The Nowcasting of Economic Conditions in Japan Using Machine Learning

Taketo Muroya

Quantitative Methods in the Social Sciences,

Graduate School of Arts and Sciences,

Columbia University

May. 14, 2020

Abstract

Since official economic statistics in Japan are usually published more than one month later than the period to which they refer, it is difficult to assess the current economic conditions in a timely and accurate manner. Therefore, this study tries to nowcast the Index of Business Conditions (IBC), which is supposed to represent monthly economic conditions, using the text data from economic reports, and the numeric values from Google searches and electricity usage. The Japanese government's Monthly Economic Report, which is published slightly earlier than the IBC, is used as the text data in this study. Using the Natural Language Processing (NLP) technique, this text data is transformed into numeric values, such as text vectors and sentiment scores. The nowcasting of the IBC is mainly made by machine learning methods, especially the Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM), in addition to the traditional econometrics model. As a result, the text vectors and sentiment scores from economic reports nowcasts the IBC with high test scores, though it is not very timely data. The Google search and electricity usage data, which are available on a daily basis, nowcast the IBC with relatively high performances in a very timely fashion. Moreover, if the industry electricity usage, which is more highly correlated with economic activities than household one is, could become available, the nowcasting of economic conditions would improve.

Introduction

To make better decisions for policymaking or investment, a timely and accurate assessment of current economic conditions would be essential. Most economic statistics are produced by governments, which takes a long time to collect data and compile information. For example, the quarterly GDP data is usually published one to four months later than the period to which it refers. To be more specific, the first-quarter estimate (January - March) of the Japanese GDP is released in May. As a result, it is difficult for governments to implement the desirable fiscal or monetary policy at the time when the economy is actually in decline. Moreover, without seeing what is really going on under current economic conditions, companies might not be able to make investment decisions. Especially, when the economy confronts an unprecedented crisis, such as COVID-19, it is extremely important to assess its actual impact on economic conditions as soon as possible. Therefore, governments or investors need a timelier indicator to examine current economic conditions without waiting for the lagged official economic statistics.

This study selects economic reports as the text data to nowcast economic conditions, specifically, the Monthly Economic Report by the Cabinet Office. This data is relatively easy to extract text data from the websites and has consistent phrases to describe economic conditions within the reports over time. The Natural Language Processing (NLP) technique allows us to translate this text data on this report into numeric values, which is useful for generating the economic index for nowcasting the economic conditions. Recently, the NLP technique has developed rapidly, which allows the machine to interpret the human language and provide useful information for quantitative analysis.

Other than this text data from economic reports, the Google search and electricity usage data are prospective sources for the economic assessment. These data would be correlated with economic activities to some extent because people tend to search certain words when the economy is in a recession or booming, and also most industries use electricity when operating their factories or offices. Since both Google search and electricity usage data are available even on a daily basis, it would be possible to nowcast the economic conditions in a much timelier way than the traditional statistics.

As a target variable that represents the economic conditions in Japan, this study chooses the Index of Business Conditions (IBC) because it is available on a monthly basis, which is more frequent than GDP publications. In addition, this index is officially supposed to represent the Japanese economy because of being used for deciding the recession or booming of the economy by the Government. However, this index has not published in a timely manner, taking more than one month after the reference period. Therefore, if other indicators or information could nowcast this index accurately before the index is actually published, it would improve the assessment of current economic conditions in terms of timeliness.

In addition to the traditional econometrics model, such as the Autoregressive Integrated Moving Average (ARIMA) model, this study develops machine learning models, the Random Forest and Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM), to nowcast Japanese economic conditions. By using the power of calculation and deep understanding of relationships between predictors and targets, the machine learning models would generate more accurate output than traditional econometric models.

In the following chapters, first, this study examines the previous studies that researchers have conducted thus far, and then figures out issues worthy of further study. Second, it briefly mentions the methodologies and data that are used for the nowcasting economic conditions in a timely and accurate manner. Third, it spells out how variables are generated and then used for the predictions and what results are obtained from the models and the data. Finally, it discusses the limitations and conclusions of this research.

Literature Review

Scholars have selected various monthly, weekly, and daily indicators to nowcast economic conditions. The monthly traditional economic statistics that are published ahead of the GDP release, were used to predict the U.S. and Japanese GDP (Bok et al., 2017; Urasawa, 2014). The monthly text data in economic reports was used to produce an economic sentiment index of Japan (Yamamoto & Matsuo, 2016). The weekly Google search data, which Google provides with the European Central Bank, was used for the prediction of German GDP (Götz & Knetsch, 2019). The daily financial indicators, such as the Nikkei stock average and nominal exchange rate, were used to nowcast the state of the Japanese economy (Bragoli, 2017). The daily electricity usage was used to assess manufacturing and regional economic indexes of Japan (Suimon et al., 2019; Nitta, 2019).

However, some of the variables are not effective as predictors. For example, adding a Nikkei stock average and the nominal exchange rate into models had little impact on the performance of models (Bragoli, 2017). Even though the daily financial data is available in a very timely manner, it does not contribute to the improvement of predictions because of its much fluctuation and noisiness. This study focuses on using the economic reports, Google search, and electricity usage data, as mentioned before. In previous studies (Yamamoto & Matsuo, 2016; Götz & Knetsch, 2019; Suimon et al., 2019; Nitta, 2019), these indicators were confirmed to somewhat correlate with economic activities.

While most of the previous studies selected either the Dynamic Factor (Bok et al., 2017; Urasawa, 2014; Bragoli, 2017) or Bridge Equation (Götz & Knetsch, 2019) model to nowcast economic conditions, it is still difficult to predict the economic index well enough for actual investment or policymaking. For example, there is more than a 2 percent point diversion between the actual GDP and the predicted value in the model (Bok et al., 2017). Thus, instead of these econometric approaches, my research tries to use machine learning methods, such as the neural network model, to nowcast the economic conditions more accurately. Machine learning models could be more appropriate for predictions when they deal with many variables and represent the complicated relationships between predictors and targets.

Among machine learning models, Yamamoto and Matsuo (2016) used the Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM) model. This model is one of the recurrent neural network methods that perform well for time-series or text data due to taking into consideration the sequential characteristics of the dataset. Since Yamamoto and Matsuo (2016) concluded that the composite index from economic text data correlates well with other economic indicators, the RNN-LSTM could work for this study as well. Therefore, this study tries to apply the RNN-LSTM to not only the text data but also Google search and electricity usage data.

All in all, this study tries to develop a machine learning model that nowcasts economic conditions with high performance regarding the Japanese economy because such a model has not been constructed yet. However, if too many variables are included in one model to predict economic conditions, it might produce unnecessary noise. Therefore, it would be important to select only meaningful variables as predictors. In addition, in order to use the nowcasting method for actual investment or policy decisions, the performance of the projections should be improved. Thus, my study tries to use the RNN-LSTM to improve nowcasting.

Methodology

Natural Language Processing (NLP)

For the text analysis of economic reports, the NLP technique allows us to convert the text data into numeric values. The first way to do this is the vectorization of the text data. The NLP algorithm counts the frequencies of words and then generates the vectors that correspond to the text data, which is called Count Vectorizing. If it also assigns weights on the word frequencies (tf) based on the importance of the word (idf), it is called Tf-idf Vectorizing, which in general would improve the transformation of text data into numeric vectors. These vectors generated from economic reports would be predictors to nowcast current economic conditions.

$$Tf\ idf\ =\ tf\ *\ idf$$

$$tf\ =\frac{Term\ X\ Frequency\ in\ a\ Document}{Sum\ of\ all\ Term\ Frequency\ in\ a\ Document}$$

$$idf\ =log(\frac{number\ of\ all\ documents}{number\ of\ documents\ which\ include\ term\ X})$$

The alternative way is the sentiment analysis that simply counts positive and negative words in the text data and then generates the sentiment index for each text data. If the economic reports are well written about the economic conditions, the sentiment index of them should be well correlated with the IBC.

Autoregressive Integrated Moving Average (ARIMA)

One of the traditional econometric approaches for forecasting is the Autoregressive integrated moving average (ARIMA) model. Since all data that this study deals with are time-series data, it is necessary to deal with a serial correlation problem. The serial correlation means that residuals in the regression are correlated with each other over time, which would make the estimators biased and non-consistent because the expectation value of residuals conditioning independent variables would be non-zero. The ARIMA model would fix the serial correlation problem incorporating autoregression term (AR) and moving average term (MA). In addition, if variables are non-stationary, taking differences is needed as well.

$ARIMA\ model\ (AR\ (1)\ \&\ MA\ (1)\ \&\ First\ Difference)$:

$$\Delta IBC_t = \beta_0 + \alpha_1 \Delta IBC_{t-1} + \beta_1 \Delta X_{1t} + \beta_2 \Delta X_{2t} + \ldots + e_t + \theta_1 e_{t-1}$$

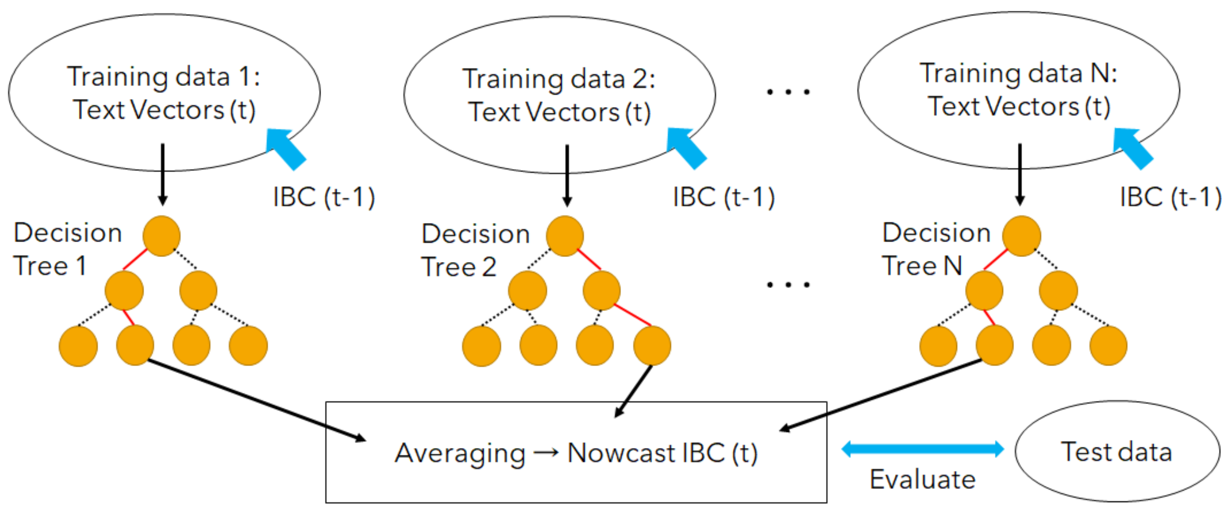
$$(where\ \ X = text\ vectors\ or\ other\ indicators)$$

Random Forest

The first machine learning model that this study uses for nowcasting economic conditions is the Random Forest method, which is the ensemble algorithm of the decision tree method (Figure 1). The decision tree method builds up the best tree diagram that regresses outputs well by minimizing the Residual Sum of Squares (RSS) among the results. The random forest model enhances the performance of the decision tree method by randomly implementing the tree build-up process again and again, and then taking the average of the prediction outputs, which would improve the performance of one decision tree method significantly. Although the one decision tree regression is usually not superior to the ordinary least square regression, the random forest regression would score a higher performance as the number of decision trees increases.

Though the Random Forest model is not the simple regression model, such as Ordinary Least Square (OLS), the serial correlation would be a problem as well. To deal with the serial correlation problem, this study incorporates the AR term into the Random Forest model. To be more specific, IBC_{t-1} is included in the independent variables in the Random Forest prediction, which would fix the serial correlation problem.

Figure 1 *Illustration of Random Forest model*



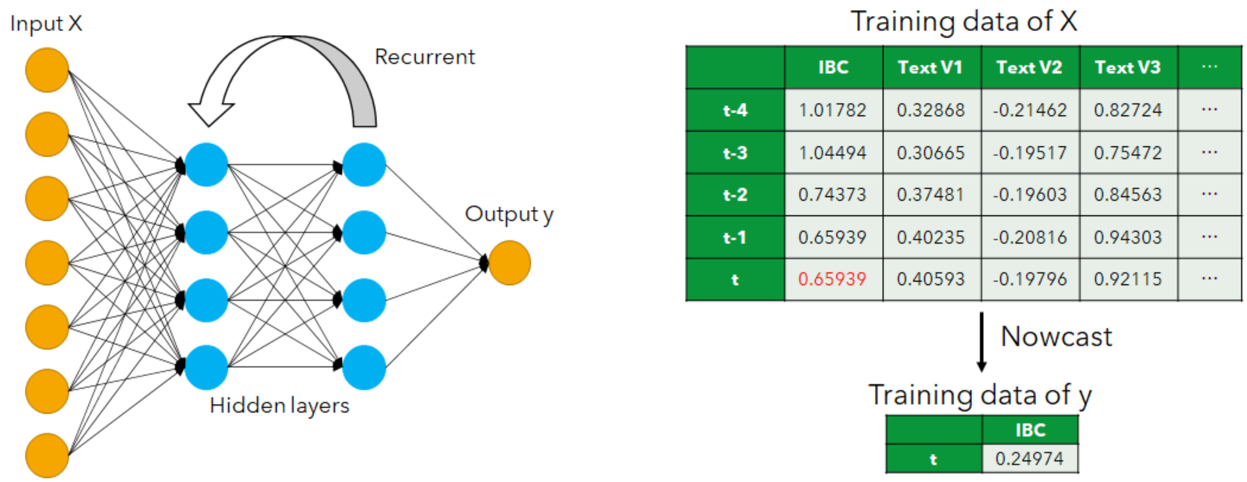
Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM)

The second machine learning method is the Neural Network model, which has a high performance for the prediction recently, especially when this model deals with the non-linear relationships in data. Since the relationship between big data and economic conditions could be complex, the neural network model would be appropriate for understanding such a relationship. First, the model assigns weights (the degree of the impact on the output) to each variable randomly. Then, it estimates the output through several hidden layers which are set up initially as parameters. After that, the algorithm calculates the performance of the model and then adjusts weights on variables slightly based on the previous results, which is called backpropagation. Similarly, the machine re-calculates the output and evaluates the performance again and again. Through this process, the performance of the prediction would improve significantly as the number of backpropagations increases.

Among the Neural Network models, this study selects the Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM) because this model can deal with the time-series data properly even in the machine learning methods (Figure 2, Left). Compared to the other neural network models, RNN-LSTM can judge which information is important for the prediction and then memorize that information for a longer time without unnecessary updates, and finally forget it when it becomes useless over time. This recurrent algorithm would fit the training a time-series data and improve the prediction for the future or current values.

To nowcast the IBC rather than forecast it, the training data of all predictors (X) in this model consists of not only the past IBC and the past other predictors, but also the current other predictors (Figure 2, Right). However, the algorithm produces an error when the current IBC in the training data is missing. Therefore, this model temporary uses the previous value of the IBC as the missing predictor. The training data of the target variable (y) is the current IBC.

Figure 2 *Illustration of RNN-LSTM*



	IBC	Text V1	Text V2	Text V3	···
t-4	1.01782	0.32868	-0.21462	0.82724	···
t-3	1.04494	0.30665	-0.19517	0.75472	···
t-2	0.74373	0.37481	-0.19603	0.84563	···
t-1	0.65939	0.40235	-0.20816	0.94303	···
t	0.65939	0.40593	-0.19796	0.92115	···

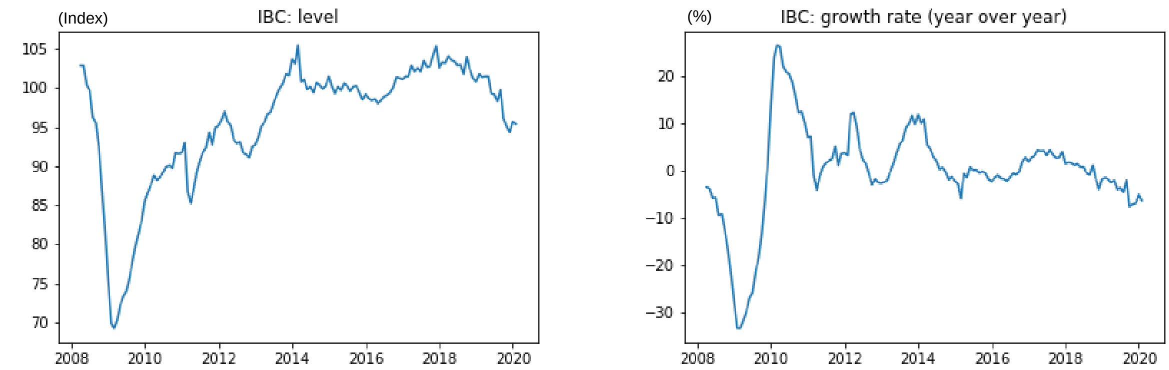
	IBC
t	0.24974

Data

Index of Business Conditions (IBC)

The target variable in this study is the Index of Business Conditions (IBC), which is generated and published by the Cabinet Office at the government of Japan. There are three types of indexes in the IBC, which are “leading index,” “coincident index,” and “lagging index.” This study uses the “coincident index” among other indexes because it represents current economic conditions rather than future or past. Since this index is available from 1985 on a monthly basis and is already seasonally adjusted to represent the economic trend. Either the level or growth rate (Year over Year) of IBC is used as a target variable depending on the model structure (Figure 3).

Figure 3 *Index of Business Conditions (level and growth rate)*



Economic Report

This study uses the Monthly Economic Report as the text data source. This report is issued by the Cabinet Office of Japan every month. The government economists monthly research the current economic conditions observing the available economic statistics as many as they can at each time. This report has a topic sentence that mentions briefly the current economic conditions in Japan (e.g. The Japanese economy is recovering at a moderate pace while weakness lasting longer mainly in exports.). It also mentions the details about each component of the Japanese economy by sector, such as private consumption, business investment, exports and imports, industrial production, corporate profits, employment situation, and consumer prices (Figure 4). Thus, this report would capture the various aspects of the Japanese economy. Since this report is usually released earlier than the IBC data publication, this study utilizes them for the prediction. This study uses the pdf version of the reports available from April 2008 (143 observations).

Figure 4 *Word Cloud of Monthly Economic Report (April 2008)*

(translation: economy, increase, decrease, production, labor, and so on)

Google Search

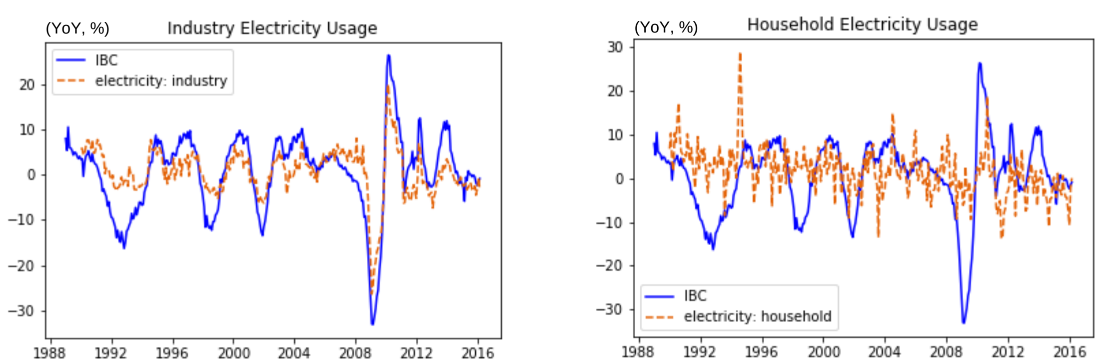
The second data source for nowcasting economic conditions is the number of Google searches over time. The number of web searches regarding a certain word would be correlated with economic conditions. For example, people might tend to query “unemployment” on the website when workers are losing their jobs along with an economic recession. In addition, the Google search data can be extracted at most on a daily basis, which is a very timely data source. However, since the daily data is difficult to deal with, this study aggregates it into monthly data. Using the Google API in the software Python, which is “pytrends,” the data can be extracted from January 2004 to the present (194 observations). Moreover, the seasonality on the search is eliminated using the other Python package, “statsmodels.”

Electricity Usage

The third data source is electricity usage. Since most economic activities in Japan use electricity as a primary energy, the volume of electricity used in a certain period should be correlated with economic conditions. This study focuses on 10 large regional electricity companies in Japan (Hokkaido, Tohoku, Tokyo, Hokuriku, Chubu, Kansai, Chugoku, Shikoku, Kyushu, and Okinawa Electric Power Company) because they provide most of the electric power in Japan. The Federation of Electric Power Companies (FEPC), consisted of these 10 companies, had provided the monthly usages of both industry and household electricity data separately from January 1989 to March 2016 (327 observations). The seasonality on the data is eliminated using the Python package, “statsmodels.” The correlation score between the growth rate (YoY) of IBC and industry electricity usage is 0.80, while the score with household electricity usage is only 0.19 (Figure 5). This is because the industry electricity is directly consumed for economic activities, while the household electricity is mainly used for daily activities in houses, such as using air conditioners, which would be affected by the fluctuations of weather conditions.

However, after March 2016, only total electricity usage that combines industry and household usage, has been offered by each electricity company. Although we can get the usage data on a daily basis, the ratio of industry and household is no longer available.

Figure 5 *Electricity Usage (growth rate) by March 2016: Industry and Household (YoY)*



Variables

Economic Report

The text data in economic reports have to be transformed into numeric values to work as predictors. First, the web scraping algorithm in software Python can extract the contents of each report from the website of Monthly Economic Reports, and then store them into the data-frames as time-series data. As the reports are uploaded as pdf files, it is needed to convert pdf files into text files using the Python package, “pdfminer.six.”

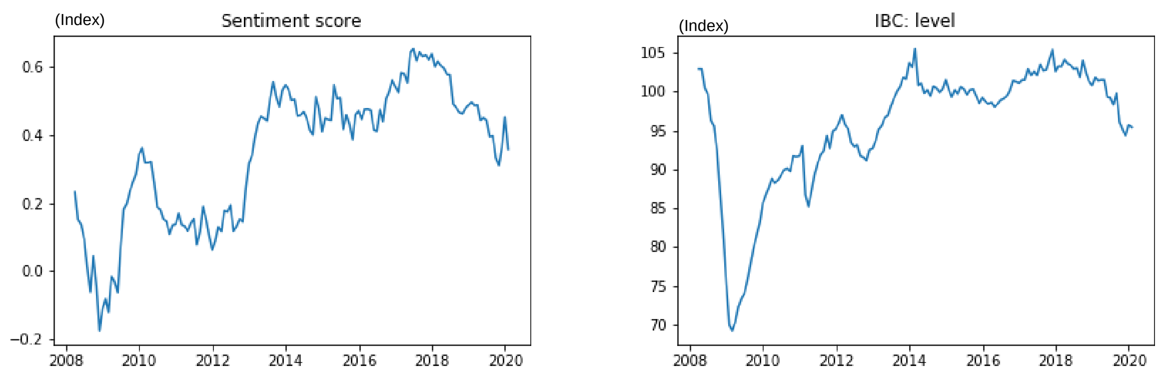
After the web scraping, the morphological analysis is required for the Japanese language as the preprocessing of the text data. This is because the Japanese language inherently does not have spaces between words in sentences and is therefore not ready for being transformed into numeric values. Using the Python package, “MeCab,” the morphological analysis function for Japanese language extracts only nouns and verbs from the text. This procedure would detect each meaningful word in sentences and also eliminate useless words, such as ‘on,’ ‘to,’ and ‘from’ at the same time. Since such stop words are not relevant to the current economic conditions, these words should be eliminated. These processes would be helpful for the machine to understand and classify the words in economic reports.

After preprocessing the dataset of text data over time, the NLP can transform the text data into vectors using the Tf-idf Vectorizer, which captures various features in the text as numeric terms. If the number of features in vectors has many noisy factors as predictors, the Principal Component Analysis (PCA) technique reduces the dimensions of vectors data by minimizing the variance within each generated component, to keep the information as much as it can keep. In this study, setting the n-gram from 1 to 3 words, the Tf-idf Vectorizer produces 15,142 text vectors from economic reports, and then the PCA reduced the dimensions of text vectors into 9. As a result, the total number of variables in text vectors is 1,287 (143 rows × 9 columns). After these processes, text vectors are ready to be independent variables for prediction models.

As an alternative, this study generates the sentiment score by simply counting negative and positive words in each report, using the Japanese Sentiment Dictionary (Kobayashi, 2005; Higashiyama, 2008). For example, if the negative word, such as 'weak,' appears in the text, the score counts as -1. Similarly, if the positive word, such as 'improve,' appears in the text, the score counts as +1. By summing up all scores and dividing it by the total number of counts, the normalized index would represent economic sentiments over time, which should be correlated with the economic conditions. Actually, the shape of the sentiment score over time is similar to the IBC and the correlation between them is 0.77 (Figure 6).

The independent variables from economic reports should have some correlations with the target variable. The relationship of each text vector on the IBC could be positive or negative because one vector might represent positive words, such as 'go up,' and another might represent negative words, such as 'go down,' in the economic reports.

Figure 6 *Comparison between Sentiment Score and IBC*

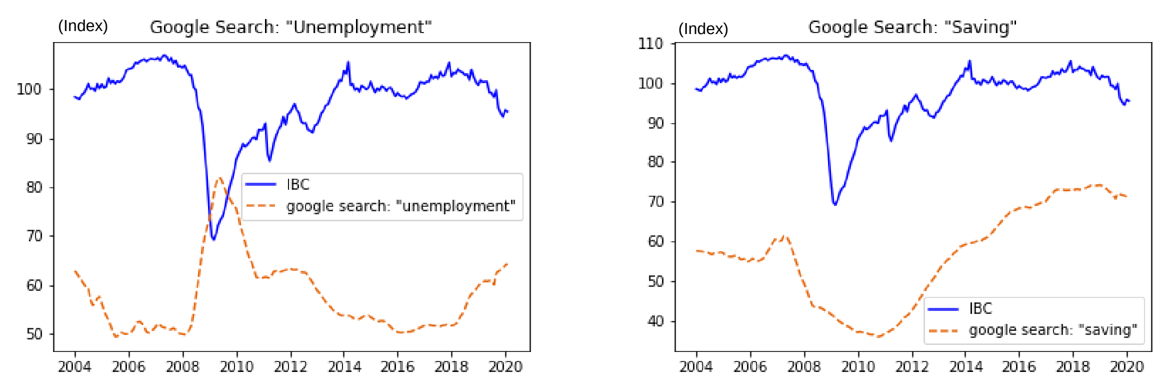


Google Search

Among many possible options for searching words that could be useful for predictors, this study selects "unemployment" and "saving" as a searching word in Google search engine. This is because these two words show relatively high correlations with the IBC, which are -0.90 for the "unemployment" and 0.65 for the "saving" respectively (Figure 7).

The index from "unemployment" moves as if it is an upside-down version of IBC. People may search "unemployment" on websites when they lose their jobs and want to get new ones. Thus, the coefficient of the index from "unemployment" on the IBC would be negative. On the other hand, the index from "saving" has a similar trend as the IBC because people may search "saving" on websites when they have enough money to save for the future. Thus, the coefficient of the index from "saving" on the IBC would be positive.

Figure 7 *Google Search: "unemployment" and "saving" (level)*

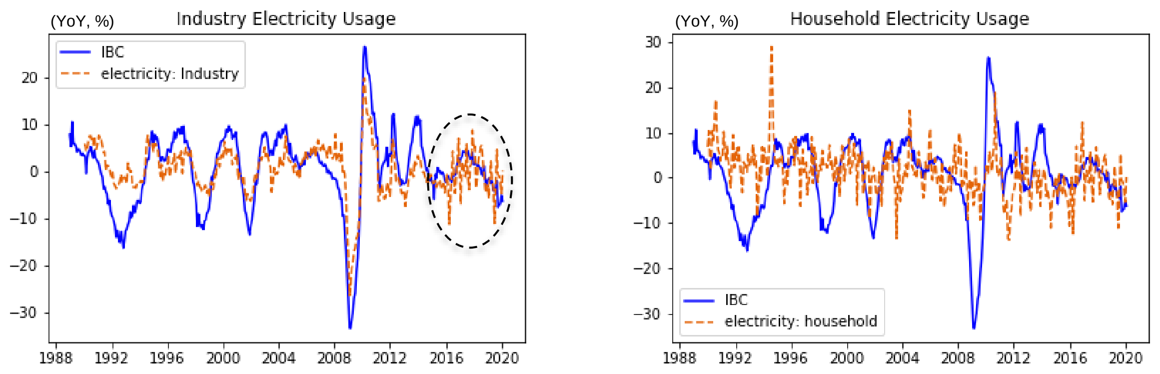


Electricity Usage

As for the data after March 2016, this study has to generate electricity usage data using the following procedure because the FEPC no longer provides the electricity data on their website. First, each generated electricity data has to be collected from each regional company's website and then aggregated into the national generated electricity. Second, as the generated electricity does not match electricity usage, it has to be adjusted by the linking coefficient between the supply side and the demand side in electricity. Third, this study divides the total electricity usage into industry and household ones based on the average ratio of industry or household usage to total usage in the past 4 years from 2012 to 2015.

Compared to the electricity data before March 2016, the usage of industrial electricity after March 2016 does not change similar to the IBC due to the limitation mentioned before (Figure 8). However, in general, electricity usage has a strong positive relationship with the IBC.

Figure 8 *Electricity Usage (growth rate): Industry and Household (YoY)*



Results

Using variables prepared so far, this chapter nowcasts the IBC based on the three methods, which are ARIMA, Random Forest, and RNN-LSTM. To train and evaluate these models, all data is split into a training dataset and test dataset. Since variables are all time-series data, it is not appropriate to randomly split data into a training dataset and test dataset. Thus, the first 80% of the data (e.g. Apr. 2008 - Sep. 2017) is defined as the training data and left 20% of the data (e.g. Oct. 2017 - Feb. 2020) is defined as the test data. Each model is trained by the training dataset based on each algorithm and evaluated by the test dataset based on R-squared metrics across datasets and methods.

The overview of the results is summarized in the table below (Table 1). As for the text data on economic reports, the text vectors and sentiment score nowcast the IBC with higher test score (0.83-0.89) using machine learning methods. As for Google search and electricity usage data, in the RNN-LSTM model, both data nowcasts the IBC with relatively high test scores (0.83-0.87). The electricity usage up to March 2016 performs the best test score (0.91) based on RNN-LSTM.

Table 1 *Overview of Results*

		ARIMA	***Random Forest***	***RNN-LSTM***
Economic Report	**Text Vectors**	0.36 (YoY)	0.83 (level)	0.89 (level)
	Sentiment Score	0.17 (level)	0.83 (level)	0.88 (level)
Google Search		-0.44 (level)	0.78 (level)	0.87 (level)
Electricity Usage	**Up to March 2016**	0.41 (YoY)	0.86 (level)	0.91 (level)
	After March 2016	-0.46 (YoY)	0.69 (level)	0.83 (level)

Note: values are the test scores measured by R-squared.

The rest of this chapter spells out the nowcasting process and the detailed results for each data source and methodology.

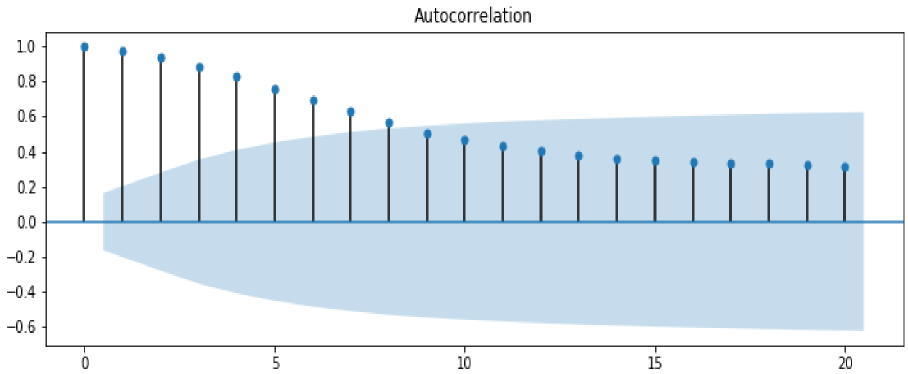
Text Vectors from Economic Report

1. Autoregressive Integrated Moving Average (ARIMA)

Before the actual prediction, it is worthwhile to examine the basic status of the time-series data. First, this study detects a clear autocorrelation in the IBC data, which means that the past

IBC data is correlated with the current IBC data (Figure 9). Thus, models have to deal with this sequential characteristic to nowcast the IBC and the result would be biased otherwise.

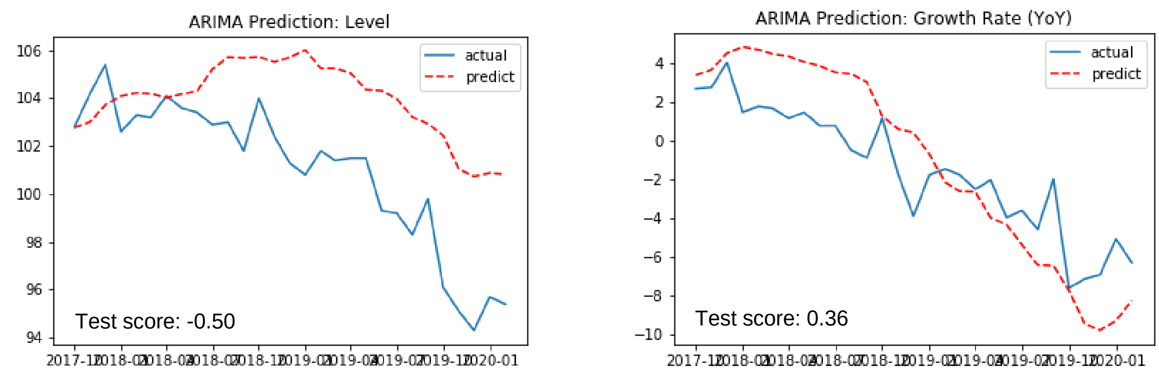
Figure 9 *Autocorrelation in IBC*



Second, the stationarity matters the way of ARIMA model. The unit-root test (Augmented Dickey-Fuller test) shows that both IBC and numeric vectors from economic reports would be non-stationary, which means that there would be a suspicious regression problem unless they are cointegrated. Although the cointegration test is also implemented for these data, the result shows that there would be no cointegration (The p-value of non-cointegration is 0.74). Thus, for these data, taking the first difference or using growth rate (YoY) in the IBC is necessary for the prediction in order to achieve non-stationarity and eliminate a suspicious regression problem.

By taking the first difference and adding AR (1) and MA (1), the ARIMA model predicts the level of IBC (Figure 10, Left). However, the test score is -0.50, which means that the prediction is meaningless. In addition to the level, the growth rate of IBC (YoY) is predicted by the ARIMA model (Figure 10, Right). As a result, the test score improves up to 0.36, but it is still quite low. Therefore, the ARIMA model does not work well for this prediction.

Figure 10 *ARIMA Predictions for Text Vectors*

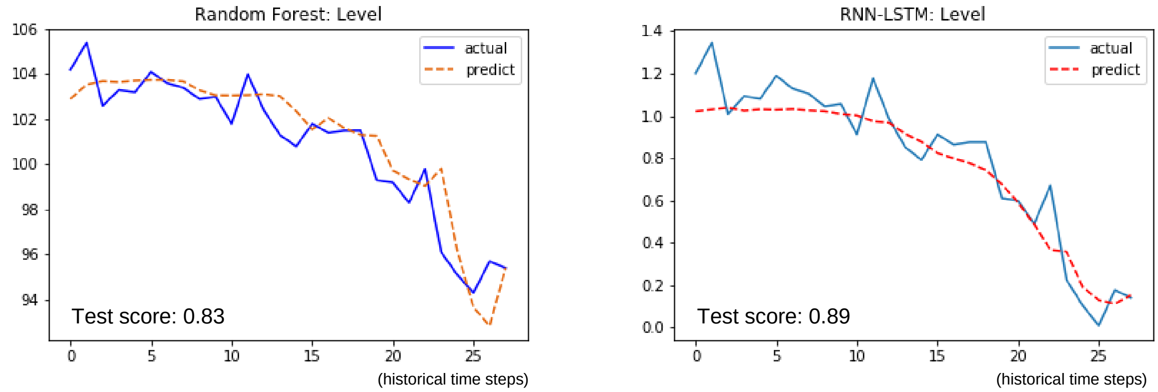


2. Machine Learning models: Random Forest and RNN-LSTM

Instead of using a simple ARIMA model, this study uses the Random Forest model as the next model. Incorporating AR (1) term into the Random Forest model and setting 1,000 as the number of random decision trees, the model predicts the level of IBC resulting in a higher test score, 0.83 (Figure 11, Left).

In addition, the other machine learning model, the RNN-LSTM predicts the level of IBC through 5 hidden layers with dropout functions. The hidden layers would improve understanding of the relationships between text vectors and the IBC. The dropout would prevent the overfitting problem by automatically drop some information on the training process (For the detailed setting of the RNN-LSTM model, see the appendix). The test score is 0.89 (Figure 11, Right), which is higher than the result of the Random Forest model.

Figure 11 *Random Forest and RNN-LSTM for Text Vectors*

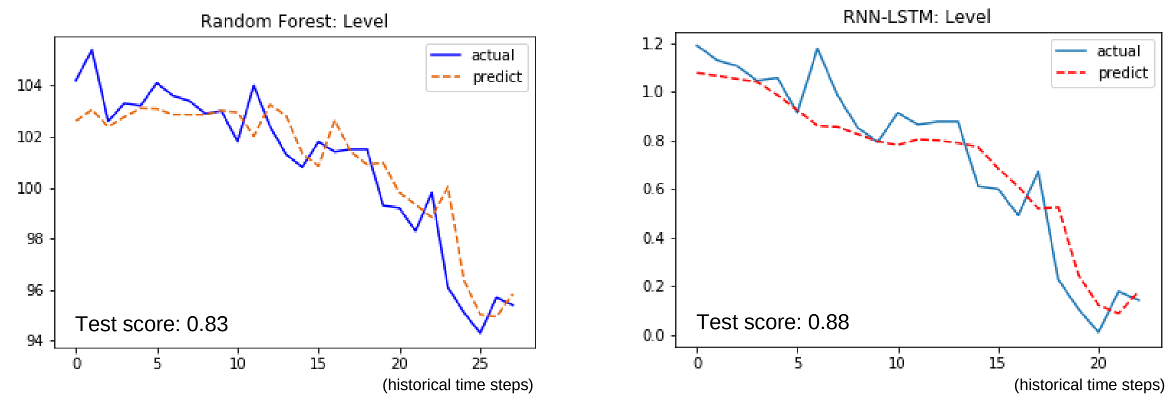


Sentiment Score from Economic Report

Although the sentiment score from economic reports is also non-stationary data according to the unit-root test, the cointegration test shows that the IBC and sentiment scores are cointegrated (The p-value of non-cointegration is 0.00). Thus, models do not need to take the first difference or use YoY to make predictions.

The ARIMA model with AR (1) and MA (1) results in a 0.17 test score. On the other hand, the Random Forest performs a 0.83 test score (Figure 12, Left). In addition, the RNN-LSTM outputs a 0.88 test score (Figure 12, Right). In conclusion, the test scores in both text vectors and sentiment scores are high when based on machine learning methods.

Figure 12 *Random Forest and RNN-LSTM for Sentiment Score*

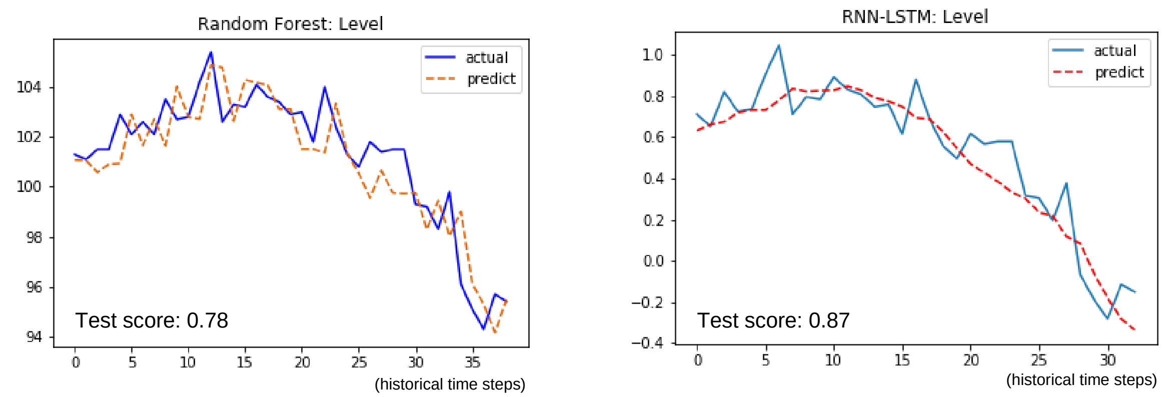


Index from Google Search

As a result of the unit-root test, the index from Google search is non-stationary data. However, the cointegration test shows that the IBC and the index of “unemployment” are cointegrated (The p-value of non-cointegration is 0.01), although the index of “saving” is not cointegrated with the IBC (The p-value of non-cointegration is 0.07). Thus, this study ends up using only “unemployment” of Google search data as a predictor.

While the ARIMA model with AR (1) and MA (1) results in an -0.44 test score, the Random Forest performs a 0.78 test score (Figure 13, Left) and RNN-LSTM outputs a 0.87 test score (Figure 13, Right). Therefore, the machine learning method is effective for Google search.

Figure 13 *Random Forest and RNN-LSTM for Google Search*

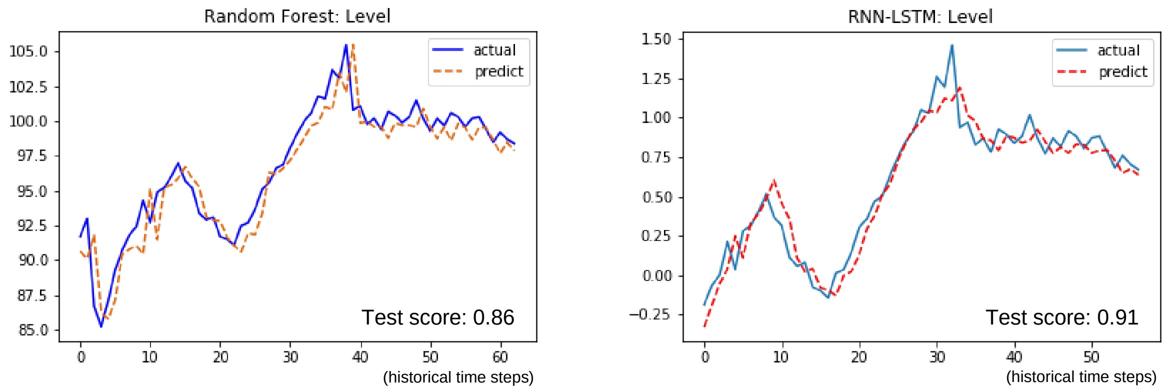


Electricity Usage up to March 2016

As the electricity-usage data is non-stationary and not cointegrated with the IBC according to the corresponding tests (The p-value of non-cointegration is 0.23), it is necessary to take the first difference for all variables or use a growth rate (YoY) as a target variable.

The ARIMA model with the first difference results in an -5.21 test score. For the growth rate (YoY), the ARIMA model performs a 0.41 test score. While the Random Forest performs a 0.86 test score (Figure 14, Left), RNN-LSTM outputs a 0.91 test score (Figure 14, Right), which is the highest among prediction models. If the separated industry and household electricity usages were available even after March 2016, the nowcasting of the IBC could have resulted in a similar performance to this score.

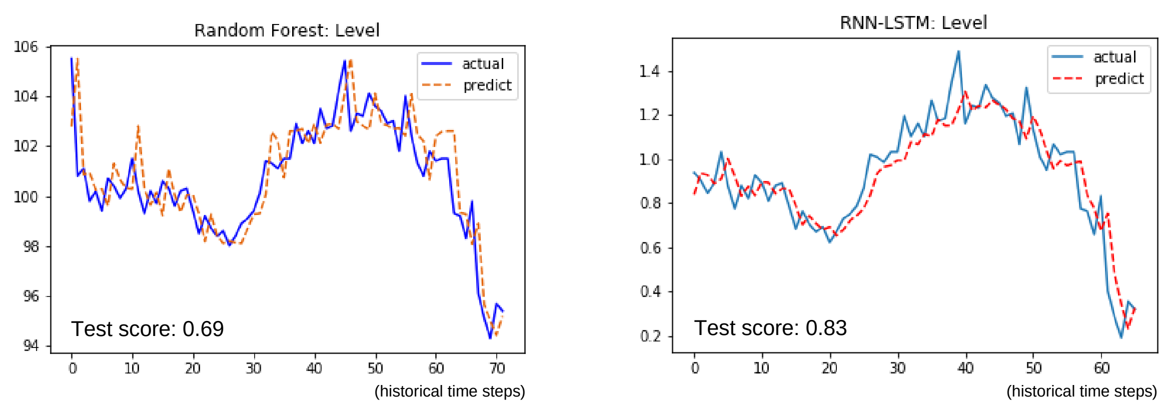
Figure 14 *Random Forest and RNN-LSTM for Electricity Usage (by March 2016)*



Electricity Usage after March 2016

While the ARIMA model taking the first difference results in an -0.68 test score, the ARIMA model to predict the growth rate (YoY) performs an -0.46 test score. The Random Forest model performs a 0.69 test score (Figure 15, Left). On the other hand, the RNN-LSTM outputs a 0.83 test score (Figure 15, Right). Although the score of RNN-LSTM is lower than the prediction for the data up to March 2016, it is still relatively high score.

Figure 15 *Random Forest and RNN-LSTM for Electricity Usages (all periods)*



Discussion

The results above show that the machine learning methods nowcast the IBC with higher test scores than the traditional ARIMA model. In addition, the RNN-LSTM models output the higher scores compared to the Random Forest models for all data sources. Also, reducing dimensions in text vectors, checking correlations of variables with the IBC, and then focusing on only meaningful predictors would contribute to the performances of the models.

Although the text data from the Monthly Economic Report predicts economic conditions with a high test score, it is not very timely data. Even though the government economists collect much data and conduct their own interviews to write this report, the information that they can utilize is limited and mostly relies on the official statistics published with lags. Thus, text data on the internet including Social Network Services (SNS), such as Twitter and Facebook data, could be another potential resource to represent economic conditions. The comments on SNS sometimes mention economic issues and these are usually available in a very timely manner. However, the users on SNS might not represent the entire population and might be tilted toward relatively younger users. Therefore, it is important to verify the population of the text data when using it as predictors for nowcasting of the entire nation's economy.

The specific words used for Google search data also leave room for discussion. This study selects "unemployment" and "saving" for the Google search words because these are well correlated with the IBC over time. However, there is no evidence that these words are the best choices among other searches. It would be more desirable to check the correlations across all possible search words in Google engine and systematically choose the best candidates for the prediction. Some web-search engine companies, such as Google or Yahoo!, would have such capabilities because they have the rich information of the query words in the system.

For electricity usage, the performance of the prediction would have increased if a separated data for industry and household electricity, which has been suspended after March 2016, could have been available as predictors. Now, electricity companies release the amount of generated electricity on a daily basis in terms of supply side. Since such a ratio does not have to be timely data, some information of the ratio should be opened for further analysis.

Conclusion

This study tries to nowcast economic conditions using economic reports, Google search, and electricity usage data. The text vectors and sentiment scores from economic reports nowcast the IBC with a high performance. In addition, Google search and electricity usage data nowcast the IBC with relatively high performance in a very timelier manner. In addiiton, the machine learning models can significantly improve the results for all cases compared to the traditional ARIMA model. Moreover, the RNN-LSTM performs noecasting more successfully than the Random Forest model for all cases.

Although there is still much room for improvement in the data and methods, the results in this study would be useful to fill the time gap in the assessment of economic conditions. Since the approaches in this study would be applicable for any other predictors in general, as more and more big data become available in the future, the coverage of application would be enlarged.

References

Bok, B., & Caratelli, D., & Giannone, D., & Sbordone, A., & Tambalotti, A. (2017). Macroeconomic nowcasting and forecasting with big data. *Federal Reserve Bank of New York Staff Reports, no. 830.*

Bragoli, D. (2017). Nowcasting the Japanese economy. *International Journal of Forecasting 33*, pp. 390-402.

Chikamatsu, K., & Hirakata, N., & Kido, Y., & Otaka, K. (2018). Nowcasting Japanese GDPs. *Bank of Japan Working Paper Series.*

Götz, T. B., & Knetsch, T. A. (2019). Google data in bridge equation models for German GDP. *Deutsche Bundesbank, Germany. International Journal of Forecasting 35.*

Higashiyama, M., & Inui, K., & Matsumoto, Y. (2008). Learning sentiment of nouns from selectional preferences of verbs and adjectives, *Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing*, pp.584-587.

Karim, B., & Olivier, D., & Laurent, F., & Bertrand, P. (2011). Monthly GDP forecasting using bridge models: Application for the French economy. *Bulletin of Economic Research. 64.*

Kobayashi, N., & Inui, K., & Matsumoto, Y., & Tateishi, K. (2005). Collecting evaluative expressions for opinion extraction, *Journal of Natural Language Processing 12*(3), 203-222.

Nitta, T. (2019). Timely assessment of regional economy using alternative data. *The Department of Economic Research, Daiwa Institute of Research.*

Suimon, Y., & Izumi, K., & Sakaji, H., & Shimada, T., & Matsushima, H. (2019). Real time economic indicator of manufacturing industry based on high frequency electricity demand data. *The Japanese Society for Artificial Intelligence. Vol 35.*

Urasawa, S. (2014). Real-time GDP forecasting for Japan: A dynamic factor model approach. *Journal of The Japanese and International Economies.*

Yamamoto, Y., & Matsuo, Y. (2016). Sentiment summarization of financial reports by LSTM RNN model with the Japan economic watcher survey data. *The 30th Annual Conference of the Japanese Society for Artificial Intelligence.*

Appendix

The settings of the RNN-LSTM models

		Past history	**Hidden layer**	**First node**	**Drop out**	**Learning rate**	**Epoch**
Economic Report	**Text Vectors**	1	5	128	Yes	0.001	10
	Sentiment Score	6	1	8	No	0.001	10
Google Search		6	1	8	No	0.0001	10
Electricity Usage	**Up to Mar. 2016**	6	1	4	No	0.001	10
	After Mar. 2016	6	1	4	No	0.0005	10

Past history: The number of past time steps that is used for training the model

Hidden layer: The number of LSTM layers that is used for training the model

First node: The number of nodes in the first LSTM layer

Dropout: The function that automatically drops some information on the training process

Learning rate: The rate of training the model

Epoch: The number of cycles that are used for training the full dataset