IFC-Bank of Italy Workshop on "Machine learning in central banking"

19-22 October 2021, Rome / virtual event

Using news sentiment for economic forecasting: a Malaysian case study[1]

Eilyn Chong, Chiung Ching Ho, Zhong Fei Ong and Hong H Ong,
Central Bank of Malaysia

[1] This presentation was prepared for the Workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the central banks and other institutions represented at the event.

Using News Sentiment for Economic Forecasting

A Malaysian Case Study

Chiung Ching Ho, Eilyn Chong, Zhong Fei Ong and Hong H. Ong[1]

Abstract

Newspaper text sentiment can be informative about the prevailing macroeconomic conditions at a high frequency level and can be used to improve forecasts of macroeconomic indicators. In this paper, we extract the sentiment from the business and financial section of local newspaper articles in Malaysia using a simple dictionary method, and then evaluate the relationship with existing survey-based sentiment measures and macroeconomic growth outcomes. Specifically, this paper investigates the forecasting power of newspaper sentiment for GDP growth and its demand-side components using linear models, non-linear machine learning models and long-short term memory (LSTM) neural network. Our findings show that the news sentiment could nowcast the survey-based business sentiment measure. Using linear regression and non-linear machine learning models, we also show that the news sentiment has a reliable predictive ability for private investment growth within the two to three-quarter forecast horizon. Nevertheless, we find no significant improvement in using news sentiment to forecast other demand-side components of GDP growth across forecast periods, suggesting that the extracted news sentiment provides limited information content for the broader economy.

Keywords: Newspaper sentiment analysis, text analysis, dictionary, macroeconomic indicator, time-series, machine learning

JEL classification: C53, C82, C45

[1] This paper was prepared for the IFC and Bank of Italy Workshop on "Data Science in Central Banking" held virtually on 19-22 October 2021. All authors are from the Central Bank of Malaysia. Emails: hocc@bnm.gov.my; eilyn@bnm.gov.my; zhongfei@bnm.gov.my; onghh@bnm.gov.my respectively. The views expressed in this paper are those of the authors and should not be interpreted as reflecting the views of the Central Bank of Malaysia or of anyone else associated with the Central Bank of Malaysia. The authors would like to thank Ong Li Ming, Mohd Zhafry Zainol, Chang Wen Huei and Gulraj Singh Sidhu for their invaluable comments on the paper.

1. Introduction

Central banks care about the sentiment of economic agents given their prospective influence on changes in real economic activity. For consumers, their expectations and sentiments about the economic conditions can affect their consumption and saving decisions. Likewise for businesses, their expectations and sentiments about the economic conditions would influence how much they invest and how they set prices and wages. One popular way of measuring sentiment is to directly survey these economic agents about their views on the current and future economic conditions. Prominent examples in Malaysia include the Malaysian Institute of Economic Research (MIER)'s Business Confidence Index and Consumer Sentiment Index. However, surveys can be challenging to conduct frequently, especially if they need to cover a representative sample of the population. Moreover, during periods of macroeconomic stress such as that caused by the COVID-19 pandemic, businesses that temporarily ceased trading may not respond to surveys, thereby affecting the quality of the survey responses.

Two recent advances have offered an alternative approach to measuring sentiment. First is the wide availability of newspapers and annual reports in digital format. These media especially digital newspapers may capture the high-frequency information that consumers and businesses refer to for decision making. The second advancement relates to developments in computational linguistic methods and the rapid growth in computing power. Put together, these enable us to process massive volume of text from newspapers and reports, and extract the sentiment contained therein.

Such techniques have been explored to provide quantitative measures of economic policy uncertainty (Alexopoulos & Cohen, 2015; Baker et al., 2016), daily economic sentiment (Buckman et al., 2020; Thorsrud, 2020), measure of political leanings of media outlets (Gentzkow & Shapiro, 2010) and central bank's objective function (Shapiro & Wilson, 2019).

The derived text-based measures are used to correlate with a variety of economic and financial outcomes. Nyman et al. (2021), for example, suggest text-based measures of excitement could pre-empt an impending financial system distress. Similarly, Manela & Moreira (2017) used The Wall Street Journal articles to construct a news implied volatility (NVIX), which peaks during the financial crises, stock market crashes, times of policy-related uncertainty and world wars. Prominent examples that link sentiment from text with financial market reaction include Calomiris & Mamaysky (2019) and García (2013) who link sentiment from news with stock returns as well as Jegadeesh & Wu (2013) and Loughran & McDonald (2011) who correlated sentiment of text in firms' annual reports with financial market reaction. Others have explored the links between text-based measures of uncertainty and the business cycle (Baker et al., 2016; Bloom, 2014; Moore, 2017). Our paper is closest to the literature that use news sentiment to track and predict a range of macroeconomic variables of interest. Papers such as Aguilar et al. (2021); Fraiberger (2016); Kalamara et al. (2020); Larsen & Thorsrud (2019); Nguyen & Cava (2020); Rambaccussing & Kwiatkowski (2020) show that newspaper text contains information on the future path of certain macroeconomic variables, including gross domestic product (GDP).

To investigate whether their findings would similarly apply to Malaysia, we extract sentiment measures for the Malaysian economy by applying text analytics on local news articles. A study by the Reuters Institute reports that around 86% of

Malaysian users rely on online media as a dominant source of news in recent years (Newman et al., 2020). We explore the feasibility of using online news articles to obtain timely and forward-looking cues about economic conditions that could inform policymaking, especially given the evolving COVID-19 situation, whose impact cannot be immediately observed from the lagging official statistics. Specifically, our news corpus comprises over 720,000 business and financial news articles from 16 major news portals, some of which have digital news archive since year 2001. We develop the sentiment measures based on the net balance of positive and negative words used in news articles based on the pre-defined word lists in dictionaries by Loughran & McDonald (2011) and Correa et al. (2017). We also leverage on the sentiment-scoring model developed by Shapiro et al. (2020) that cater specifically to economic news articles. We then aggregate the word counts or sentiment scores from individual article into monthly time-series indexes. The monthly index is found to comove with the business cycle and key economic events.

Specifically, we find that the news sentiment measures can nowcast movements in the survey-based sentiment indicators that are released on a quarterly basis by MIER. Of the growth variables we attempt to forecast using the news sentiment, we find that the news sentiment measures perform well especially in forecasting private investment growth especially within 2 – 3 quarters ahead. This is true even during periods of macroeconomic stress, highlighting the advantages of using higher frequency news sentiment to inform movements in private investment growth. Similar results are obtained when using non-linear machine learning algorithms. Nevertheless, we find no significant improvement in using news sentiment to forecast other components of economic activity, such as private consumption. This suggests the extracted news sentiment provides limited information content for the broader economy.

The rest of the paper is organised as follows: we first describe our newspaper text data in Section 2. Section 3 then discusses the methods of transforming text into time series and the nowcasting exercises that we perform. In Section 4, we look at the forecast performance of several economic variables with the text-based sentiment using simple linear regression, non-linear machine learning models as well as long short-term (LSTM) neural networks. Section 5 discusses the overall results and concludes.

2. Data

The raw data used in constructing the news sentiment consist of daily newspaper articles taken from our internal subscriptions of 16 Malaysian online newspapers portal either from the official websites or via third party services. Based on a study commissioned by the Reuters Institute in 2020, online news websites remained as one of the predominant sources of news for Malaysians in recent years. Over 720,000 online newspaper articles in the English language were used for our analytical dataset. The selection of newspaper articles was motivated by the availability of digital archives, allowing articles to be extracted from as early as 2001, all the way up to June 2021.

We were particularly interested in investigating the sentiment of newspaper articles in the year 2020, which is representative of the developments of the COVID-19 pandemic and its impact to the economy. As we intend to correlate news sentiment with economic indicators, we consider only news articles from the business- or financial-related sections to increase the signal-to-noise ratio. The 16 news portal used in our dataset, sorted by the average number of articles published per month, is shown in Table 1.

Based on the study by Reuters Institute, The Star and Astro Awani are among the most popular news portals with readership[2] of 30% and 35% respectively. It is important to note, however, that these represent the general readership pattern of Malaysians for all types of news, which is different from the goal in this study which focuses on business- or financial-related news articles. Hence, beyond the popular news portals for general news, we also include the more business-centric news portals such as The Edge Markets and i3investor that are not featured in the Reuters Institute's study.

On top of English news, Malay, Tamil and Chinese are also common languages used in news portals in Malaysia. We decided to analyse only English news articles for two main reasons. The first is the limited availability of well-established algorithms to analyse the sentiment of text in vernacular languages. The second is to avoid double counting news articles that are written in multiple languages in the same news portal.

[2] Based on the Reuters Institute Digital News Report 2020 (Newman et al. 2020), readership is defined as the share of respondents who consumed the media at least once a week.

Descriptive statistics of articles from selected local news portals

Table 1

	Average number of articles per month	Date of first online article	Readership[3]
i3investor	7898	03/03/2020	-
The Star	949	01/01/2003	30%
The Edge Markets	912	16/01/2009	-
Malay Mail	827	18/06/2013	8%
Bernama	654	04/03/2020	-
The Malaysian Reserve	487	10/01/2017	-
Free Malaysia Today	384	31/12/2015	15%
The Borneo Post	284	23/12/2009	-
The Sun Daily	280	15/11/2017	-
SoyaCincau	273	29/01/2020	-
New Straits Times	238	20/05/2014	10%
MPOB Palm News	165	01/03/2020	-
paultan.org	78	25/03/2007	-
Astro Awani	59	01/01/2013	35%
Daily Express	54	15/01/2001	-
MARC	18	20/05/2020	-

Sources: Authors' calculation

3. Methodology

Text pre-processing

Text pre-processing is a common practice of cleaning and preparing text data for subsequent natural language processing tasks. We took the common steps of cleaning the raw newspaper text, including:

- Removal of punctuations, hyperlinks, hypertext markup language (HTML) tags, special characters and extra white spaces;
- Dropping of common stop words using the word list by Nothman et al. (2019) - words that are not by themselves informative and differentiative of sentiment, such as *and*, *is* and *the*;
- Setting all words to lowercases.

Given that we extract sentiment using dictionaries that include the stem words and their inflections (for example, *decline, declining* and *declined*), we do not use stemming or lemmatisation.

3

Sentiment scoring

Sentiment from text is not directly observable and would require text analytical approaches to extract and quantify the sentiment contained in the text. There are two such general approaches. The first is the lexical- or dictionary-based approach that associates predefined lists of words with specific scores indicating how positive or negative it is, without any element of learning. Generally, these dictionaries have ternary classifications of 1, 0, and -1 for positive, neutral, and negative sentiment respectively, but certain lexicons such as Vader have a range of scores. While such word-matching method measures the sentiment of a given corpus of text based on the prevalence of negative vs positive words, it ignores the word's context and compositionality.

To capture the specific contextual characteristics and nuances in human language beyond heuristic rules, machine learning (ML) techniques can be employed to probabilistically predict the sentiment of any given set of text. An ML model is typically trained on a large set of text containing a mapping between textual utterances and sentiment ratings assigned by humans. These have been applied for example on social media data, such as tweets on Twitter combined with user feedback to identify the sentiment of the tweets. While this approach can better capture the nuances in sentiment expression, constructing a large, labelled training dataset is time-consuming and expensive.

In this paper, we adopt the simpler dictionary-based method to measure the sentiment of the news corpuses and construct the news sentiment index. We use the financial stability dictionary by Correa et al. (2017) (hereafter Correa) and the finance-oriented dictionary by Loughran & McDonald (2011) (hereafter LM). Our news sentiment is constructed by counting the number of times that negative and positive words appear in the cleaned text of articles and measuring the net balance of words. When the news contains more positive words and/or fewer negative words, it indicates better sentiment in the economy.

In addition to the lexicons above, we also leveraged on the lexicon created by Shapiro et al. (2020) (hereafter SSW), who scored a corpus of U.S. economic news articles with Vader[4]. This consists of 20,000 words labelled from -4 to +4, corresponding to most negative to most positive. While specific to the context of the U.S., we adopt this alternative lexicon as an attempt to incorporate words that are more specific to the economics rather than financial domain, and that have a wider scoring scale that differentiates between weaker and heightened sentiments. For example, the word *declined* is assigned a score of -0.12, *dropped* -0.16, *downturns* -1.22 and *sluggish* -1.97.

We also took an additional step to swap the sentiment for the words *positiv*e and *negative* that are in the dictionaries but may be associated with COVID-19 in the more recent newspaper text as it will contribute falsely to the economic sentiment that we intend to construct.

To construct an index of the news sentiment, the articles are sorted by the date of publication. In the case of Correa and LM, we compute the sentiment score for each news article i by subtracting the count of negative words from the count of positive words and then dividing by total word count. In the case of SSW, we take a

[4] An open-source Vader python package developed by Hutto & Gilbert (2014)

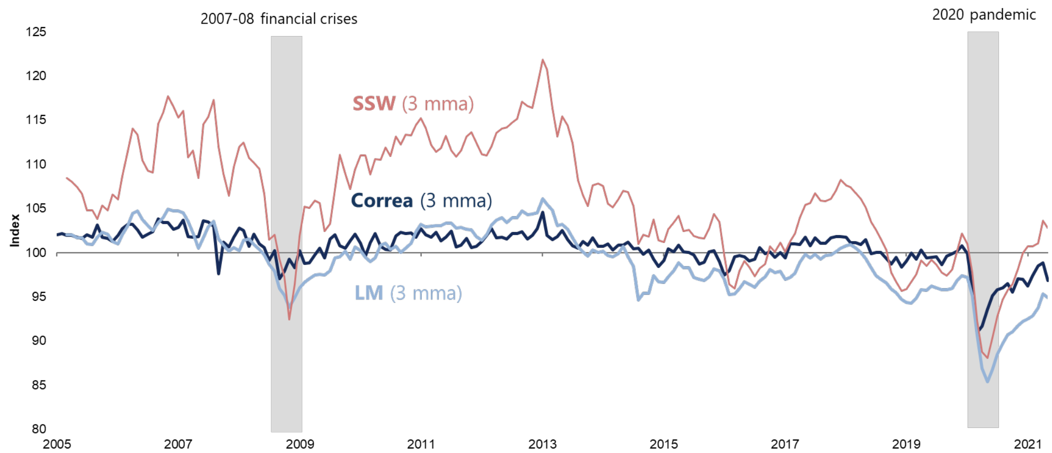
sum of the sentiment scores associated with the words w in each news article i and then dividing by total word count. We then express these sentiment scores as net count per 1000 words and convert them into an index. An index above 100 indicates better sentiment, and index below 100 indicates otherwise.

Correa and LM: $$sentiment\ index_i = 100 + \frac{\sum Positive_i - \sum Negative_i}{Word\ count_i} \times 1000$$

SSW: $$sentiment\ index_i = 100 + \sum_w \frac{Score_{w,i}}{Word\ count_i} \times 1000$$

Given that the overall volume of articles varies across newspapers and time, we scale the index by the total number of articles in the same newspaper and time period, which yields a news sentiment time series for each news. We then take average across the 16 news portals by a chosen frequency, whether it be daily, monthly or quarterly. Figure 1 shows the monthly news sentiment indices. The sentiment measures exhibit a pronounced drop during the economic downturns that happened during the 2007-08 financial crisis and the onset of COVID-19 pandemic in Malaysia, as well as the subsequent increase following the economic recovery from these crises.

Figure 1: Monthly news sentiment for Malaysia



Nowcasting survey-based measures of sentiment

To investigate the information content of the news sentiment, we first explore whether news sentiment can help nowcast the Business Condition Index (BCI) and Consumer Sentiment Index (CSI) published by the Malaysian Institute of Research (MIER). While these survey-based measures are closely followed by economic analysts, they are released only every quarter and with a lag of 2 months after the end of the reporting quarter. In contrast, the news sentiment can be constructed at a higher frequency, thereby providing information about economic sentiment between releases of these survey-based measures.

In this nowcasting exercise, we test whether the monthly news sentiment within the quarter can help predict the current quarter's BCI_t and CSI_t. We estimate the following for $m = 1, 2, 3$:

$$BCI_t = \alpha + \beta BCI_{t-1} + \eta x_{t,m} + \varepsilon_t \quad (1)$$

$$CSI_t = \alpha + \beta CSI_{t-1} + \eta x_{t,m} + \varepsilon_t \quad (2)$$

where $x_{t,1}$ is the value of the sentiment in the first month of the current quarter t, $x_{t,2}$ is the average value for the first two months of the quarter t and $x_{t,3}$ is the average value for the full quarter.

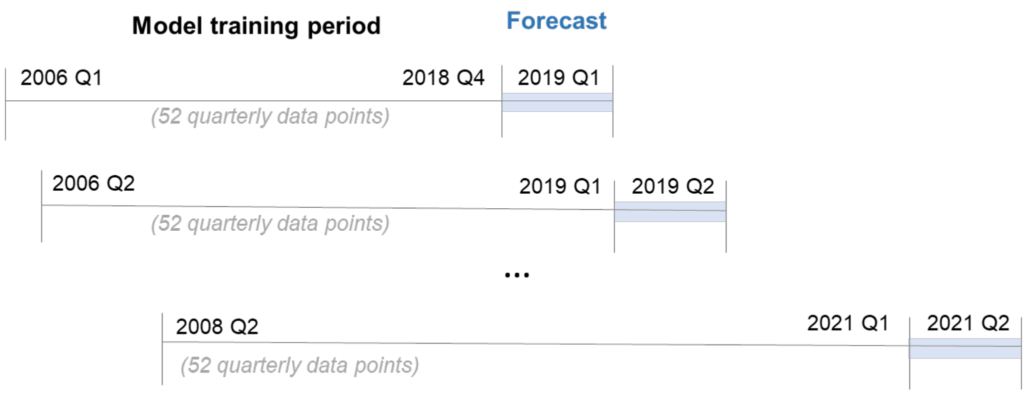
Forecasting GDP components using linear models with news sentiment

Next, we assess the forecasting ability of the news sentiment using linear models. Our forecasting targets variables are aggregated real GDP year-on-year growth and some of its demand-side components, namely year-on-year growth in private consumption, private investment, exports and imports. All target variables are at quarterly frequency and are standardised to z-scores for the forecasting exercises.

Our forecast exercises involve estimating a model over a specified training period using each of the three sentiment measures in turn. We train the models recursively on a rolling window, followed by producing out-of-sample predictions of target variables at horizon *h* = 1; 2; 3 quarters ahead. In other words, our forecast exercise seeks to mimic a scenario in which policymakers at time t have historical data of the target variable (i.e., $y_{t-1}, y_{t-2,...}$) and are anticipating/forecasting the official statistics y_{t-1+h} while having access to the news sentiment x_t.

For *h* =1, this replicates an actual forecasting situation starting from 2019 Q1 and moving forward a quarter at a time through to 2021 Q2 (or 2018 Q3 onwards for *h* = 3). For example, for the first vintage of the data, the models are estimated over the period 2006 Q1 to 2018 Q4 using data for both the chosen target variable, its lag and the news sentiment. The fitted models are then used to nowcast the response variable in 2019 Q1. As an example, Figure 2 illustrates the selected horizons for model training and 1-quarter ahead forecast period. Overall, we generate 10 real-time nowcasts of the response variables. The chosen period will also put to test the informational content of the proxy indicators during the recent macroeconomic stress caused by the pandemic.

Figure 2: Training period and 1-quarter ahead forecast horizon



We use the ordinary least square (OLS) linear regression method, and also include two more modified versions of linear regression:

- The Ridge regression which performs L2 regularisation where the model is penalised for the sum of squared value of the magnitude of the coefficients. λ is the parameter that determine the relative impact of the penalty terms. When λ = 0, we have the standard OLS approach.

$$\beta^{Ridge} = argmin \left[\sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{p} \mathrm{x}_{ij} \beta_j \right) + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$$

- The Huber regression which is robust to outlier by reducing the weight of large residuals. In Huber weighting, observations with small residuals get a weight of 1 and the larger the residual, the smaller the weight. This is defined by the weight function:

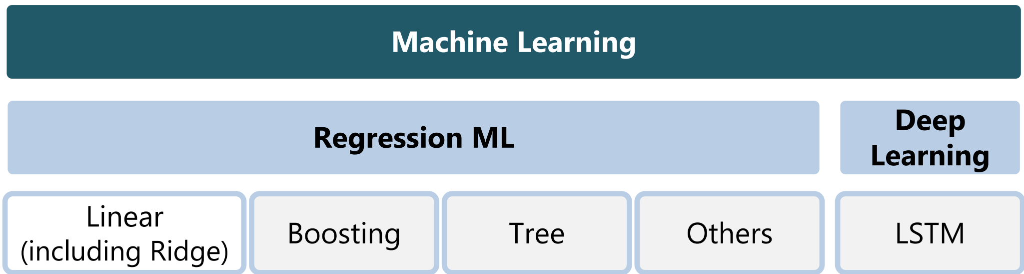
$$w(e) = \begin{cases} 1 & for \quad |e| \leq \mathrm{k} \\ \frac{k}{|e|} & for \quad |e| > \mathrm{k} \end{cases}$$

Forecasting GDP components using non-linear machine learning models with news sentiment

Beyond linear regression models, we also run a set of non-linear machine learning models. More generally, machine learning is a subset of artificial intelligence which aims to learn representation of knowledge of data in order to generate meaningful insights concerning both the data on hand and also on data that is unknown from the future. Supervised machine learning algorithms take an input (e.g., the sentiment of newspaper articles) to predict a future outcome, e.g., an economic indicator such as a country's GDP. Supervised regression depends on labelled data, which are pairs of inputs and outputs that has been sampled from historical data. The parameters of a supervised machine learning algorithm are tuned in order to minimise in the prediction error, for example the root mean squared error (RMSE).

In our paper, we wanted to investigate the viability of using supervised non-linear regression machine learning techniques in order to predict the growth variables for h = 1; 2; 3 quarters ahead. As with the linear regression models, we train the models recursively with news sentiment used as a feature on a rolling window, followed by producing out-of-sample predictions of the target variables, but with two distinct approaches: a suite of non-linear machine learning regression models and long short-term memory (LSTM) (only for h = 1).

Figure 3: Machine learning groups



The regression machine learning algorithms used are Light Gradient Boosting Machine, Random Forest Regressor, Extra Tree Regressor, Orthogonal Matching Pursuit, Gradient Boosting Regressor, Decision Tree Regressor, Adaboost Regressor and Passive Aggressive Regressor, which are grouped into distinct groups as shown in Figure 3.

One group of non-linear machine learning algorithms is called Boosting, and it groups weak machine learning algorithms in an ensemble in order to reduce bias and variance which reduces predictive power. Algorithms in this group include Light Gradient Boosting Machine, Gradient Boosting Regressor and the Adaboost Regressor. The following group, Tree, are highly interpretable algorithms which divides the prediction boundaries into simpler regions. In this group, we have algorithms such as the Random Forest Regressor, Extra Tree Regressor and the Decision Tree Regressor. The final group Others, contains algorithms such as Orthogonal Matching Pursuit.

Recurrent neural network (RNN) is a type of neural network with an internal memory. Because of this memory, RNN can remember information about the input that they have received, and this helps RNN to predict what future values precisely. This attribute is very useful for predicting the values of data that is organised in sequences, including time-series data. Long term short memory (LSTM) is an extension to RNN, in that it extends the memory. It is therefore very well suited to learn from data that has very long time-gaps in between. LSTM has been used to successfully predict macroeconomic indicators such as global merchandise exports value and volume (Hopp, 2021). LSTM is considered to be a type of deep learning algorithm. The learning process in deep learning algorithm is 'deep' because the underlying neural network has many layers that captures data (input layer), process data (hidden layer) and contains the predicted state of the data (output layer)

Regression machine learning algorithms and deep learning algorithms were chosen as we wanted to explore the fundamental differences in how prediction is made. Regression machine learning algorithms needs to be told how to represent data (the independent variables) while deep learning algorithms are able to create new data representation through data processing.

In the LSTM experiments, we wanted to ascertain the suitability of this technique given the frequency of data available in our context. Different combinations of independent variables were used to predict macroeconomic indicators, using different length of data.

We use the settings in Table 2 for the LSTM machine learning experiments. The evaluation measure used in the experiments is the average root mean squared error for the sliding window that we are using. The sliding window will start from 2006Q1 and move 1 quarter at each step. From this table, we can conceptually understand that the LSTM deep learning network will consider the past four quarter of observations in order to predict the following one-quarter ahead. We further considere different combinations of macroeconomic economic indicators and sentiment scores in order to predict a target macroeconomic indicator.

LSTM Experimental Setting

Table 2

Time period	2006Q1 to 2021Q2
Sliding window size	4
Number of quarters to predict ahead	1
Number of layers	5
Optimizer	Adam
Loss function	Huber

Forecast evaluation

For all models mentioned above, the out-of-sample root mean squared error (RMSE) is used as a metric to compare the nowcasting properties of the sentiment measures. RMSE is calculated on the out-of-sample forecast period with the standard formula:

$$RMSE = \sqrt{\sum_{t=1}^{N} \frac{(\hat{y}_t - y_t)^2}{N}}$$

where $\hat{y}_t$ is the predicted value for the time period t, y_t is the actual value and N is the total number of predicted observations.

We compare the performance of the model with news sentiment to a pure OLS-AR(1) model, which is the baseline model without the sentiment measure x_t. For example, in the case of the linear models, we compare

AR(1) with sentiment: $y_{t+h} = \alpha + \beta\, y_{t-1} + \eta\, x_t + \epsilon_t$

OLS- AR(1) baseline: $y_{t+h} = \alpha + \beta\, y_{t-1} + \epsilon_t$

where $x_t \in \{Correa_t, LM_t, SSW_t\}$

For each forecast, we calculate the ratio RMSE which is the model's RMSE relative to the OLS-AR(1) model.

$$Ratio\ RMSE = \frac{RMSE}{RMSE_{AR1}}$$

A ratio RMSE of less than 1 indicates that the model performed better compared to the benchmark model. Conversely, a ratio RMSE of greater than 1 indicates that the model has performed worse compared to the benchmark model. For a given target variable, we calculate the average ratio RMSE across the out-of-sample periods.

4. Results and Discussion

Pearson correlation coefficient

Table 3

	Full sample 2006 Q1 – 2021 Q2			Non-crisis 2011 Q1 – 2020 Q4		
	Correa	LM	SSW	Correa	LM	SSW
Macroeconomic variables:						
Aggregate GDP	0.57	0.52	0.41	0.31	0.27	0.32
Private investment	0.49	0.51	0.48	0.56	0.61	0.65
Private consumption	0.60	0.53	0.38	0.23	0.16	0.18
Exports	0.25	0.22	0.22	-0.03	-0.11	-0.13
Imports	0.28	0.28	0.28	0.23	0.18	0.17

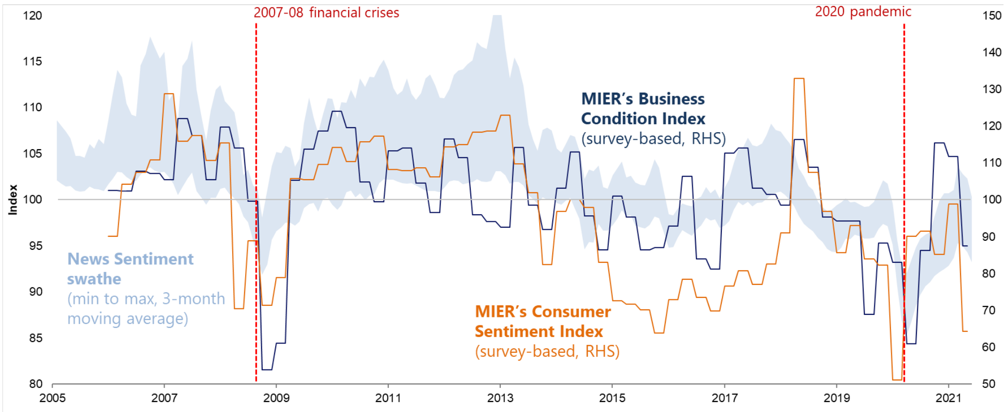
Sources: Authors’ calculation

As a starting point, we compare each news sentiment to the time series of aggregate GDP growth and its components by looking at their contemporaneous correlations as shown in Table 3. On average, the news sentiments’ correlation with aggregate GDP growth as well as private sector economic activities are relatively higher and of the expected sign. Trade activities (exports and imports) appear to be weakly correlated with the news sentiment measures, possibly reflecting the coverage of topics in our news corpuses that may lean towards domestic-oriented economic activities. Another possibility is a stronger lead-lag relationship between news sentiment and trade activities. After excluding the 2007-08 financial crisis and the recent pandemic-induced crisis, there is a noticeable decline in the correlations between the macroeconomic growth variables and news sentiments, with the exception of private investment. This suggests that even during non-crisis period, the news sentiment may contain information regarding investment activities in the private sector.

Nowcasting survey-based measures of sentiment

We now assess the information content of the news sentiment vis-à-vis the survey-based measures of economic sentiment. Both news and survey-based sentiment appear to move in tandem and exhibit sharp declines during the 2007-08 financial crisis and onset of COVID-19.

Figure 4: Monthly news sentiment and survey-based sentiment measures



Tables 4a documents the results of using news sentiment to nowcast the survey-based MIER's BCI. Before adding the news sentiment measures, we find that the prior quarter's release of the survey-based sentiment is statistically significant in explaining the current quarter's release. Beyond column (1), when we add the news sentiment measure, reflecting the information set available with each passing month in the quarter, we find that this news sentiment measure is statistically significant throughout. Perhaps not surprisingly, the coefficient of the sentiment measure and adjusted R-squared also generally increase as more news within the quarter are incorporated into the news sentiment measure, suggesting increasing information content in nowcasting MIER's BCI in the current quarter. Similar results are also obtained when we repeat the exercise with MIER's CSI in Table 4b, but the statistical significance is weaker compared to that of BCI.

Nowcasting Quarterly Survey-Based Sentiment

Dependent variable: MIER's Business Conditions Index (BCI)
Sample: 2006 Q1 – 2021 Q2

Table 4a

	(1) AR1	(2) Correa	(3) Correa	(4) Correa	(5) LM	(6) LM	(7) LM	(8) SSW	(9) SSW	(10) SSW
MIER's Sentiment (1 quarter prior)	0.49***	0.36***	0.33***	0.33***	0.35***	0.34***	0.35***	0.35***	0.31***	0.33***
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
News sentiment (first month of quarter)		0.30**			0.32**			0.41***		
		(0.14)			(0.14)			(0.14)		
News sentiment (average of first 2 months)			0.41***			0.39***			0.52***	
			(0.15)			(0.14)			(0.15)	
News sentiment (average for the quarter)				0.45***			0.40***			0.48***
				(0.14)			(0.13)			(0.14)
Constant	-0.01	-0.01	-0.01	0.02	-0.03	-0.02	0.00	-0.02	-0.03	-0.02
	(0.11)	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)
Adjusted R^2	0.22	0.29	0.34	0.35	0.30	0.32	0.32	0.35	0.39	0.35
Observations	61	61	61	61	61	61	61	61	61	61

Heteroscedastic and autocorrelation robust (HAC) standard errors in parentheses. * p < 0:10 ** p < 0:05 *** p < 0:01

Source: Authors' calculation

Nowcasting Quarterly Survey-Based Sentiment

Dependent variable: MIER's Consumer Sentiment Index (CSI)
Sample: 2006 Q1 – 2021 Q2

Table 4b

	(1) AR1	(2) Correa	(3) Correa	(4) Correa	(5) LM	(6) LM	(7) LM	(8) SSW	(9) SSW	(10) SSW
MIER's Sentiment (1 quarter prior)	0.71***	0.65***	0.59***	0.55***	0.54***	0.54***	0.52***	0.50***	0.44***	0.42***
	(0.09)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.13)
News sentiment (first month of quarter)		0.10			0.27*			0.32***		
		(0.14)			(0.14)			(0.08)		
News sentiment (first 2 months of quarter)			0.22			0.30**			0.44***	
			(0.16)			(0.14)			(0.10)	
News sentiment (full data for the quarter)				0.30*			0.35***			0.48***
				(0.15)			(0.13)			(0.11)
Constant	-0.01	-0.01	-0.01	0.00	-0.03	-0.02	-0.00	-0.02	-0.03	-0.02
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
Adjusted R^2	0.47	0.47	0.49	0.51	0.51	0.51	0.52	0.51	0.55	0.55
Observations	61	61	61	61	61	61	61	61	61	61

Heteroscedastic and autocorrelation robust (HAC) standard errors in parentheses. * p < 0:10 ** p < 0:05 *** p < 0:01

Source: Authors' calculation

Forecasting GDP components using linear models

We now look at the predictive ability of the news sentiment for GDP components in a linear model setting. Figure 5 shows the average ratio RMSEs (across forecast period and news sentiment measures) for the out-of-sample forecast relative to the OLS-AR(1) baseline model for each target variable. Bars below the red line indicate ratio RMSE values of less than 1, which may be interpreted as an improvement of the forecast over the OLS-AR(1) model

Across the OLS, Ridge, and Huber regressions, we observe that the forecasts of private investment over the 2- and 3-year forecast horizon have ratio RMSEs less than 1, which indicates improvement in performance with the addition of news sentiment compared to the baseline model. Bars with ** further indicate ratio RMSEs that are below 1 for at least 90% of the chosen forecast periods. Furthermore, Figure 6 shows that despite worsening performance across all models following the onset of the pandemic in Q2 2020, news sentiment still offers a significant improvement in the forecast performance relative to the benchmark OLS-AR(1) model.

For aggregate GDP and other GDP components, however, the results are relatively weaker across all forecast horizons compared to the AR(1) model, except for trade activities at the 3-quarter horizon. These suggest that news sentiment has a relatively stronger influence on investment decision but not so much on the broader economic activity.

Figure 5: Ratio RMSE using linear regression models

Figure 5 (continued): Ratio RMSE using linear regression models

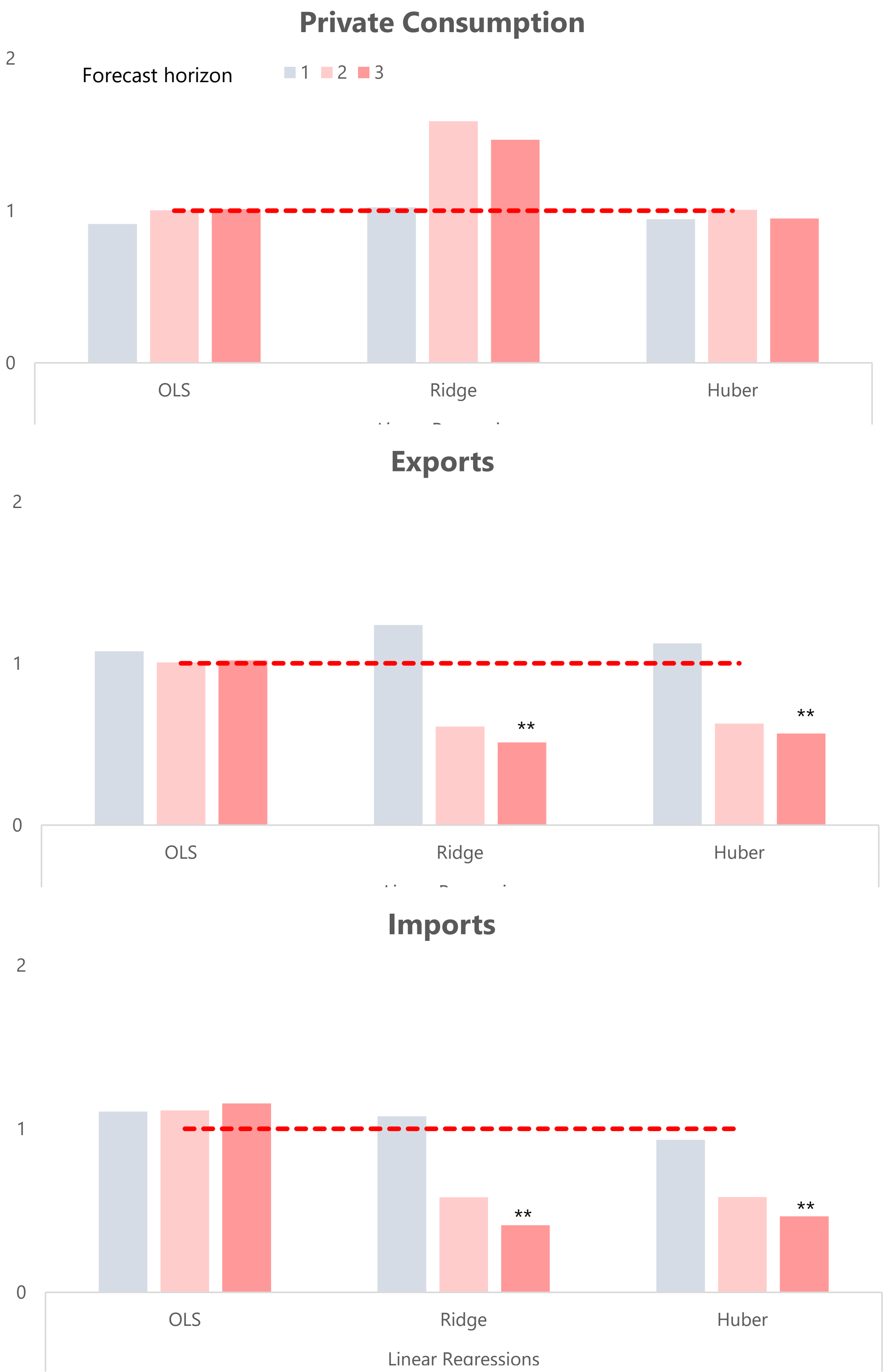
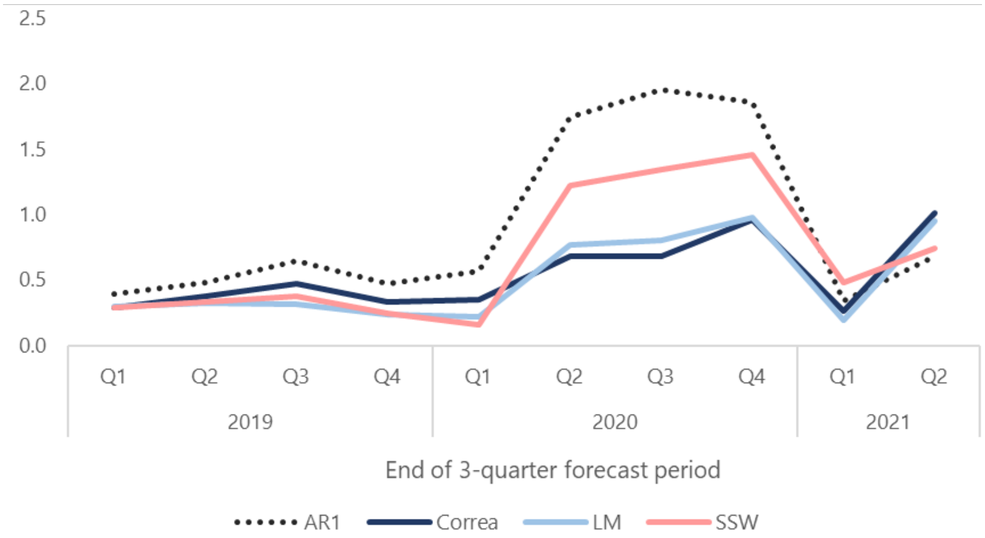


Figure 6: Comparison of out-of-sample RMSEs (3-quarter ahead) between the AR(1) model and the OLS model with news sentiment.



Forecasting GDP components using non-linear machine learning models

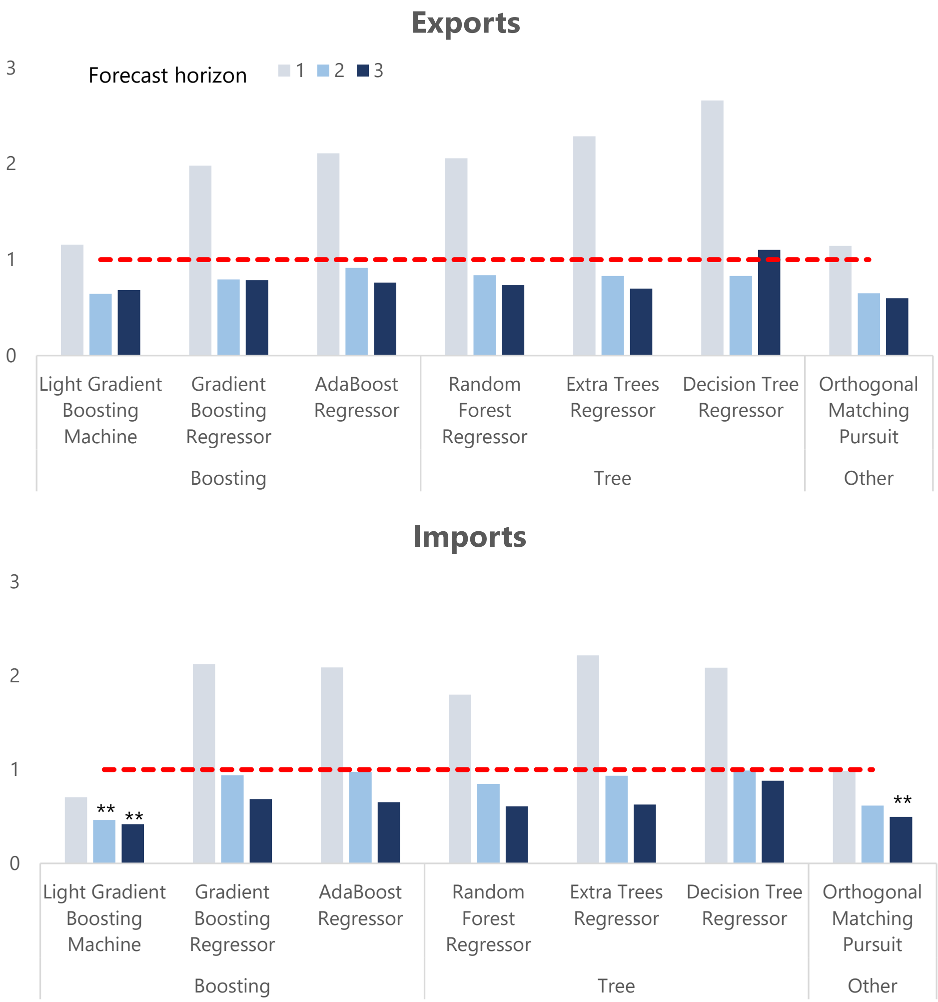
Figure 7 expands the earlier results of predicting GDP and its various components using a range of non-linear machine learning algorithms from the *Boosting*, *Tree* and *Others* groups with news sentiment. Again, bars below the red line can be interpreted as an improvement of the forecast over the OLS-AR(1) model, and bars marked with ** further indicate ratio RMSEs that are below 1 for at least 90% of the chosen forecast periods.

Similar to the results from linear regressions, we observe notable improvements in the performance of forecasting private investment growth 2- and 3-quarter ahead. In this regard, the AdaBoost Regressor, Random Forest Regressors and Extra Trees Regressors show reliable improvement in forecasting private investment growth across the forecast periods. As with the linear regressions, the forecast improvements shown by the machine learning models with news sentiment persist even during the crisis in 2020 (results not shown).

The machine learning algorithms also exhibit improvements for the 2- and 3-quarter predictions of exports and imports growth, although with less consistency across time period compared to private investment growth. The Light Gradient Boosting Machine nevertheless performs well for predicting imports growth. Overall, non-linear machine learning algorithms performed poorly when it comes to predicting GDP and private consumption. Further investigation on the usage of machine learning to estimate macroeconomic indicators is however warranted due to good results achieved recently (Cicceri et al., 2020; Richardson et al., 2018).

Figure 7: Ratio RMSE using non-linear machine learning algorithms

Figure 7 (continued): Ratio RMSE using non-linear machine learning algorithms



Experimental Results using LSTM

Average RMSE across sliding windows for LSTM

Table 5

Input variable(s) in LSTM (with past four quarter of observations)	Target variables				
	GDP	Private investment	Private consumption	Exports	Imports
(1) GDP	0.58				
(2) Private investment		0.32			
(3) Private consumption			0.29		
(4) Exports				0.95	
(5) Imports					0.80
(6) GDP and Correa	0.55				
(7) GDP and LM	0.53				
(8) GDP and SSW	0.56				
(9) Private investment and Correa		0.30			
(10) Private investment and LM		0.27			
(11) Private investment and SSW		0.18			
(12) Private consumption and Correa			0.19		
(13) Private consumption and LM			0.18		
(14) Private consumption and SSW			0.18		
(15) Exports and Correa				1.20	
(16) Exports and LM				1.16	
(17) Exports and SSW				1.08	
(18) Imports and Correa					1.06
(19) Imports and LM					0.97
(20) Imports and SSW					0.88
(21) GDP, Private investment, Private consumption, Exports, Imports, Correa, LM, SSW	0.55	0.19	0.14	1.02	0.90

In Table 5, we show our experimental results using LSTM. Our LSTM experiments were focused on investigating the effect of newspaper sentiment on predicting macroeconomic indicators. The shaded cells indicate the lowest value for RMSE for each target variable.

Rows 6 till 21 show the effect of newspaper sentiment on predicting macroeconomic indicators and here, the results are mixed but broadly similar to the findings from the linear and non-linear models above. Compared to the RMSEs of models where the lagged values of the target variable are the only input in the LSTM network (rows 1 to 5), there is an improvement for GDP, private consumption and especially for private investment when newspaper sentiment is included as an additional input to the LSTM. For exports and imports, however, the prediction results worsened after incorporating newspaper sentiment. What happens if we were to incorporate more variables in the LSTM model? In Row 21, eight features were used as inputs and the prediction results improved for private investment, private consumption and GDP but worsened again for exports and imports.

Nevertheless, this study deals with only a relatively short time series for LSTM. The architecture of the LSTM model, the number of times the entire time series is passed through the network (epoch), the choice of the optimizer and loss function, batch size and network constructs (two autoencoders in this case) would all need to be investigated thoroughly for a conclusive answer.

5. Conclusion

Our observations suggest that the sentiment embodied in the business and financial news from online newspaper portals corresponds to the survey-based business sentiment measure and can provide forward-looking indication of investment activities in Malaysia. Specifically, we find that the monthly news sentiment – even in the early part of the quarter – can explain movements in the quarterly survey-based business sentiment indicators that are often published with a lag.

We investigated the extent to which news sentiment can help predict economic growth outcomes, finding that the news sentiment can consistently forecast private investment growth better than the benchmark OLS-AR(1) model, especially within 2 – 3 quarters ahead, but not other components of economic activity. The forecast gains for private investment holds true even during the recent period of macroeconomic stress following the pandemic. This suggests the extracted news sentiment can provide a timelier read of investment activities in Malaysia even during economic turning points but has limited information content for the broader economy.

There are several avenues for future research. First, we only explore and compare the sentiment in English local news in Malaysia, but to be truly representative of the Malaysia's newspaper readership in terms of ideological predisposition, one needs to also draw text from vernacular-based newspapers (e.g. Malay, Chinese and Tamil). To our knowledge, well-established algorithms to analyse the sentiment of such text remain limited. Yet another avenue for future research is to consider the ideology of a news article as a feature for machine learning models. In addition, one may also consider identifying the sentiment in news articles from other sections. For example, while we find that business and financial news are associated with investment activity, the language used in non-business articles is plausibly associated with changes in consumer sentiments as well as household spending. Finally, using non-linear machine learning models such as LSTM to improve forecasting performance seems to be hindered by the limited data points as we collapse article text into a single time series. To get the most out of text, it may be worthwhile to have more granular data points, similar to work by Kalamara et al. (2020) who retain thousands of terms retained from text as a larger set of time series that can be used with machine learning models.

References

Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. (2021). Can news help measure economic sentiment? An application in COVID-19 times. *Economics Letters*, *199*, 109730. https://doi.org/10.1016/j.econlet.2021.109730

Alexopoulos, M., & Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. *International Review of Economics and Finance*, *40*. https://doi.org/10.1016/j.iref.2015.02.002

Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, *35*(4), 1370–1386. https://doi.org/10.1016/j.ijforecast.2018.10.010

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, *28*(2). https://doi.org/10.1257/jep.28.2.153

Buckman, S. R., Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). News Sentiment in the Time of COVID-19. *FRBSF Economic Letter*, *08*.

Calomiris, C. W., & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, *133*(2). https://doi.org/10.1016/j.jfineco.2018.11.009

Cicceri, G., Inserra, G., & Limosani, M. (2020). A Machine Learning Approach to Forecast Economic Recessions—An Italian Case Study. *Mathematics*, *8*(2), 241. https://doi.org/10.3390/math8020241

Correa, R., Garud, K., Londono-Yarce, J.-M., & Mislang, N. (2017). Constructing a Dictionary for Financial Stability. *IFDP Notes*, *2017*(33), 1–7. https://doi.org/10.17016/2573-2129.33

Fraiberger, S. P. (2016). *News Sentiment and Cross-Country Fluctuations*.

García, D. (2013). Sentiment during Recessions. *Journal of Finance*, *68*(3). https://doi.org/10.1111/jofi.12027

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1). https://doi.org/10.3982/ECTA7195

Hopp, D. (2021). Economic Nowcasting with Long Short-term Memory Artificial Neural Networks (LSTM). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3855402

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). Making Text Count: Economic Forecasting Using Newspaper Text. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3610770

Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, *210*(1), 203–218. https://doi.org/10.1016/j.jeconom.2018.11.013

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Moore, A. (2017). Measuring Economic Uncertainty and Its Effects. *Economic Record*, *93*(303). https://doi.org/10.1111/1475-4932.12356

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Kleis Nielsen, R. (2020). *Reuters Institute Digital News Report 2020* (p. 99). Oxford: Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Nguyen, K., & Cava, G. L. (2020). *Start Spreading the News: News Sentiment and Economic Activity in Australia*. https://www.rba.gov.au

Nothman, J., Qin, H., & Yurchak, R. (2019). *Stop Word Lists in Free Open-source Software Packages*. https://doi.org/10.18653/v1/w18-2502

Nyman, R., Kapadia, S., & Tuckett, D. (2021). News and narratives in financial systems: Exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control*, *127*. https://doi.org/10.1016/j.jedc.2021.104119

Rambaccussing, D., & Kwiatkowski, A. (2020). Forecasting with news sentiment: Evidence with UK newspapers. *International Journal of Forecasting*, *36*(4), 1501–1516. https://doi.org/10.1016/j.ijforecast.2020.04.002

Richardson, A., Mulder, T., & I Vehbi, T. (2018). Nowcasting New Zealand GDP using machine learning algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3256578

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.07.053

Shapiro, A. H., & Wilson, D. J. (2019). Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis. *Federal Reserve Bank of San Francisco, Working Paper Series*. https://doi.org/10.24148/wp2019-02

Thorsrud, L. A. (2020). Words are the New Numbers: A Newsy Coincident Index of the Business Cycle. *Journal of Business & Economic Statistics*, *38*(2), 393–409. https://doi.org/10.1080/07350015.2018.1506344

Using News Sentiment for Economic Forecasting

A Malaysian Case Study

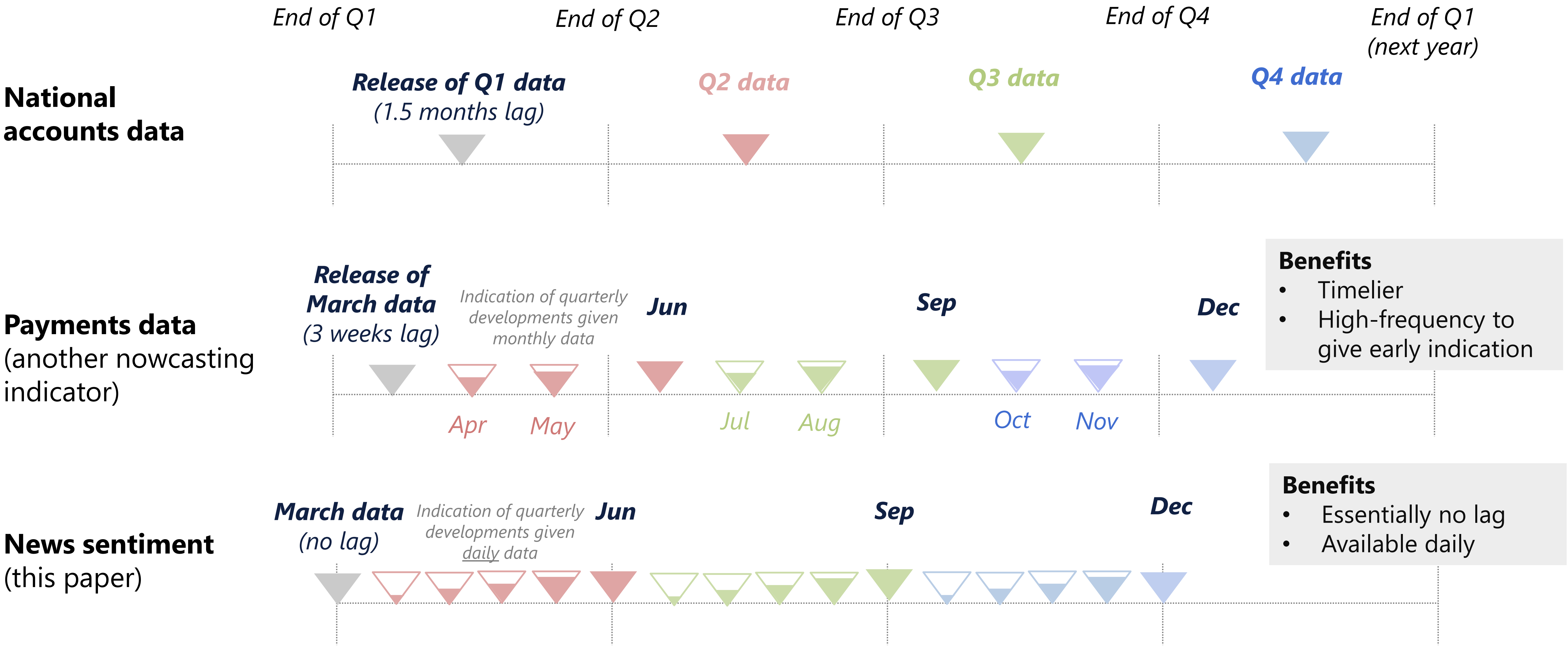
Eilyn Chong, Chiung Ching Ho, Zhong Fei Ong and Hong H. Ong

Bank Negara Malaysia

Outline

- **Motivation**
- **Literature Review**
- **Data**
- **Methodology**
- **Selection of results and conclusions**

Increasingly, central banks have been relying on timelier indicators to assess the near-term developments of the economy in advance of the release of official statistics



A growing literature has made use of the wide availability of text in digital format, developments in computational linguistic methods and better computing power

- **Computational linguistic methods have been applied to text in digital format to provide quantitative measures**
  - Economic policy uncertainty (Alexopoulos and Cohen, 2015; Baker, Bloom and Davis, 2016)
  - Daily economic sentiment (Buckman et al., 2020; Thorsrud, 2020)
  - Others: Media slant (Gentzkow & Shapiro, 2010) and central bank's objective function (Shapiro & Wilson, 2019)

- **The papers also relate the text-based measures to a variety of economic and financial outcomes.**
  - Text from news with imminent financial distress (Manela & Moreira, 2017; Nyman et al. , 2021)
  - Text from news/firms' annual reports with stock returns (Calomiris & Mamaysky, 2019; García, 2013; Jegadeesh & Wu, 2013,' Loughran & McDonald, 2011)
  - Text-based measures of uncertainty with business cycle (Bloom, 2014; Moore, 2017)

- **This paper is closest to literature that use news sentiment to track and predict a range of macroeconomic variables.**
  - Aguilar et al. (2021); Fraiberger (2016); Kalamara et al. (2020); Larsen & Thorsrud (2019); Nguyen & Cava (2020); Rambaccussing & Kwiatkowski (2020) show that newspaper text contains information on the future path of certain macroeconomic variables, e.g. GDP.

Data: Our news corpus consists of over 720 thousands economic and financial news articles from major news portals since early 2000s

Descriptive statistics of articles from selected local news portals

Table 1

	Average number of articles per month	Date of first online article	Readership (selected news media)[1]
klse.i3investor	7898	03/03/2020	
The Star	949	01/01/2003	30%
The Edge Markets	912	16/01/2009	
Malay Mail	827	18/06/2013	8%
Bernama	654	04/03/2020	
The Malaysian Reserve	487	10/01/2017	
Free Malaysia Today	384	31/12/2015	15%
The Borneo Post	284	23/12/2009	
The Sun Daily	280	15/11/2017	
SoyaCincau	273	29/01/2020	
New Straits Times	238	20/05/2014	10%

Note: 1. Based on the Reuters Institute Digital News Report 2020 (Newman et al. 2020). Readership is defined as the share of respondents who consumed the media at least once a week.
Sources: Authors' calculation

Methodology – Sentiment scoring: We apply the dictionary-based method on the news corpus to construct the news sentiment index

- Leverage on the lexicon approach using:
  - Financial stability dictionary by Correa et al. (2017) (hereafter Correa)
  - Finance-oriented dictionary by Loughran & Mcdonald (2011) (hereafter LM)
  - Lexicon created by Shapiro, Sudhof and Wilson (2020) (hereafter SSW), who scored a corpus of U.S. economic news articles with Vader (scores ranging from -4 to 4)

- Construction of news sentiment index using words w in each news article i :
  - Correa and LM: $$sentiment\ index_i = 100 + \frac{\sum Positive_i - \sum Negative_i}{Word\ count_i} \times 1000$$
  - SSW: $$sentiment\ index_i = 100 + \sum_w \frac{Score_{w,i}}{Word\ count_i} \times 1000$$

- $sentiment\ index_i$ is then averaged for a chosen frequency, whether it be daily, monthly or quarterly.

Methodology: We assess whether news sentiment can nowcast survey-based sentiment & compare the forecasting performance of models with news sentiment to AR (1) model

Various daily news sources

Text pre-processing

- Remove punctuations, hyperlinks, HTML tags, special characters, extra white spaces
- Drop stop words
- Set words in lowercases

Sentiment indices using lexicons, x_t
(2006 – 2021 Q2)

- Correa
- LM
- SSW

1 **Nowcasting quarterly survey-based sentiment**
(2006 – 2021 Q2)

- Business conditions index
- Consumer sentiment index

2 **Forecasting target variables,**
y_t (2006 – 2021 Q2)

- Aggregate GDP
- Private consumption
- Private investment
- Imports
- Exports

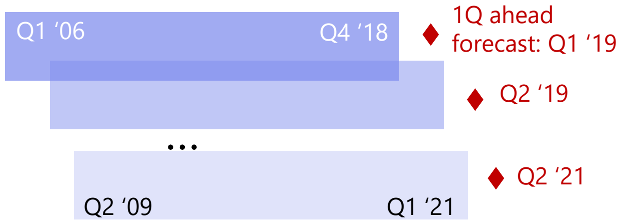
Split into train-test period

Train

Test

Rolling-window estimations
comparing forecasts from…

- **OLS-AR1**
  $y_{t+h} = \alpha + \beta\, y_{t-1} + \epsilon_t$
- **AR1 with sentiment**
  $y_{t+h} = f\,(y_{t-1}, \boldsymbol{\eta}\, \boldsymbol{x_t})$
  where $x_t \in \{Correa_t, LM_t, SSW_t\}$

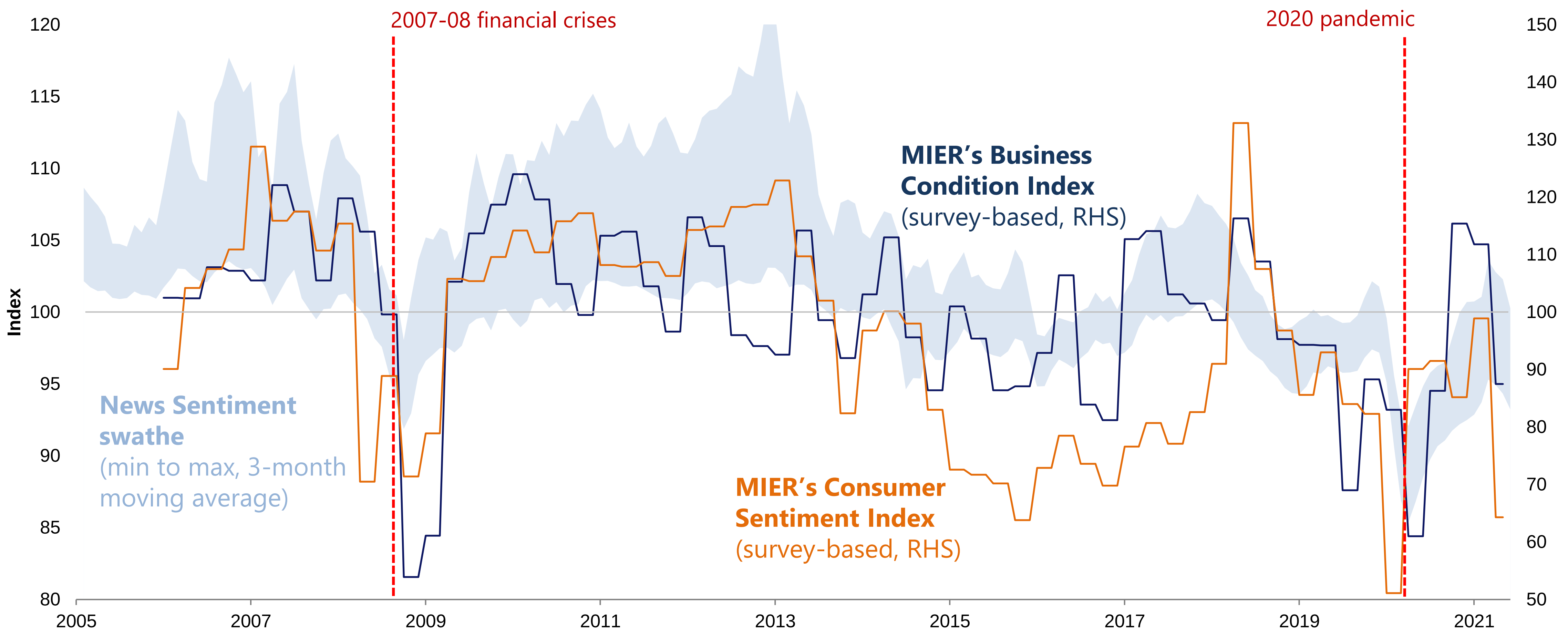


Evaluation of h-quarter ahead forecast

$$Ratio\ RMSE = \frac{RMSE_{model_i}}{RMSE_{AR1}}$$

$$for\ h = 1,2,3$$

The monthly news sentiment measures move with fluctuations in economic conditions and exhibit the sharp declines during the 2007-08 financial crisis and onset of COVID19



Monthly news sentiment provides information in nowcasting movements in the quarterly survey-based measures of business sentiment

Nowcasting Quarterly Survey-Based Sentiment

Dependent variable: MIER's Business Conditions Index (BCI) in the same quarter
Sample: 2006 Q1 – 2021 Q2

Table 3

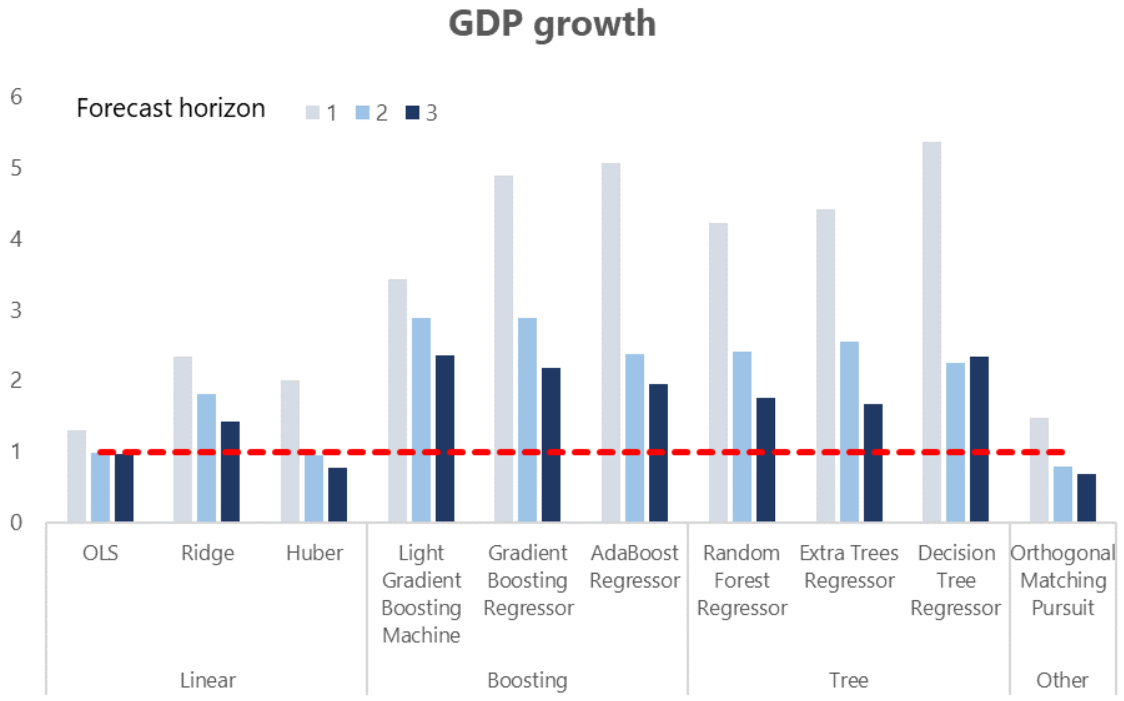
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	AR1	Correa	Correa	Correa	LM	LM	LM	SSW	SSW	SSW
MIER's Sentiment (1 quarter prior)	0.49***	0.36***	0.33***	0.33***	0.35***	0.34***	0.35***	0.35***	0.31***	0.33***
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
News sentiment (first month of the quarter)		0.30**			0.32**			0.41***		
		(0.14)			(0.14)			(0.14)		
News sentiment (average first 2 months of the quarter)			0.41***			0.39***			0.52***	
			(0.15)			(0.14)			(0.15)	
News sentiment (full data for the quarter)				0.45***			0.40***			0.48***
				(0.14)			(0.13)			(0.14)
Constant	-0.01	-0.01	-0.01	0.02	-0.03	-0.02	0.00	-0.02	-0.03	-0.02
	(0.11)	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)
Adjusted R^2	0.22	0.29	0.34	0.35	0.30	0.32	0.32	0.35	0.39	0.35
Observations	61	61	61	61	61	61	61	61	61	61

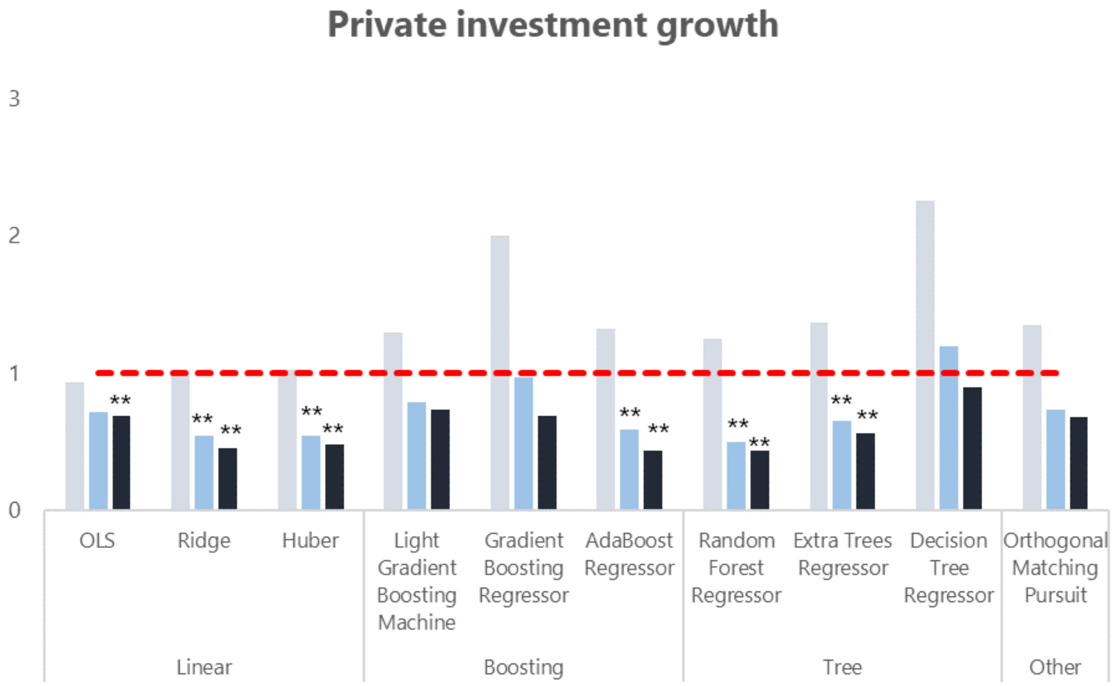
Heteroscedastic and autocorrelation robust (HAC) standard errors in parentheses. * p < 0:10 ** p < 0:05 *** p < 0:01

Sources: Authors' calculations

Among the GDP components, the news sentiment is predictive of private investment growth 2-3 quarters ahead

Ratio of RMSE of models with news sentiment vs. RMSE of OLS-AR(1)
(average across time period and sentiment measures)





Bars below the dashed red line indicate an improvement in forecast performance, and those with ** exhibit a forecast improvement for at least 90% of the forecast periods when compared to the OLS with just the AR(1) term only

Conclusion

1. The news sentiment measures move with the fluctuations in economic conditions and can provide information in nowcasting movements in the less timely, quarterly survey-based business sentiment.

2. Among the GDP components, the news sentiment is able to reliably forecast private investment growth 2 – 3 quarters ahead.

Future work

Explore and compare the sentiment in news of vernacular languages (e.g. Malay, Chinese and Tamil).

Generate new models to show better predictive accuracy than existing models by taking into the account of articles' ideological predisposition.