this improvement: while we note that

[18] The confusion matrix and signaling indicators for other values of the threshold (e.g., 40% or 60%) are available upon request.

the inclusion of the raw MRI mildly improves the percentage of correct signals (from 81.22% to 96.53%), the most significant improvement is observed in the percentage of recessions detected, which increases from 0.00% in model 1 to 55.81%. The caveat is that the inclusion of the raw MRI increases the percentage of false alarms. In unreported results, we specifically look at where those false alarms occur. We find that in most cases, those false alarms occur within 6 months after a recession has ended. This is not a source of concern at this point, because the purpose of the news-based indices is to detect the onset of recessions. Also, given that news-based indices are constructed in a frequency-based methodology, false positives are expected: the specialized news media is likely to still discuss the aftermath of a recession shortly after the recession has ended. Comparing the AUROC across different models confirms that the raw MRI provides a substantial improvement in signaling recessions, with the AUROC increasing from 65.35% to 85.33% between models 1 and 2. The comparisons of models 1 and 3 shows the incremental impact of augmenting the baseline model with the scaled MRI. We find that the scaled MRI also improves the baseline model and the improvements are of similar magnitude when using the raw MRI.

We also use a second approach to compare the predictive ability across different signals. Specifically, we investigate how early each signal crosses 50% threshold for the last three U.S. recessions.[19] Table XI reports the results. We clearly see that the inclusion of the MRI improves the performance of the baseline model: the signal based on model 1 does not spike prior to any of the 3 recessions considered, and only spikes as the 2020 recession starts. When the model is augmented with the raw MRI, the signal spikes 3 months prior to the 2001 and 2008 recession while still spiking coincidentally with the 2020 recession. Here again, the scaled MRI also improves the baseline model and the differences with the raw MRI are not economically significant. We conclude that there is value in the MRI signals when interested in predicting recession probabilities.

B. Imbalanced Data

The binary classification we have used has an unequal distribution of the two output classes. In other words, the data is imbalanced as there are fewer occurrences of being in a recession than in an expansion. In such cases, some standard metrics, such as the AUROC, may be misleading

[19]Results for other values of the threshold are available upon request.

(Hand, 2009). We perform three sets of additional tests to complement our baseline work in order to check whether our results are not suffering from data imbalance problems.

First, we construct two new measures: the area under the precision-recall curve (AUPRC) and the balanced accuracy score (BAS). The AUPRC considers the trade-off between the sensitivity (recall) of the classifier (defined as the rate of true positives to the sum of true positives and false negatives) and its precision (defined as the rate of rate of true positives to sum of true positives and false positives). Similarly to the ROC, we plot a precision-recall curve (PRC) which describes the relationship between the sensitivity on the x-axis and the precision on the y-axis. We shall measure AUPRC as a measure of performance: the baseline AUPRC is the fraction of positives in our sample. In our case, the percentage of months with a recession is 11% of the total number of months. As such, a model which is better than a random classifier should have an AUPRC greater than 0.11. Why do we use the AUPRC? Unlike the ROC, the PRC does not use true negatives at all, since they do not feature in the construction of either the recall or the precision. As such, the AUPRC will completely "ignore" the substantial proportion of true negatives in the data and will only "focus" on the small proportion of positives. AUPRC is therefore considered a useful metric in imbalanced data.

The BAS measure deals with the problem that the accuracy measure (the ratio of true predictions to total predictions) can be very misleading in an imbalanced data set. A naive predictor which always predicts "no recession" will have a very high accuracy, but will not be useful in predicting a recession. We therefore construct a metric which does not suffer from this problem: the BAS which measures the average of the true positive rate and the true negative rate. It is defined as:

$$BAS = \frac{TPR + TNR}{2} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{18}$$

Now, the naive classifier that only signals recessions (or only signals expansions) all the time would get a BAS of 0.5; the maximum theoretical BAS is 1 for a perfect classifier.

Results are in Table XII. The inclusion of the raw MRI significantly improves the AUPRC from 29.60% in model 1 to 68.15% in model 2. We also find that the BAS greatly improves when the raw MRI is included as an additional predictor, increasing from 49.26% to 74.44%. The results also show an improvement for the scaled MRI (the AUPRC increases to 48.50% and the BAS to 62.32%).

Second, another criticism for imbalanced data is that the AUROC measure may not be a good classifier of which model is better. If the ROC curves of two different classifiers cross each other, it is not clear which classifier is better (Hand, 2009; Chawla et al., 2011). An easy way to check whether this is the case is to plot the ROCs obtained using models 1, 2 and 3 and we do so in Figure 9. The Figure clearly shows that the ROC obtained from model 2 (with the raw MRI) and the ROC obtained from model 3 (with the scaled MRI) are both above the ROC from model 1 at all points, and do not cross it at all. This demonstrates that the inclusion of either the raw or the scaled MRI is a clear improvement to the baseline model. Because the ROC curves for models 2 and 3 intersect, we conclude that the two MRI variables offer similar predictive ability.

Third, we follow the approach developed by Chawla et al. (2011) to address the issue of imbalanced dataset.[20] Briefly, this approach consists of oversampling the minority class by synthesizing new examples from the existing dataset and is called Synthetic Minority Oversampling Technique (SMOTE). We use the SMOTE algorithm and reproduce the results of the signaling approach; the results are set out in Table XIII and should be read in juxtaposition with their non-SMOTE counterparts in Table X. We first find that all models predict far more recessions with SMOTE than without. This is reflected in an increase in PRD and PFA with SMOTE. The AUROC also slightly improves with SMOTE. Importantly for us, even with the SMOTE algorithm, the results point to the fact that the inclusion of the raw MRI significantly improves the baseline model. Similar findings arise with the use of the scaled MRI.[21]

VII. Conclusion

We propose news-based indices that can be used as leading indicators of real economic activity growth and recessions. The indicators are based on articles in the *Financial Times* and the *Wall Street Journal*, and use different simple ways of weighting articles that include the key word "recession." The constructed indicators outperform predictor variables that are well established in the literature. Our key contribution arises from showing that a simple measure that relies on the number of times the word "recession" shows up in the financial press can be a useful signal

[20] We thank an anonymous referee for the suggestion to do so.

[21] In unreported results, we also find that the AUPRC and BAS also improve with the use of SMOTE.

for future macroeconomic activity. We believe our findings can be useful for central banks and investors interested in business cycles and investment management.

References

Algaba, A., Borms, S., Boudt, K. and Verbeken, B. (2023), 'Daily news sentiment and monthly surveys: A mixed-frequency dynamic factor model for nowcasting consumer confidence', *International Journal of Forecasting* **39**(1), 266–278.

Baker, S., Bloom, N. and Davis, S. (2016), 'Measuring economic policy uncertainty', *Quarterly Journal of Economics* **131**(4), 1593–1636.

Bakshi, G., Panayotov, G. and Skoulakis, G. (2011), 'Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios', *Journal of Financial Economics* **100**(3), 475–495.

Bañbura, M., Giannone, D., Modugno, M. and Reichlin, L. (2013), Now-casting and the real-time data flow (Chapter 4), *in* 'Handbook of Economic Forecasting', Vol. 2, pp. 195–237.

Barbaglia, L., Consoli, S. and Manzan, S. (2023), 'Forecasting with economic news', *Journal of Business & Economic Statistics* **41**(3), 708–719.

Berge, T. J. and Jorda, S. (2011), 'Evaluating the classification of economic activity into recessions and expansions', *American Economic Journal: Macroeconomics* **3**(2), 246–277.

Bernanke, B. S. and Blinder, A. S. (1992), 'The federal funds rate and the channels of monetary transmission', *The American Economic Review* **82**(4), 901–921.

Blinder, A. S. (2023), 'Landings, soft and hard: The Federal Reserve, 1965–2022', *Journal of Economic Perspectives* **37**(1), 101–120.

Bybee, L., Kelly, B., Manela, A. and Xiu, D. (2024), 'Business news and business cycles', *The Journal of Finance* **79**(5), 3105–3147.

Caldara, D. and Iacoviello, M. (2022), 'Measuring geopolitical risk', *American Economic Review* **112**(4), 1194–1225.

Chatterjee, U. K. and Bazzana, F. (2024), 'Do corporate credit spreads predict the real economy?', *International Review of Economics & Finance* **91**, 272–286.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2011), 'Smote: Synthetic minority over-sampling technique', *The Journal of Artificial Intelligence Research* **16**, 321–357.

Clark, T. E. and Mccracken, M. W. (2005), 'Evaluating direct multistep forecasts', *Econometric Reviews* **24**(4), 369–404.

Doms, M. E. and Morin, N. J. (2004), 'Consumer sentiment, the economy, and the news media', *FRB of San Francisco Working Paper* (2004-09).

Eggers, A. C., Ellison, M. and Lee, S. S. (2021), 'The economic impact of recession announcements', *Journal of Monetary Economics* **120**, 40–52.

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

Erdogan, O., Bennett, P. and Ozyildirim, C. (2015), 'Recession prediction using yield curve and stock market liquidity deviation measures', *Review of Finance* **19**(1), 407–422.

Estrella, A. and Hardouvelis, G. A. (1991), 'The term structure as a predictor of real economic activity', *Journal of Finance* **46**(2), 555–576.

Evans, M. (2005), 'Where are we now?: Real time estimates of the macro economy', *International Journal of Central Banking* **1**(2), 127–175.

Faccini, R., Konstantinidi, E., Skiadopoulos, G. and Sarantopoulou-Chiourea, S. (2019), 'A new predictor of U.S. real economic activity: The S&P 500 option implied risk aversion', *Management Science* **65**(10), 4927–4949.

Fama, E. F. and French, K. R. (1995), 'Size and book-to-market factors in earnings and returns', *The Journal of Finance* **50**(1), 131–155.

Fendel, R., Mai, N. and Mohr, O. (2021), 'Recession probabilities for the eurozone at the zero lower bound: Challenges to the term spread and rise of alternatives', *Journal of Forecasting* **40**(6), 1000–1026.

Ferrara, L. and Simoni, A. (2023), 'When are Google data useful to nowcast GDP? An approach via preselection and shrinkage', *Journal of Business & Economic Statistics* **41**(4), 1188–1202.

Friedman, B. M. and Kuttner, K. N. (1992), 'Money, income, prices, and interest rates', *The American Economic Review* **82**(3), 472–492.

Gardner, B., Scotti, C. and Vega, C. (2022), 'Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements', *Journal of Econometrics* **231**(2), 387–409.

Giannone, D., Reichlin, L. and Small, D. (2008), 'Nowcasting: The real-time informational content of macroeconomic data', *Journal of Monetary Economics* **55**(4), 665–676.

Hand, D. J. (2009), 'Measuring classifier performance: a coherent alternative to the area under the ROC curve', *Machine learning* **77**(1), 103–123.

Hansen, L. P. and Hodrick, R. J. (1980), 'Forward exchange rates as optimal predictors of future spot rates: An econometric analysis', *Journal of Political Economy* **88**(5), 829–853.

Hastie, T., Tibshirani, R. and Friedman, J. H. J. H. (2009), *The elements of statistical learning : data mining, inference, and prediction*, Springer series in statistics, 2nd ed. edn, Springer, New York.

Kaminsky, G. L. and Reinhart, C. M. (1999), 'The twin crises: The causes of banking and balance-of-payments problems', *American Economic Review* **89**(3), 473–500.

Kelly, B., Manela, A. and Moreira, A. (2021), 'Text selection', *Journal of Business & Economic Statistics* **39**(4), 859–879.

Larsen, V. and Thorsrud, L. (2019), 'The value of news for economic developments', *Journal of Econometrics* **210**(1), 203–218.

Liew, J. and Vassalou, M. (2000), 'Can book-to-market, size and momentum be risk factors that predict economic growth?', *Journal of Financial Economics* **57**(2), 221–245.

Loughran, T. and Mcdonald, B. (2016), 'Textual analysis in accounting and finance: A survey', *Journal of Accounting Research* **54**(4), 1187–1230.

Manela, A. and Moreira, A. (2017), 'News implied volatility and disaster concerns', *Journal of Financial Economics* **123**(1), 137–162.

Ng, S. and Wright, J. H. (2013), 'Facts and challenges from the great recession for forecasting and macroeconomic modeling', *Journal of Economic Literature* **51**(4), 1120–1154.

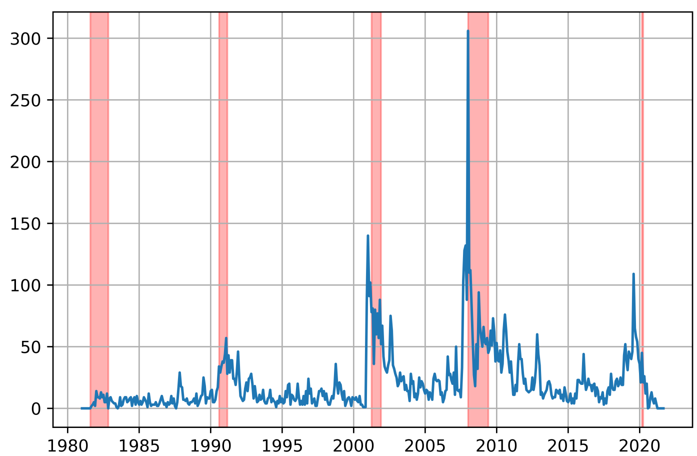
Shapiro, A. H., Sudhof, M. and Wilson, D. J. (2022), 'Measuring news sentiment', *Journal of Econometrics* **228**(2), 221–243.

Sims, C. A. (1980), 'Comparison of interwar and postwar business cycles: Monetarism reconsidered', *The American Economic Review* **70**(2), 250–257.

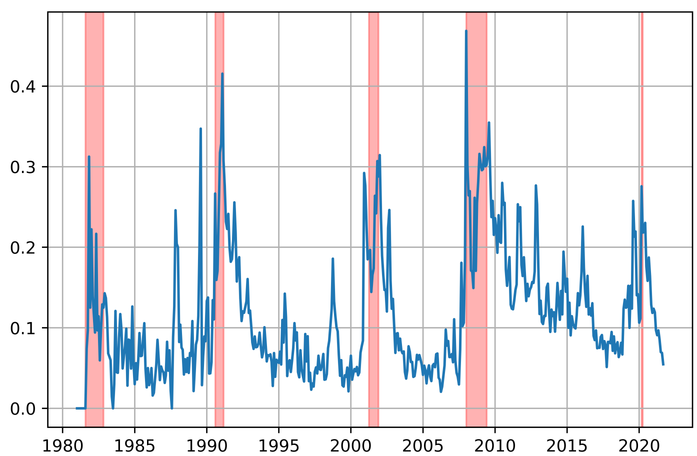
Soo, C. K. (2018), 'Quantifying sentiment with news media across local housing markets', *The Review of Financial Studies* **31**(10), 3689–3719.

Stock, J. H. and Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *NBER Macroeconomics Annual* **4**, 351–394.

Stock, J. H. and Watson, M. W. (2003), 'Forecasting output and inflation: The role of asset prices', *Journal of Economic Literature* **41**(3), 788–829.

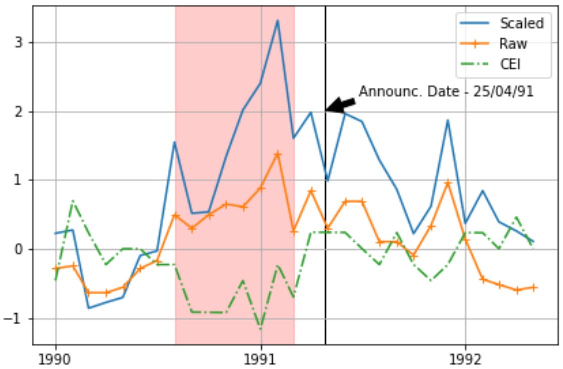


Panel A. Raw

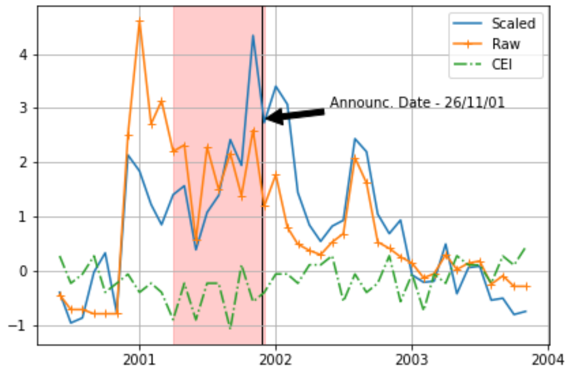


Panel B. Scaled

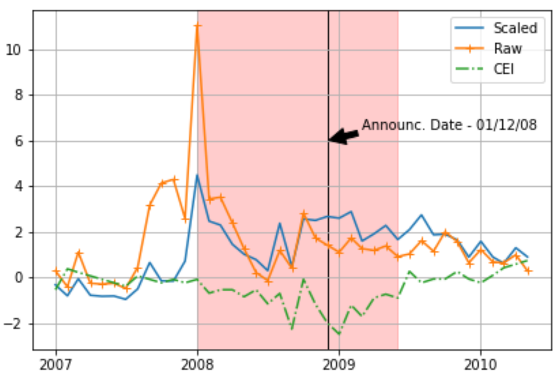
Figure 1. Raw and scaled media recession indices (MRIs). The raw and scaled MRIs are based on the WSJ and FT and are defined as in Equations (1) and (2), respectively. Shaded areas indicate NBER-dated U.S. recessions.



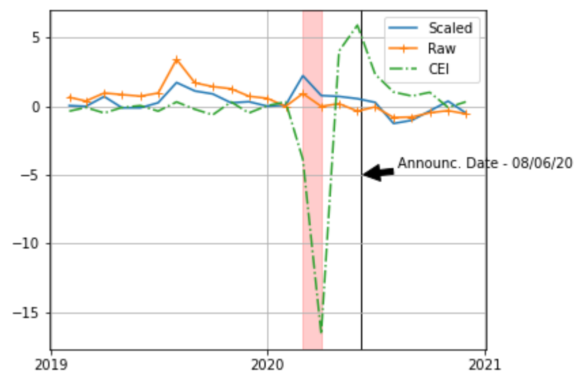
Panel A. 1990–1991 U.S. Recession



Panel B. 2001 U.S. Recession



Panel C. 2007–2009 U.S. Recession



Panel D. 2020 U.S. Recession

Figure 2. Comparison of the raw and scaled MRIs. The raw and scaled MRIs are based on the WSJ and the FT and are defined in Equations (1) and (2), respectively. CEI is a widely followed indicator constructed by The Conference Board. All indices are standardized to have a unit standard deviation to be comparable. Shaded areas indicate NBER-dated U.S. recessions.

Figure 3. Word Cloud around the Term "Recession". This figure presents the word cloud of the words occurring within a distance of 10 words before or after the term "recession" in articles published in FT between 1980 and 2020. Terms are assigned a font size proportional to their frequency in the corpus of documents.

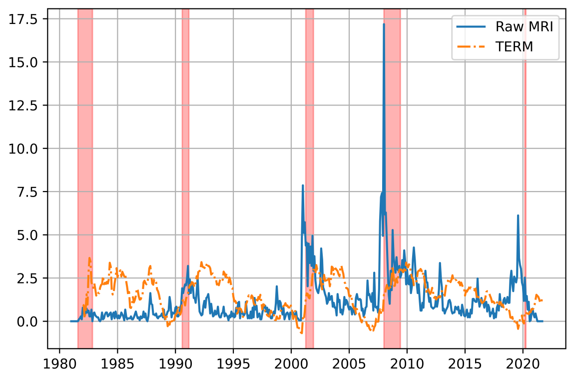
Panel A. 1980 to 1989

Panel B. 1990 to 1999

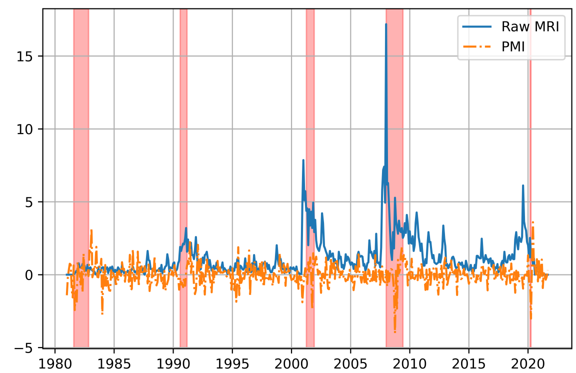
Panel C. 2000 to 2009

Panel D. 2010 to 2020

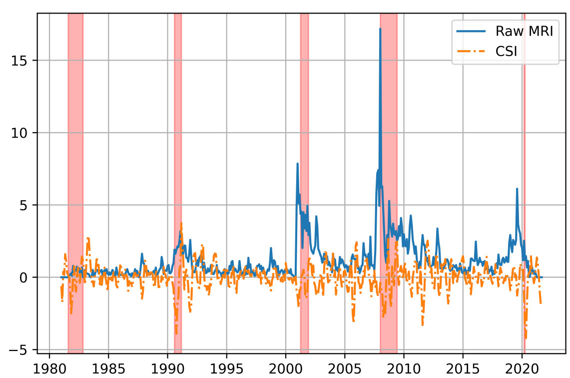
Figure 4. Word Cloud around the Term "Recession" by Decade. This figure presents the word cloud of the words occurring within a distance of 10 words before or after the term "recession" in articles published in FT between 1980 and 2020, presented by decade. Terms are assigned a font size proportional to their frequency in the corpus of documents.



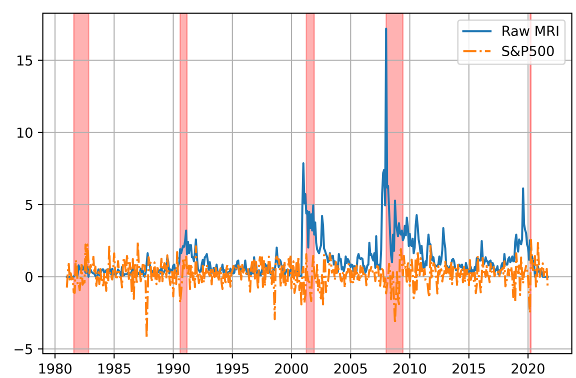
Panel A. TERM versus Raw MRI



Panel B. PMI versus Raw MRI

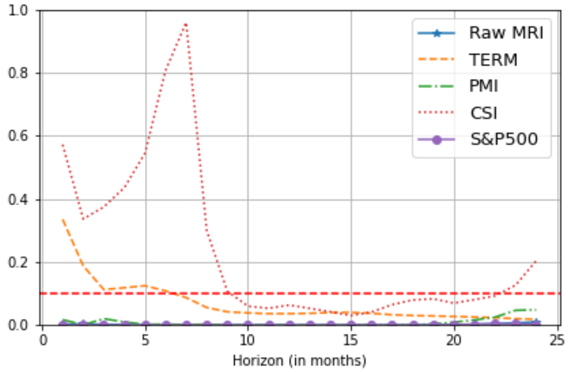


Panel C. CSI versus Raw MRI

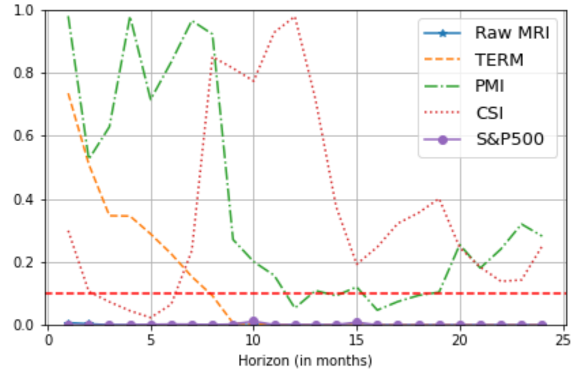


Panel D. S&P500 versus Raw MRI

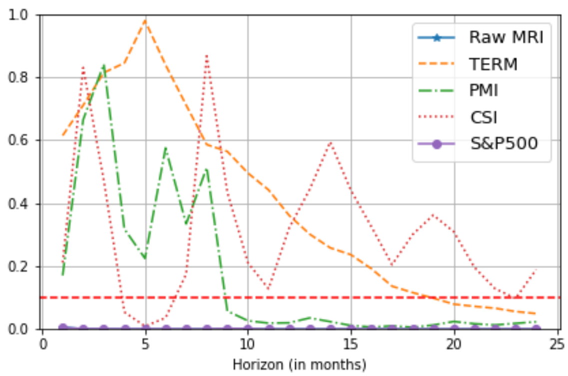
Figure 5. Raw MRI and REA Predictors. The raw MRI is based on the FT and the WSJ and are defined in Equation (1). TERM is the term premium, defined as the difference between the 10-year Treasury bond and the 3-month Treasury-bill rate; PMI is the Purchasing Managers' Index; CSI is the Consumer Sentiment Index; S&P500 is the return on the S&P500 index. All variables are standardized to have a unit standard deviation to be comparable. Shaded areas indicate NBER-dated U.S. recessions.



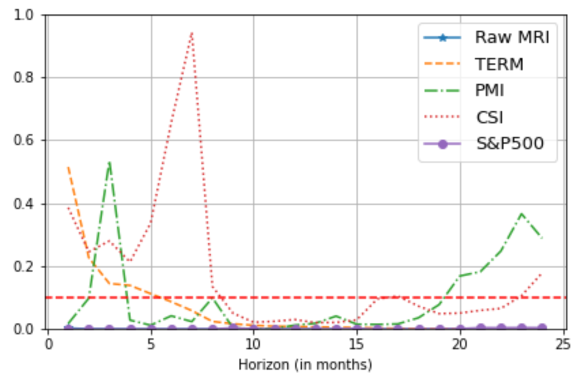
Panel A. IP p-values



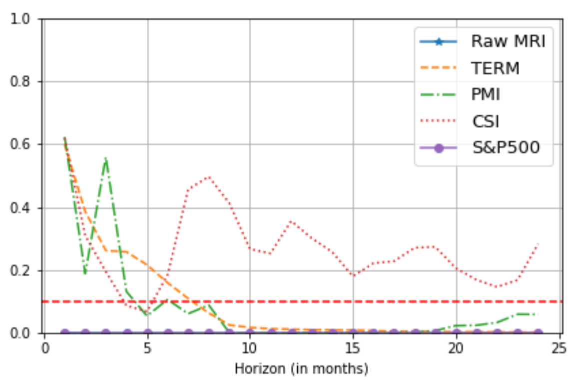
Panel B. NFP p-values



Panel C. RPILTP p-values

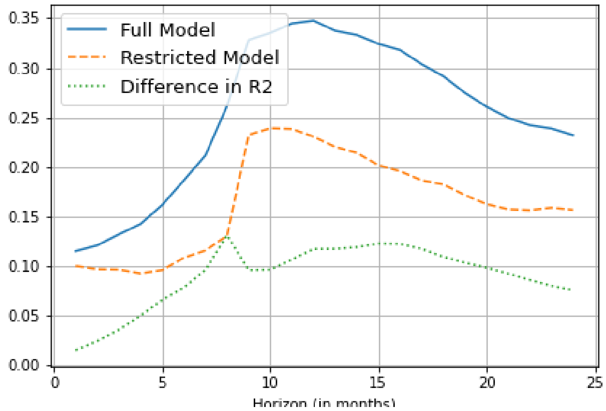


Panel D. RMTS p-values

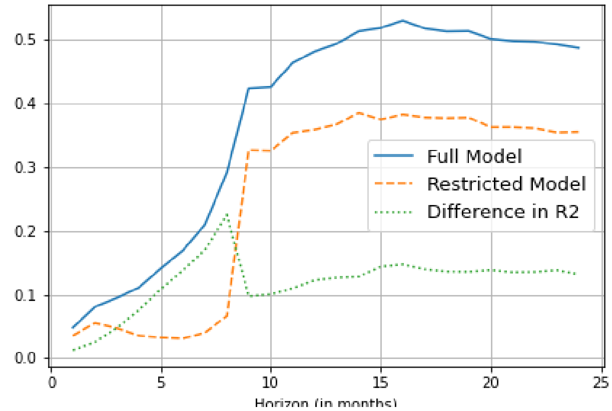


Panel E. CEI p-values

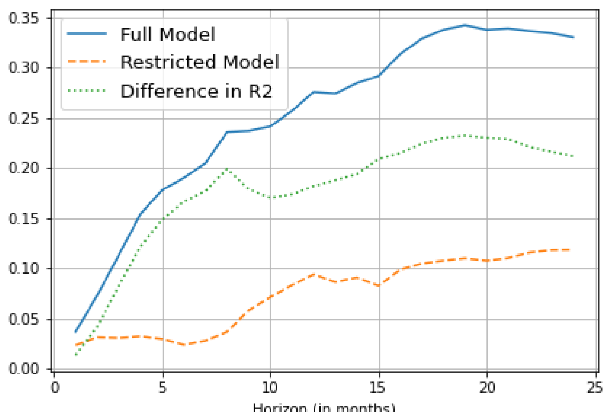
Figure 6. In-sample regression summary (p-values). This figure presents Hansen-Hodrick p-values for different monthly horizons for a model predicting five proxies of REA: IP (Panel A), NFP (Panel B) RPILTP (Panel C), RMTS (Panel D) and CEI (Panel E) estimated based on Equation (3). The horizontal dotted line represents the 10% cutoff p-value.



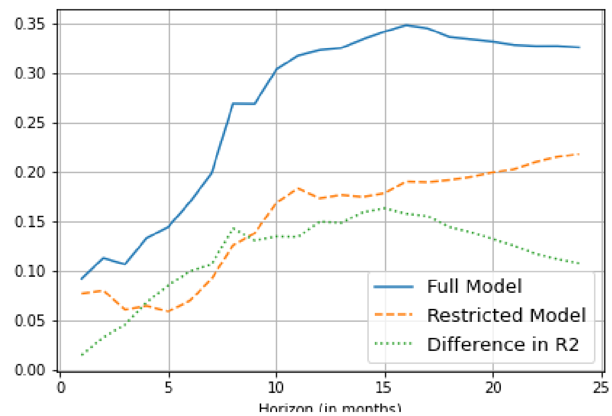
Panel A. IP R-Squared



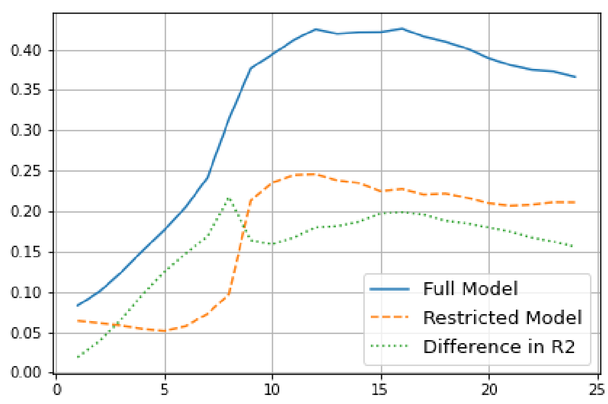
Panel B. NFP R-Squared



Panel C. RPILTP R-Squared



Panel D. RMTS R-Squared



Panel E. CEI R-Squared

Figure 7. In-sample regression summary (R-squared). This figure compares the adjusted R-squared for the restricted model of Equation (3) that excludes the raw MRI with the adjusted R-squared for the full model (that includes the raw MRI) at different horizons. Panel A, B, C, D and E respectively presents the results for IP, NFP, RPILTP, RMTS and CEI as the proxies for REA growth.

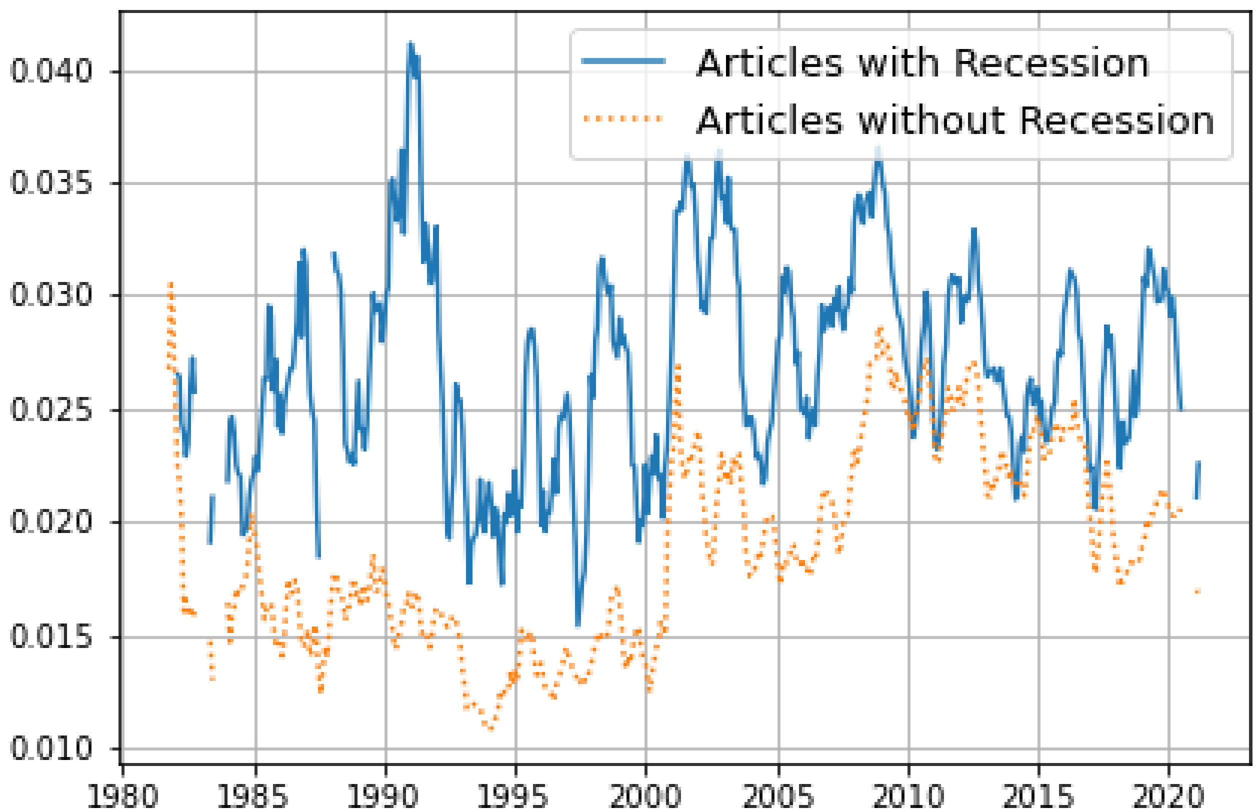


Figure 8. Negative Sentiment of FT Articles Discussing the U.S. Economy. The graphs plots 6-month rolling-mean sentiment of articles published in the FT. The sentiment constructed from articles that mention (do not mention) the term "recession" is plotted in solid (dotted) line.

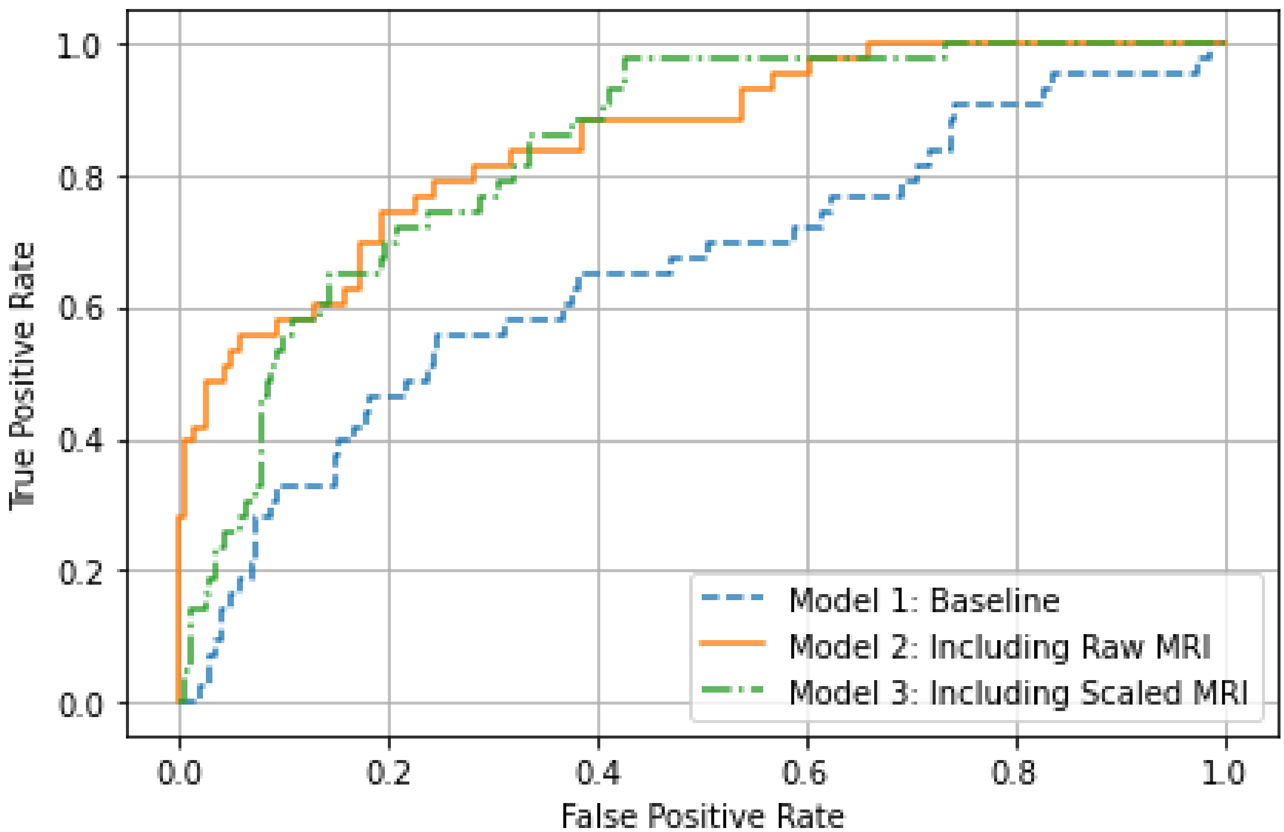


Figure 9. Receiver Operating Curves (ROCs). The graph plots ROCs obtained by using Models 1, 2 and 3 defined in section VI.A

Table I. MRIs: Descriptive Statistics

This table reports descriptive statistics for the raw and scaled MRIs. Descriptive statistics are computed for the full sample, during NBER-dated U.S. recessions, and expansions. The sample spans January 1981 to December 2020.

	Full Sample		Recessions		Expansions		Difference in Means	
	Raw MRI	Scaled MRI	Raw MRI	Scaled MRI	Raw MRI	Scaled MRI	Raw MRI	Scaled MRI
N	480	480	52	52	428	428		
Mean	21	0.12	46	0.22	18	0.10	28***	0.12***
Std. Dev.	26	0.08	48	0.10	20	0.07		
25%-Percentile	7	0.06	12	0.14	6	0.05		
Median	14	0.09	38	0.22	12	0.08		
75%-Percentile	25	0.15	59	0.30	23	0.14		

Table II. Variable Definitions

The table lists the variables and reports the definition, source, starting date of data availability, and any transformation performed on the variable. All variables are at a monthly frequency. FRED refers to the Federal Reserve Economic Data database, maintained by the Federal Reserve Bank of St. Louis' research division; TCB refers to The Conference Board; CRSP refers to the Center for Research in Security Prices.

Variable Name	Definition	Source	Starting Date	Transformation
Panel A: REA Proxies				
IP	Industrial production index	FRED	January 1919	Log difference
NFP	Nonfarm payroll	FRED	January 1939	Log difference
RPILTP	Real personal income less transfer payments	TCB	January 1959	Log difference
RMTS	Real manufacturing and trade sales	TCB	January 1959	Log difference
CEI	Coincident economic indicator	TCB	January 1959	Log difference
Panel B: REA Predictors and Other Variables				
TERM	Term premium defined as the difference between the rate of the 10-year Treasury bond and that of the 3-month Treasury bill (end-of-month)	FRED	January 1982	None
PMI	Purchasing managers' index	Institute for Supply Management (ISM World)	January 1948	Log difference
CSI	Index of consumer sentiment	Survey of Consumers (University of Michigan)	March 1978	Log difference
S&P500	Monthly returns of the S&P500 index	CRSP	January 1926	None
GDPNow	Nowcasting model for gross domestic product (GDP) growth	Atlanta Fed	February 1967	None
Recession	NBER Recession indicator (equal to 1 during U.S. recessions)	FRED	December 1854	None

Table III. Descriptive Statistics

This table reports descriptive statistics for two U.S. real economic activity (REA) monthly growth proxies (IP, NFP, RPILTP, RMTS and CEI) and for four REA predictors (TERM, PMI, CSI, S&P500). Descriptive statistics are computed for the full sample (Panel A), during NBER-dated U.S. recessions (Panel B), and expansions (Panel C). For the difference in means between recessions and expansions p-value statistical significance is: *p <0.1; **p <0.05; ***p <0.01. The sample spans January 1981 to December 2020 for all variables except for TERM, for which the sample is from January 1982 to December 2020.

	REA Proxies					REA Predictors			
Panel A: Full Sample	IP	NFP	RPILTP	RMTS	CEI	TERM	PMI	CSI	S&P500
N	480	480	480	480	480	468	480	480	480
Mean	0.0013	0.0010	0.0022	0.0020	0.0016	1.7337	0.0003	0.0002	1.0131
Standard Deviation	0.0100	0.0069	0.0073	0.0114	0.0070	1.1266	0.0401	0.0267	4.4643
25%-Percentile	-0.0022	0.0004	-0.0001	-0.0031	0.0000	0.8100	-0.0204	-0.0136	-1.6625
Median	0.0021	0.0014	0.0025	0.0020	0.0019	1.7700	0.0000	0.0020	1.4050
75%-Percentile	0.0056	0.0022	0.0047	0.0076	0.0038	2.6225	0.0207	0.0139	3.8900
Panel B: During U.S. Recessions	IP	NFP	RPILTP	RMTS	CEI	TERM	PMI	CSI	S&P500
N	52	52	52	52	52	47	52	52	52
Mean	-0.0104	-0.0050	-0.0037	-0.0080	-0.0061	2.0396	-0.0105	-0.0112	-0.1704
Standard Deviation	0.0216	0.0188	0.0102	0.0198	0.0161	0.8884	0.0611	0.0424	6.8298
25%-Percentile	-0.0096	-0.0032	-0.0044	-0.0131	-0.0060	1.4500	-0.0365	-0.0384	-5.5800
Median	-0.0061	-0.0019	-0.0011	-0.0068	-0.0038	2.0900	0.0026	-0.0105	-0.6900
75%-Percentile	-0.0033	-0.0010	0.0003	0.0013	-0.0012	2.5850	0.0204	0.0160	4.8875
Panel C: During U.S. Expansions	IP	NFP	RPILTP	RMTS	CEI	TERM	PMI	CSI	S&P500
N	428	428	428	428	428	421	428	428	428
Mean	0.0028	0.0017	0.0029	0.0032	0.0025	1.6995	0.0016	0.0016	1.1569
Standard Deviation	0.0062	0.0025	0.0065	0.0092	0.0040	1.1460	0.0367	0.0238	4.0736
25%-Percentile	-0.0009	0.0008	0.0006	-0.0021	0.0000	0.7700	-0.0194	-0.0119	-1.2325
Median	0.0027	0.0015	0.0028	0.0026	0.0021	1.7200	0.0000	0.0021	1.4900
75%-Percentile	0.0059	0.0022	0.0050	0.0079	0.0041	2.6200	0.0208	0.0133	3.7975
Difference in Means Between Recessions and Expansions	-0.0131***	-0.0066***	-0.0065***	-0.0112***	-0.0086***	0.3400**	-0.0121**	-0.0128***	-1.3273**

Table IV. In-Sample Evaluation

This table reports estimation results for the in-sample predictive OLS regression (3). This regression is estimated for five proxies of U.S. real economic activity (REA) growth (IP, NFP, RPILTP, RMTS and CEI). For each dependent variable, we estimate the model using the two raw and scaled MRIs constructed per equations (1) and (2), respectively. The model includes the following control variables under the vector $\boldsymbol{X_t}$: TERM, PMI, CSI and S&P500. Panels A and B report the coefficients, Newey-West p-values (in parentheses), and Hansen and Hodrick (1980) p-values (in brackets) when using the raw and scaled MRI, respectively. Newey-West standard errors were estimated with a number of lags equal to $T^{0.25}$. *p <0.1; **p <0.05; ***p <0.01 (Hansen and Hodrick (1980) p-values). The sample spans from January 1986 to December 2020.

	Panel A: Raw					**Panel B: Scaled**				
Horizon	IP	NFP	RPILTP	RMTS	CEI	IP	NFP	RPILTP	RMTS	CEI
1 Month	-0.1078 ***	-0.0897 ***	-0.0387 ***	-0.0887 ***	-0.0974 ***	-0.1432 ***	-0.1504 ***	-0.0893 ***	-0.0939 ***	-0.1432 ***
	(0.0326)	(0.1134)	(0.3384)	(0.0181)	(0.0559)	(0.0489)	(0.1266)	(0.1275)	(0.1040)	(0.0748)
	[0.0029]	[0.0073]	[0.0062]	[0.0033]	[0.0012]	[0.0000]	[0.0000]	[0.0000]	[0.0011]	[0.0000]
3 Months	-0.1662 ***	-0.1640 ***	-0.1271 ***	-0.1481 ***	-0.1730 ***	-0.1671 ***	-0.1958 ***	-0.1420 ***	-0.1004 ***	-0.1759 ***
	(0.0238)	(0.0537)	(0.0524)	(0.0074)	(0.0197)	(0.0244)	(0.0313)	(0.0646)	(0.1147)	(0.0221)
	[0.0022]	[0.0005]	[0.0000]	[0.0006]	[0.0000]	[0.0017]	[0.0000]	[0.0000]	[0.0086]	[0.0000]
6 Months	-0.2318 ***	-0.2714 ***	-0.1908 ***	-0.2142 ***	-0.2573 ***	-0.1927 ***	-0.2798 ***	-0.1753 ***	-0.1389 ***	-0.2240 ***
	(0.0038)	(0.0188)	(0.0046)	(0.0001)	(0.0016)	(0.0118)	(0.0033)	(0.0371)	(0.0627)	(0.0045)
	[0.0002]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0023]	[0.0000]	[0.0000]	[0.0033]	[0.0000]
9 Months	-0.2677 ***	-0.2956 ***	-0.1876 ***	-0.2514 ***	-0.2889 ***	-0.1945 ***	-0.2667 ***	-0.1831 ***	-0.1708 ***	-0.2335 ***
	(0.0002)	(0.0212)	(0.0034)	(0.0003)	(0.0001)	(0.0145)	(0.0180)	(0.0191)	(0.0281)	(0.0016)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0061]	[0.0000]	[0.0000]	[0.0018]	[0.0000]
12 Months	-0.2917 ***	-0.3210 ***	-0.1924 ***	-0.2755 ***	-0.3038 ***	-0.1884 **	-0.2711 ***	-0.1851 ***	-0.1605 ***	-0.2260 ***
	(0.0000)	(0.0004)	(0.0013)	(0.0045)	(0.0000)	(0.0174)	(0.0021)	(0.0156)	(0.0577)	(0.0012)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0164]	[0.0000]	[0.0000]	[0.0084]	[0.0000]

Table V. Out-of-Sample Evaluation

This table reports out-of-sample R-squared (Equation (8)) that compares the forecasting errors between the restricted model (Equation (4)) and the full model (Equation (5)). The regressions are estimated for five proxies of U.S. real economic activity (REA) growth (IP, NFP, RPILTP, RMTS and CEI). For each dependent variable, we estimate the model using the raw and scaled MRIs (with results presented in Panels A and B, respectively). The model includes the following control variables under the vector $\boldsymbol{X_t}$: TERM, PMI, CSI and S&P500. For all regressions, the estimation period ends on December 1999 and the out-of-sample forecasting period starts on January 2000. Out-of-sample forecasts are produced recursively. The table reports the coefficient out-of-sample R-squared and p-values in brackets. p-values are computed from the nonstandard distribution of Γ_1 (defined in the Online Appendix) obtained from 5,000 Monte Carlo simulations. *p <0.1; **p <0.05; ***p <0.01. The sample spans from January 1986 to December 2020.

	Panel A: Raw					Panel B: Scaled				
Horizon	IP	NFP	RPILTP	RMTS	CEI	IP	NFP	RPILTP	RMTS	CEI
1 month	2.67%***	6.41%***	2.35%***	2.29%***	4.29%***	3.65%***	11.73%***	4.68%***	0.25%**	6.49%***
	(0.0004)	(0.0000)	(0.0002)	(0.0032)	(0.0000)	(0.0004)	(0.0000)	(0.0000)	(0.0368)	(0.0000)
3 months	4.88%***	11.37%***	10.80%***	5.75%***	10.09%***	4.39%***	18.72%***	10.49%***	2.01%***	11.15%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0042)	(0.0000)	(0.0000)	(0.0044)	(0.0000)
6 months	10.18%***	17.19%***	19.44%***	12.70%***	18.66%***	6.31%***	25.09%***	19.16%***	5.49%***	17.87%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
9 months	14.92%***	13.17%***	23.83%***	15.42%***	25.67%***	4.56%***	9.46%***	25.01%***	6.94%***	15.57%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0010)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
12 months	14.41%***	21.51%***	30.66%***	13.35%***	29.87%***	3.98%***	12.52%***	24.55%***	5.12%***	15.16%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table VI. Sentiment and MRI

This table reports estimation results for the regression (9). $Sentiment_i$ is the negative sentiment as described above and $Recession_i$ is an indicator variable equal to 1 if article i mentions the word "recession" at least once and 0 otherwise. Standard errors are robust to heteroskedasticity. $^{***}p<0.01$.

Coefficient	Value
α	0.0190***
β	0.0107***
Adjusted R^2	0.036

Table VII. Out-of-Sample Evaluation - Sentiment (While Controlling for the Raw MRI)

This table reports out-of-sample R-squared (Equation (8)) that compares the forecasting errors between the restricted model (Equation (10)) and the full model (Equation (11)). The regressions are estimated for five proxies of U.S. real economic activity (REA) growth (IP,NFP, RPILTP, RMTS and CEI). The model includes the following control variables under the vector $\boldsymbol{X_t}$: TERM, PMI, CSI and S&P500. For all regressions, the estimation period ends on December 1999 and the out-of-sample forecasting period starts on January 2000. Out-of-sample forecasts are produced recursively. The table reports the coefficient out-of-sample R-squared and p-values in brackets. p-values are computed from the nonstandard distribution of Γ_1 (defined in the Online Appendix) obtained from 5,000 Monte Carlo simulations. *p <0.1; **p <0.05; ***p <0.01. The sample spans from January 1981 to December 2020.

Horizon	IP	NFP	RPILTP	RMTS	CEI
1 month	3.40%***	0.64%	1.55%***	-0.28%	2.64%***
	(0.0014)	(0.1248)	(0.0044)	(0.4512)	(0.0056)
3 months	4.98%***	4.13%***	1.08%***	-0.23%	4.14%***
	(0.0002)	(0.0000)	(0.0000)	(0.1950)	(0.0002)
6 months	5.10%***	6.39%***	3.79%***	-0.42%*	6.62%***
	(0.0000)	(0.0000)	(0.0000)	(0.0648)	(0.0000)
9 months	-0.49%**	0.04%**	6.39%***	-0.77%**	2.11%***
	(0.0128)	(0.0308)	(0.0000)	(0.0464)	(0.0000)
12 months	-1.70%	0.33%***	5.00%***	-2.22%	1.09%***
	(0.1052)	(0.0020)	(0.0000)	(0.8998)	(0.0000)

Table VIII. Out-of-Sample Evaluation - Raw MRI (While Controlling for the Sentiment)

This table reports out-of-sample R-squared (Equation (8)) that compares the forecasting errors between the restricted model (Equation (12)) and the full model (Equation (13)). The regressions are estimated for five proxies of U.S. real economic activity (REA) growth (IP,NFP, RPILTP, RMTS and CEI). The model includes the following control variables under the vector $\boldsymbol{X_t}$: TERM, PMI, CSI and S&P500). For all regressions, the estimation period ends on December 1999 and the out-of-sample forecasting period starts on January 2000. Out-of-sample forecasts are produced recursively. The table reports the coefficient out-of-sample R-squared and p-values in brackets. p-values are computed from the nonstandard distribution of Γ_1 (defined in the Online Appendix) obtained from 5,000 Monte Carlo simulations. *p <0.1; **p <0.05; ***p <0.01. The sample spans from January 1981 to December 2020.

Horizon	IP	NFP	RPILTP	RMTS	CEI
1 month	0.13%*	3.50%***	0.11%*	1.44%***	0.95%**
	(0.0946)	(0.0008)	(0.0938)	(0.0066)	(0.0198)
3 months	0.55%*	5.69%***	5.78%***	2.77%***	3.79%***
	(0.0654)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
6 months	3.12%***	9.17%***	10.22%***	6.98%***	8.46%***
	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
9 months	7.55%***	9.42%***	11.62%***	8.90%***	16.31%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
12 months	5.63%***	16.66%***	17.73%***	5.94%***	19.70%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table IX. Out-of-Sample Evaluation - Sentiment Weighted

This table reports out-of-sample R-squared (Equation (8)) that compares the forecasting errors between the restricted model (Equation (4)) and the full model (Equation (5)). The regressions are estimated for five proxies of U.S. real economic activity (REA) growth (IP,NFP, RPILTP, RMTS and CEI). The model includes the following control variables under the vector $\boldsymbol{X_t}$: TERM, PMI, CSI and S&P500). For all regressions, the estimation period ends on December 1999 and the out-of-sample forecasting period starts on January 2000. Out-of-sample forecasts are produced recursively. The table reports the coefficient out-of-sample R-squared and p-values in brackets. p-values are computed from the nonstandard distribution of Γ_1 (defined in the Online Appendix) obtained from 5,000 Monte Carlo simulations. *p <0.1; **p <0.05; ***p <0.01. The sample spans from January 1986 to December 2020.

	SW-MRI Raw					SW-MRI Scaled				
Horizon	IP	NFP	RPILTP	RMTS	CEI	IP	NFP	RPILTP	RMTS	CEI
1 month	3.25%***	6.93%***	2.08%***	2.26%***	4.62%***	5.44%***	13.27%***	5.34%***	0.88%***	8.23%***
	(0.0000)	(0.0000)	(0.0000)	(0.0008)	(0.0006)	(0.0000)	(0.0000)	(0.0000)	(0.0076)	(0.0000)
3 months	5.71%***	12.16%***	11.15%***	6.05%***	10.85%***	6.44%***	21.19%***	13.11%***	3.32%***	13.81%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
6 months	11.40%***	18.09%***	19.23%***	13.07%***	19.74%***	8.58%***	28.09%***	22.05%***	6.84%***	21.08%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
9 months	15.58%***	12.40%***	23.94%***	16.43%***	25.87%***	5.17%***	9.93%***	28.32%***	7.78%***	16.79%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
12 months	15.04%***	20.10%***	31.53%***	14.43%***	29.85%***	4.15%***	12.99%***	26.97%***	5.69%***	15.80%***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0014)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table X. Assessing Recession Likelihood - Signalling Approach

This table reports the results of assessing the out-of-sample predictive ability of three models over recessions through the signaling approach. Model 1 includes: TERM, PMI, CSI and S&P500. Model 2 (3) augments Model 1 by including the Raw (Scaled) MRI. The table reports the confusion matrix (true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN)), the percentage of correct signals (PCS), the percentage of recessions detected (PRD), the percentage of false alarms (PFA) and the area under the receiver operating curve (AUROC). These are defined in Section VI precisely. For the confusion matrix and for PCS, PRD and PFA, the table reports results using a 50% threshold to determine whether a signal indicates a recession. Results are presented for the predictive ability over a 6-month horizon. The sample spans from January 1986 to December 2020; the estimation period spans from January 1986 to December 1999, and the forecasting period spans from January 2000 to December 2020.

	Confusion Matrix				Signaling Indicators			
	TP	FP	TN	FN	PCS	PRD	PFA	AUROC
Model 1	0.00%	1.22%	81.22%	17.55%	81.22%	0.00%	1.49%	65.35%
Model 2	9.80%	5.71%	76.73%	7.76%	86.53%	55.81%	6.93%	85.33%
Model 3	5.71%	6.53%	75.92%	11.84%	81.63%	32.56%	7.92%	83.74%

Table XI. Assessing Recession Likelihood - Lead Time to Recessions

This table reports the results of assessing the out-of-sample predictive ability of three models. Model 1 includes: TERM, PMI, CSI and S&P500. Model 2 (3) augments Model 1 by including the Raw (Scaled) MRI. The lead time to recessions is defined as the difference between the time the probability of recession signals cross a given threshold and the start of the recession. The table reports the results using a 50% threshold to determine whether a signal indicates a recession. ND denotes that the recession is not detected by the signal, and C denotes that the recession is detected coincidentally, with no lead time. The sample spans from January 1986 to December 2020; the estimation period spans from January 1986 to December 1999, and the forecasting period spans from January 2000 to December 2020.

	2001	2008	2020
M1	ND	ND	ND
M2	3m	3m	C
M3	3m	ND	C

Table XII. Assessing Recession Likelihood - Addressing Imbalanced Data Concerns

This table reports the results of assessing the out-of-sample predictive ability of three models to alleviate concerns regarding the imbalanced nature of the recession data. Model 1 includes: TERM, PMI, CSI and S&P500. Model 2 (3) augments Model 1 by including the Raw (Scaled) MRI. The table reports the the area under the precision recall curve (AUPRC) and the balanced accuracy score (BAS). These are defined in Section VI precisely. For the BAS, the table reports results using a 50% threshold to determine whether a signal indicates a recession. Results are presented for the predictive ability over a 6-month horizon. The sample spans from January 1986 to December 2020; the estimation period spans from January 1986 to December 1999, and the forecasting period spans from January 2000 to December 2020.

	AUPRC	BAS
Model 1	29.60%	49.26%
Model 2	68.15%	74.44%
Model 3	48.50%	62.32%

Table XIII. Assessing Recession Likelihood - Signalling Approach (with SMOTE Algorithm)

This table reports the results of assessing the out-of-sample predictive ability of three models over recessions through the signaling approach and using the SMOTE algorithm. Model 1 includes: TERM, PMI, CSI and S&P500. Model 2 (3) augments Model 1 by including the Raw (Scaled) MRI. The table reports the percentage of correct signals (PCS), the percentage of recessions detected (PRD), the percentage of false alarms (PFA) and the area under the receiver operating curve (AUROC). These are defined in Section VI precisely. For PCS, PRD and PFA, the table reports results using a 50% threshold to determine whether a signal indicates a recession. Results are presented for the predictive ability over a 6-month horizon. The sample spans from January 1986 to December 2020; the estimation period spans from January 1986 to December 1999, and the forecasting period spans from January 2000 to December 2020.

	Signaling Indicators			
	PCS	PRD	PFA	AUROC
Model 1	61.22%	67.44%	40.10%	70.31%
Model 2	66.53%	90.70%	38.61%	86.81%
Model 3	72.24%	86.05%	30.69%	85.45%