No. 2023-10

BOK Working Paper

Econometric Forecasting Using Ubiquitous News Text: Text-enhanced Factor Model

Beomseok Seo

2023. 6

BANK OF KOREA

Economic Research Institute
Bank of Korea

Publisher
Changyong Rhee
(Governor of Bank of Korea)

Editor
Taehyoung Cho
(Acting Director General of the Institute)

Requests for copies of publications, or for addition/changes to the mailing list, should be sent to:

Economic Research Institute
Bank of Korea
39 Namdaemunno Jung-Gu
Seoul, 110-794, Korea

E-mail: eso@bok.or.kr

Fax: 82-2-759-5410

This publication available on the BOK Economic Research Institute website (http://imer.bok.or.kr)

BOK Working Paper is occasionally published by the Economic Research Institute, Bank of Korea. This is circulated in order to stimulate discussion and comments. Articles include research achievement by the staff and visiting scholars, and selected works sponsored by the Institute.

BOK WP No.2023-10

Econometric Forecasting Using Ubiquitous News Text: Text-enhanced Factor Model

Beomseok Seo*

June, 2023

* Economist, Office of Economic Modeling and Policy Analysis, Bank of Korea, Tel: +82-2-759-4248, E-mail: bsseo@bok.or.kr, Web: https://seo-beomseok.github.io

This work is a substantially revised version of the author's working paper (Seo, 2022) written in Korean. I would like to thank Byoungho Bae, Hyoungseok Oh, Kyounghoon Park, Minsu Chung, and the surveillance forum participants at the Bank of Korea for their helpful comments.

Contents

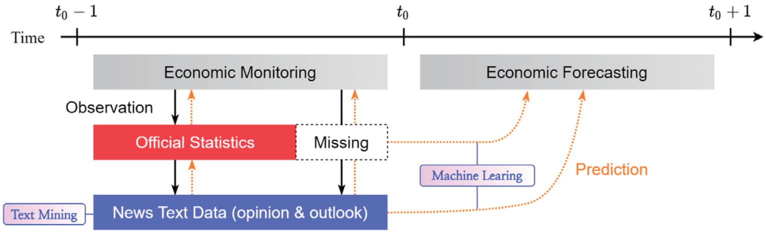
Econometric Forecasting Using Ubiquitous News Text: Text-enhanced Factor Model

The use of news text as a novel source for econometric forecasting is gaining increasing attention. This paper revisited the way of incorporating narrative information into econometric forecasting by effectively quantifying sector-specific textual information without requiring training data. We exploit *Theme Frequency Indices*(TFI) utilizing domain-specific subject-predicate patterns to gauge public perception of the economy. TFIs of 15 sectors, including production, inflation, employment, capital investment, stock and house prices, and others, were examined and integrated into *Text-enhanced Factor Model*(TFM) using latent factor structures. Empirical analysis, based on over 18 million news articles in Korea, reveals that TFM improves the accuracy of near-term GDP forecasts, demonstrating simple text-mining techniques along with domain knowledge are capable of leveraging qualitative information from news without costly training. The proposed method is applicable to a wide range of subjects for measuring narrative information of the economy, offering a rapid and cost-effective approach.

Keywords: dynamic factor model; text mining; machine learning; economic forecasting; nowcasting

JEL Classification: C45, C53, C55, C82

Figure 1. Schematic diagram for economic forecasting with news text data.



I. Introduction

Over the last decade, nowcasting has been popular among central banks and economists due to the increasing accessibility to big data. Many recent works point out that text data is crucial to improve the prediction accuracy of nowcasting models (Thorsrud, 2020; Babii et al., 2022; Kalamara et al., 2022; Barbaglia et al., 2022). Various approaches have been studied to incorporate narrative information from text into quantitative research. However, most of the works rely on topic modeling, i.e., text clustering (Thorsrud, 2020; Babii et al., 2022; Ferreira et al., 2021), using unsupervised methods like latent Dirichlet allocation (LDA), or a few sector text indices (Kalamara et al., 2022) such as news sentiment index (NSI) (Shapiro et al., 2022; Seo et al., 2022) and economic policy uncertainty (EPU) (Baker et al., 2016; Lee et al., 2020). Despite its popularity, however, unsupervised approaches have inherent drawbacks. One major challenge is that it can be difficult to verify the results as each cluster may not necessarily connect to a specific economic topic[1]. On the other hand, supervised methods using machine learning (Seo et al., 2022) can be expensive for developing even a single index due to the need for labeled training data.

To address these challenges, we propose a novel text-mining approach to extract quantitative information from news text and incorporate it into an econometric forecasting model. We generate text-based economic indicators, called *Theme Frequency Indices* (TFI), for 15 sectors that are important for economic

1) Additionally, for the unsupervised method, information that appears with low frequency in the text is often ignored as small clusters tend to merge with larger clusters, resulting in clustering outcomes being vulnerable to the addition or removal of a few data points (Li et al., 2019).

forecasts, including macroeconomic variables such as production, inflation, employment, capital investment, and stock and house prices, as well as industry-related microeconomic variables such as semiconductor, and wholesale & retail. We then propose a forecasting model, *Text-enhanced Factor Model* (TFM), that employs the textual information to forecast economic variables.

Specifically, our proposed method utilizes a lexical approach to extract quantified information from news text, based on predefined domain-specific subject-predicate patterns of the 15 sectors. This approach does not require any training data for computation and can be applied in a timely manner. The method generates essential sector indices for the economy, hence it is not only practical but also can be verified by comparing the results to official statistics. Our empirical analysis has shown a strong correlation between the text indices and the corresponding official statistics for up to 10 months, where the text indices lead the official statistics. Additionally, the proposed model, which incorporates textual information, has also been demonstrated to improve the generalized testing accuracy for near-term forecasting compared to the method that uses only official statistics. Through texts, the model leverages qualitative information as well as takes advantage of using the latest information ubiquitous in public when official statistics are not available.

Textual information is particularly valuable for economic forecasting as it addresses the challenge of limited availability of data for statistical models. Currently, economic judgment heavily relies on expert opinions, who qualitatively evaluate various information (Mæhle et al., 2021). The limitation of statistical models is primarily due to the difficulty in immediately incorporating sufficient information. Firstly, outside of the financial sector, there are few high-frequency economic indicators available on a daily or weekly basis. Additionally, official statistics used as quantitative indicators often have a time lag in publication, making them difficult to use in rapidly changing economic conditions. Moreover, integrating qualitative information into quantitative models poses further difficulties as variable importance varies depending on the economic situation and cannot be easily quantified.

Textual information, however, can address the above-mentioned limitations, as vast amounts of qualitative information are conveyed in a swift manner through

Figure 2. The relationship between news and economic perceptions of the public.

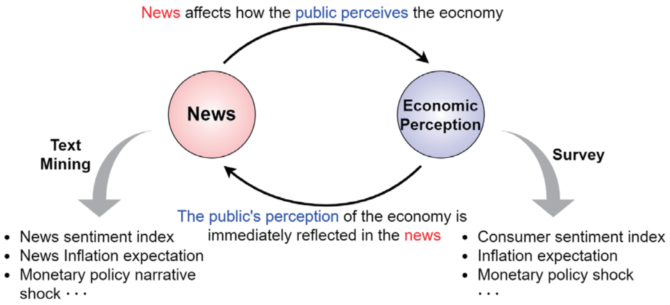


Table 1. The characteristics of news text data.

	Information	Advantages & Disadvantages
Survey Statistics	• Information on the topic	• Managed by a quality control • Cost and time-consuming • Cannot be surveyed frequently
News Text Information	• Information on the topic • Experts' opinion and expectation • Interpretation by journalists • Degree of attention from the public	• Big data (volume, velocity, & variety) • Include qualitative information • May include noise

Notes: Text information obtained from the news is comparable to the information based on survey statistics.

texts. News text, in particular, has the characteristic of delivering a wide variety and volume of information quickly. Therefore, it is expected that faster economic judgment is possible when appropriately incorporating the news information into forecasting models (Babii et al., 2022; Bybee et al., 2021; Thorsrud, 2020). News articles not only present factual information, quoting statistics but also convey expert opinions and insights through interviews and journalists' interpretations of the economy. Additionally, news stories that capture public interest are more likely to be widely distributed, reflecting the evolving interests of the public over time. This stands in contrast to official statistics that rely on fixed metrics.

The information gleaned from news texts can be likened to that obtained through surveys. The news texts have a two-fold effect on the public's perception of the economy: news journalists capture the public's perception of the economy while their reporting also impacts it. Surveys, an experimental study method, are a reliable method for gouging public opinion, managed under stringent quality control. However, they are costly, time-consuming, and not frequently feasible.

In contrast, text mining, an observational study method, can extract similar information as surveys with significantly lower costs and time investment at any time, given that news texts are associated with public economic perceptions.

Therefore, our proposed method has the potential to contribute to several areas of future research. Firstly, the proposed text-mining technique is applicable to a wide range of subjects to measure the narrative information in a rapid and cost-effective manner. TFIs can complement existing survey indices and serve as supplementary indices themselves based on observational studies. Secondly, the proposed framework for forecasting near-term gross domestic product(GDP) provides an approach to develop a real-time statistical model to leverage timely news information into econometric prediction. This framework, automated through a web-scraping program and text-mining algorithm, can be effectively utilized to combine narrative information with existing sources, enabling timely predictions.

The remainder of this paper is organized as follows: Section 2 provides an overview of relevant literature. Following Section 3 presents the proposed method in detail, covering data collection, the text-mining method used to create text indicators, and the text-enhanced factor model. In Section 4, we analyze the empirical results of the proposed indices and the model incorporating text information. Finally, Section 5 summarizes the implications of this paper and provides avenues for future research.

II. Related Works

Economic analysis using text has been actively conducted since the mid-2010s, utilizing sources such as news texts (Baker et al., 2016; Lee et al., 2019; Shapiro et al., 2022; Seo et al., 2022) that reflect press opinion, social media and search information (Sun et al., 2016) that communicates public view, and corporate accounting and evaluation reports (Lewis and Young, 2019; Seo, 2023) for industrial information. Most studies analyze texts using sentiment analysis and topic modeling based on a lexical approach or statistical model. In recent years, there are research utilizing more complex models like neural networks (Hájek and Olej, 2013; Seo et al., 2022). Text is unstructured data, and there is virtually no limit to the range of information it can convey. Therefore, existing studies have analyzed text to predict market prices such as inflation and stock prices (Kalamara et al., 2022; Li et al., 2020), identify new economic information such as crisis indicators and uncertainty indicators (Li et al., 2009; Baker et al., 2016; Caldara and Iacoviello, 2022), detect fraud (Humpherys et al., 2011), corporate sustainability classification (Te Liew et al., 2014), credit scoring (Yap et al., 2011), and for various other purposes.

For Korean text, the Bank of Korea is leading the use of text data for economic analysis. Lee et al. (2020) calculated the monetary policy surprise index through news text analysis, and Seo et al. (2022) developed a news sentiment index of Korea using a state-of-the-art transformer model. Besides news data, various text analyses, such as using analyst reports (Seo, 2023), are also being attempted in Korea.

Certainly, both new data and models are crucial for generating unprecedented insights in economic forecasting. Alongside new data such as text, recent advancements in statistical models and machine learning techniques have played a pivotal role in overcoming various challenges, such as enabling the utilization of ultra-high-dimensional data in forecasting.

Dynamic Factor Model (DFM) (Bańbura et al., 2010, 2013; Bok et al., 2018) has been gaining growing popularity for nowcasting, as it can handle more variables than samples by utilizing unsupervised factor analysis. Using the factor analysis and EM algorithm, DFM can predict a large number of variables simultaneously, and incorporate mixed frequency data (Mariano and Murasawa, 2010).

The real-time prediction has been explored using a range of machine-learning models as well, including kernel quantile regression for private consumption (Shin and Seo, 2022), k-nearest neighbor regression and support vector regression (SVR) for GDP (Richardson et al., 2021), and neural networks for global merchandise trade exports (Hopp, 2022).

III. Data and Methodology

1. Preparing Texts

Using web-scraping techniques, we collected economic news articles posted on the internet news portal sites in Korea after January 2005. Our web-scraping method automatically downloads publicly available news articles and constructs a news database every morning. The method was built using *Python* programming language to automate the entire process of collecting news articles and generating TFI textual information without human intervention. The collected data consists of approximately 4,000 articles per weekday, which corresponds to about 1 million articles per year (in total, 18,343 thousand articles between January 2005 and December 2022) from 79 media companies. Overall, the dataset encompasses about 207 million sentences and 3.5 billion words.[2)]

2. Topic Selection

In order to extract information from text effectively, it is crucial to define specific topics that need to be extracted and select an appropriate extraction method. If the topic is too general, the extracted information may lose its value, while if it is too specific, the information may not be present in the text.

We selected text topics across 15 sectors, covering major macroeconomic variables and industry-related micro variables (as shown in Table 3). The selection of these topics was based on their predictive importance for the GDP of Korea, with the understanding that news information can be collected at any time and used as a supplementary source when official statistics are not available. In addition to the 15 sector text indicators, we also utilized the news sentiment index (NSI) of Korea (Seo et al., 2022) and the economic policy uncertainty (EPU) index. The NSI was obtained from the Bank of Korea's Economic Statistics System

2) The configuration of the database is as follows.

Provider	Num. of articles	Examples
Economic newspapers	10.8m (59.0%)	Money Today, Asia Economics, EDAILY, etc.
News agencies	4.8m (26.1%)	Yonhap News, Newsis, News1, etc.
Regular news	2.5m (13.8%)	Segye Ilbo, Donga Ilbo, Kyunghyang Shinmun, etc.
Others	0.2m (1.1%)	Shindonga, Jungang Sunday, etc.

Figure 3. Information extraction and synthesis procedure from news text data.

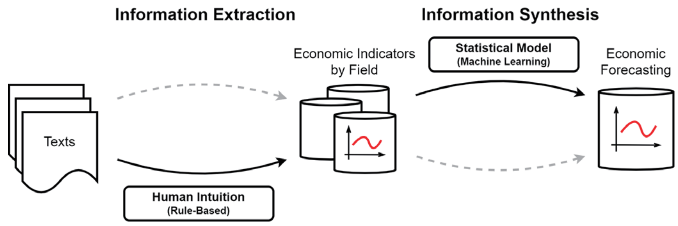


Table 2. Two main approaches for information extraction and synthesis.

Statistical Models (Machine Learning)	Human Intuition (Rule-Based)
• Algorithm finds patterns from data.	• Human defines patterns based on intuition.
• Statistical models are better at finding relationships between given multiple variables within a finite dataset.	• Human intuition can excel at identifying topics that are indirectly connected with the help of prior knowledge.

(ECOS), while the monthly EPU indices were calculated by us using the same news database used for the 15 text indices, following the methods of Baker et al. (2016) and Lee et al. (2020).

3. Choice of Information Extraction Method

Once the topic has been selected, an appropriate extraction method must be chosen. There are two types of text extraction methods: stochastic approaches that use statistical models (such as machine learning), and rule-based approaches that use human-defined extraction conditions. The choice of method depends on the nature of the source of the text data and the topic of the information being extracted from the text.

In general, statistical models tend to have better accuracy than rule-based methods when dealing with unstructured text data sources and abstract topics. This is because there are countless expressions people can use to describe an abstract topic in unstructured texts, as found, for example, in capturing economic sentiment from colloquial social media texts. In contrast, if the texts are relatively formal, such as news texts, and the topic of interest is specific, rule-based methods can offer competitive accuracy. For instance, describing the direction of inflation is relatively easier to define narrative patterns because there are formalized expressions frequently used by news journalists. Inflation is often associated with verbs like rise, increase, surge, spike, escalate, and climb. These patterns are even more apparent in the Korean language, which heavily relies on verbs originating from Chinese characters for formal writing.

More importantly, adopting the rule-based method eliminates the need to construct expensive labeled data and requires minimal preprocessing steps, resulting in computationally efficient text analysis. In general, statistical models require text standardization through tokenization, normalization, removal of stopwords, and integer embedding. This is because statistical models require input of all text characters and the transformation of text tokens to standard forms to recognize meaning even when the words have different conjugations. In contrast, by using only the root part of words[3)], the rule-based model can quickly check for the presence of the selected patterns without the need for pre-processing steps. As a result, we analyze the news text by applying human-defined rules that identify the subject and predicate components of word patterns.

4. News Sentence-Based Theme Frequency Index

News sentence-based *Theme Frequency Index* (TFI) is defined by the relative frequency of a certain topic appearing in news articles. TFI is measured based on sentences rather than articles. That is, if an article includes a sentence matching a theme, the article is added to the count to compute the frequency. Let $\Omega_t = \{A_1^{(t)}, \cdots, A_{N_t}^{(t)}\}$ be the set of N_t news articles at time t for articles $A_i^{(t)}$, $i = 1, \cdots, N_t$, and let each news article $A_i^{(t)} = \{S_{i1}^{(t)}, \cdots, S_{iM_i}^{(t)}\}$ be the set of sentences, $S_{im}^{(t)}$, $m = 1, \cdots, M_i$, for M_i being the length of sentence in i-th article. Then, for the word groups for a certain topic, $W^{[k]} = \{w_1^{[k]}, \cdots, w_{L_k}^{[k]}\}$, $k = 1, \cdots, K$, and the topic words

3) In Korean, the root part of a predicate typically maintains its fixed form as a stem and is concatenated with various endings for conjugation. This is especially noticeable in formal writing, where words originating from Chinese characters are more commonly used. Similarly, in English, the same approach can be taken by using the non-changing parts of a word for conjugation and considering additional cases for short words.

belonging to the group k, $w_l^{[k]}$, $l = 1, \cdots, L_k$, TFI is computed as follows.

$$\begin{aligned}
C_{im}^{(t)} &= \prod_{k=1}^{K} \bigvee_{l=1}^{L_k} I_{S_{im}^{(t)}}(w_l^{[k]}), && (1)\\
\hat{A}_i^{(t)} &= \bigvee_{m=1}^{M_i} C_{im}^{(t)}, && (2)\\
TFI_t &= \frac{\sum_{i=1}^{N_t}, \hat{A}_i^{(t)}}{N_t} && (3)
\end{aligned}$$

where $I_S(w) = \begin{cases} 1, & \text{if } w \in S \\ 0, & \text{o.w.} \end{cases}$ is an indicator function, and $\bigvee_{m=1}^{M} C_m = max(C_1, \cdots, C_M)$ is an elementwise max function.

When the topic consists of both positive and negative opinions, it is better to compute them separately and generate the TFI by subtracting TFI computed with negative words from TFI with positive words.

$$TFI_t = TFI_t^{(pos)} - TFI_t^{(neg)} \tag{4}$$

Table 3 describes the subject-predicate word groups used for each of the 15 sectors. Initially, these word groups were selected based on expressions commonly used in news articles to describe the economic direction of each topic. Then, in order to decide whether to use words without considering direction or with direction, the corresponding TFIs were computed using the chosen word groups and compared to official statistics for each sector (as presented in Table 5 and Figure 5). The final word groups were chosen by examining the Pearson's correlation between TFIs and official statistics. For House Construction and Government Expenditure, word groups without direction were used while directional word groups were used for the other sectors. This result is reasonable because news articles typically report the start of house construction and government expenditure when they occur. In contrast, for the other topics, the direction of change is also mainly covered by news articles rather than just reporting the occurrence.

Table 3. Subjective-predicate patterns used to compute 15 sector text indices.

<table>
<tr><th colspan="2">Topic</th><th>Topic Words</th></tr>
<tr><td rowspan="4">Industry</td><td>Product</td><td>|(product 생산), (rise 상승, increase 늘, skyrocket 급등, grow 증가, improve 개선, accelerate 가속)| − |(product 생산), (fall 하락, decrease 줄, drop 급락, shrink 감소, worsen 악화, slowdown 둔화)|</td></tr>
<tr><td>Shipbuilding</td><td>|(ship 선박), (order 수주), (rise 상승, increase 늘, skyrocket 급등, grow 증가, improve 개선) | − |(ship 선박), (order 수주), (fall 하락, decrease 줄, drop 급락, shrink 감소, worsen 악화) |</td></tr>
<tr><td>Automotive</td><td>|(car 자동차, automotive 승용차), (rise 상승, skyrocket 급등, grow 증가, improve 개선, accelerate 가속)| − |(car 자동차, automotive 승용차), (fall 하락, drop 급락, shrink 감소, worsen 악화, slowdown 둔화) |</td></tr>
<tr><td>Semiconductor</td><td>|(semiconductor 반도체), (rise 상승, skyrocket 급등, grow 증가, improve 개선, accelerate 가속)| − |(semiconductor 반도체), (fall 하락, drop 급락, shrink 감소, worsen 악화, slowdown 둔화) |</td></tr>
<tr><td colspan="2">Facilities Investment</td><td>|(facility investment 설비투자, r&d R&D), (rise 상승, skyrocket 급등, grow 증가, improve 개선, accelerate 가속, increase 늘, expand 확대)| − |(facility investment 설비투자, r&d R&D), (fall 하락, drop 급락, shrink 감소, worsen 악화, slowdown 둔화, decrease 줄, reduce 감축) |</td></tr>
<tr><td colspan="2">House Construction</td><td>|(house 주택, apartment 아파트), (construct 건축, build 건설, construction 착공, construction 시공) |</td></tr>
<tr><td rowspan="3">Employ-ment</td><td>Unemployment</td><td>|(unemployment 실업), (rise 상승, increase 늘, grow 증가, worsen 악화)| − |(unemployment 실업), (fall 하락, decrease 줄, shrink 감소, improve 개선) |</td></tr>
<tr><td>Recruitment</td><td>|(recruitment 채용, hiring 고용), (rise 상승, increase 늘, grow 증가, improve 개선)| − |(recruitment채용, hiring 고용), (fall 하락, decrease 줄, shrink 감소, worsen 악화) |</td></tr>
<tr><td>Job Search</td><td>|(employment 취업, job search 구직), (rise 상승, increase 늘, grow 증가)| − |(employment 취업, job search 구직), (fall 하락, decrease 줄, shrink 감소) |</td></tr>
<tr><td colspan="2">Wholesale and Retail</td><td>|(wholesale 도매, retail 소매, wholesale and retail 도소매), (rise 상승, skyrocket 급등, grow 증가, improve 개선, accelerate 가속, increase 늘)| − |(wholesale 도매, retail 소매, wholesale and retail 도소매), (fall 하락, drop 급락, shrink 감소, worsen 악화, slowdown 둔화, decrease 줄) |</td></tr>
<tr><td colspan="2">Government Expenditure</td><td>|(government 정부), (support 지원, subsidy 보조, expenditure 지출) |</td></tr>
<tr><td colspan="2">Inflation Outlook</td><td>|(inflation 물가), (forecast 전망, predict 예측, expect 예상), (rise 상승, skyrocket 급등, go up 올라, high 높)| − |(inflation 물가), (forecast 전망, predict 예측, expect 예상), (fall 하락, drop 급락, go down 내려, low 낮) |</td></tr>
<tr><td colspan="2">Stock Price Outlook</td><td>|(kospi 코스피, kosdaq 코스닥, stock price 주가, stock 주식), (forecast 전망, predict 예측), (rise 상승, grow 증가, improve 개선, increase 늘, high 높)| − |(kospi 코스피, kosdaq 코스닥, stock price 주가, stock 주식), (forecast 전망, predict 예측), (fall 하락, shrink 감소, worsen 악화, decrease 줄, low 낮) |</td></tr>
<tr><td colspan="2">House Price Outlook</td><td>|(house 주택, apartment 아파트), (price 가격, selling price 매매가, lease price 전세가, presale price 분양가), (forecast 전망, predict 예측), (rise 상승, skyrocket 급등, expand 확대, improve 개선, accelerate 가속, high 높)| − |(house 주택, apartment 아파트), (price가격, selling price 매매가, lease price 전세가, presale price 분양가), (forecast 전망, predict 예측), (fall 하락, drop 급락, reduce 축소, worsen 악화, slowdown 둔화, low 낮) |</td></tr>
<tr><td colspan="2">World Trade</td><td>|(global 글로벌, world 세계), (trade 교역, trade 무역, export 수출, import 수입), (rise 상승, skyrocket 급등, grow 증가, improve 개선, accelerate 가속, increase 늘, expand 확대)| − |(global 글로벌, world 세계), (trade 교역, trade 무역, export 수출, import 수입), (fall 하락, drop 급락, shrink 감소, worsen 악화, slowdown 둔화, decrease 줄, reduce 축소)</td></tr>
</table>

Notes: To apply this method to English, it is recommended to use different groups of words, as the English words used are direct translations of the corresponding Korean words. The elements in $|\cdot|$, denoted by a $(\cdot)$, indicate each word group, $W^{[k]}$, for $k = 1, ..., K$, and those in $(\cdot)$ are the topic words, $w_l^{[k]}$, for $l = 1, \cdots, L_k$. The word groups before and after − indicate the keywords patterns for $TFI^{(pos)}$ and $TFI^{(neg)}$ respectively.

5. Forecasting Models Utilizing Textual Information

5.1. Text-Enhanced Factor Model

We utilize the Dynamic Factor Model (DFM) structure, which is capable of handling mixed frequency data, and incorporate TFIs to each factor of DFM that corresponds to macroeconomic segments such as production, the external market, investment, labor market, and others, as shown in Table 4. This setup enables us to utilize textual information at the ragged edge point, where official statistics are not yet available for each sector.

For q latent factors at time t, f_t, and p-dimensional vector of the observed variables, X_t, the *Text-enhanced Factor Model* (TFM) is defined by the usual dynamic factor model with two linear functions.

$$X_t = \Lambda f_t + \varepsilon_t \tag{5}$$

$$f_t = \psi(L) f_{t-1} + \eta_t \tag{6}$$

$$\varepsilon_{i,t} = \rho_i e_{i,t-1} + e_{i,t} \tag{7}$$

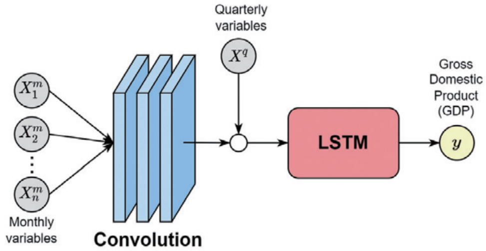
$$e_{i,t} \sim N(0, \sigma_i^2) \tag{8}$$

$$\eta_t \sim N(0, \Sigma) \tag{9}$$

where L is lag operator, and $\Lambda \in \mathbb{R}^{p\times q}$ is the matrix of factor loadings, and $\psi(L) \in \mathbb{R}^{q\times q}$ is r-th order polynomial coefficients, and $\varepsilon_t \in \mathbb{R}^p$ is idiosyncratic disturbance with $\rho_i \in \mathbb{R}$ and $\sigma_i^2 \in \mathbb{R}$ for $i = 1, \cdots, p$, and $\eta_t \in \mathbb{R}^q$ is factor innovation with $\Sigma \in \mathbb{R}^{q\times q}$.

In this research, we consider two overall factors combining all variables and 13 sector-specific factors composed of corresponding variables (see Table 4). A total of 83 variables are utilized, comprising 69 monthly variables and 14 quarterly variables. These include 52 official statistics, 14 financial indices, and 17 text indices that comprise 15 TFIs, NSI, and EPU. The financial and textual indices have no lags and are readily available, while the availability of official indices varies depending on the forecasted time. To capture the characteristics of the variables, we used year-over-year and month-over-month growth rates. For variables available in both seasonally adjusted (SA) and not-seasonally adjusted

Figure 4. CRNN structure designed for mixed frequency data.



(NSA) forms, we used month-over-month transformation for SA variables and year-over-year transformation for NSA variables. For variables released without seasonal adjustment, we applied X-13 ARIMA SEATS from Census Bureau(Sax and Eddelbuettel, 2018) and transformed them into month-over-month growth rates before using them. Consistent with previous studies, we employed an AR(4) structure for the two overall factors and an AR(1) structure for the other 13 sector-specific factors. For the observed variables, we used AR(1) structure for the idiosyncratic disturbance.

To train TFM, it is crucial to organize the training data in a proper way by utilizing the available information, using vintage data, which contains statistics with different release cycles and lags. To resolve this mixed frequency and ragged edge problems in the training of TFM, the missing values are imputed using the Kalman filter(Seabold and Perktold, 2010). Please refer to Mariano and Murasawa (2010) for more details.

5.2. Convolutional Recurrent Neural Networks (CRNN) with TFIs

It is important to choose an appropriate forecasting model that takes into account the characteristics of text data. Text indicators, which quantify various nuances, may contain more noise than official statistics and may be difficult to interpret for trend and periodicity. Although they are quantitative measures of media interest and do not exhibit obvious non-stationarity, such as upward trends (see Figure 5), they show high volatility and it is challenging to distinguish noise from economic signals. Therefore, using a linear model may result in increased prediction errors, while using a non-linear model may lead to overfitting errors by fitting noise. In light of these challenges, this study examines the predictive

Table 4. All variables and factors for Text-enhanced Factor Model.

	External	Production	Labor	Consumption	Facility Inv.	Construction	Government	Exports	Imports	Prices	Finance	Real estate	Sentiments
GDP(SA)(Q)		V											
GDP(NSA)(Q)		V											
Private consumption(SA)(Q)				V									
Government consumption(SA)(Q)							V						
Construction(SA)(Q)						V							
Facility investment(SA)(Q)					V								
Exports of goods and services(SA)(Q)	V							V					
Imports of goods and services(SA)(Q)	V								V				
Private consumption(NSA)(Q)				V									
Government consumption(NSA)(Q)							V						
Construction(NSA)(Q)						V							
Facility investment(NSA)(Q)					V								
Exports of goods and services(NSA)(Q)	V							V					
Imports of goods and services(NSA)(Q)	V								V				
Unemployment rate			V										
Employment to population ratio			V										
Number of employed people			V										
Monthly goods exports	V							V					
Monthly goods imports	V								V				
Export price index	V							V		V			
Import price index	V								V	V			
Producer price index		V								V			
Consumer price index				V						V			
Price index excluding agricultural product & oil										V			
Price index excluding food & energy										V			
Consumption & Retail sales index(SA)				V									
Service industry production index(SA)		V		V									
Consumption & Retail sales index(NSA)				V									
Service industry production index(NSA)		V		V									
Manufacturing industry production index(SA)		V											
Manufacturing industry shipment index(SA)		V											
Manufacturing inventory index(SA)		V											
Manufacturing industry production index(NSA)		V											
Manufacturing industry shipment index(NSA)		V											
Manufacturing inventory index(NSA)		V											
Facility investment index(SA)					V								
Construction completed(SA)						V							
Facility investment index(NSA)					V								
Construction completed(NSA)						V							
Manufacturing business performance BSI(SA)		V											V
Manufacturing business performance BSI(NSA)		V											V
All industries performance BSI		V											V
Service industry performance BSI		V											V
All industries sales BSI		V											V
Manufacturing export BSI								V					V
Manufacturing domestic demand sales BSI		V											V
Manufacturing new orders BSI		V											V
Manufacturing operation rate BSI		V											V
Economic sentiment index													V
Current economic judgment CSI													V
Consumer sentiment index				V									V
Consolidated fiscal balance							V						
Housing sales price index(HSPI)-Seoul												V	
Housing sales price index(HSPI)-National												V	
Housing lease price index(HLPI)-Seoul												V	
Housing lease price index(HLPI)-National												V	
Call rate											V		
CD rate											V		
KTB 3-year rate											V		
KRW exchange rate	V										V		
EUR exchange rate	V										V		
KOSPI											V		
KOSDAQ											V		
Dubai crude oil											V		
WTI futures											V		
Gold futures											V		
News sentiment index(T)													V
Economic policy uncertainty(T)													V
Production(T)		V											
Shipbuilding(T)		V											
Automotive(T)		V											
Semiconductor(T)		V											
Facility investment(T)					V								
Construction(T)						V							
Unemployment(T)			V										
Recruitment(T)			V										
Job search(T)			V										
Wholesale & retail(T)				V									
Government expenditure(T)							V						
Price outlook(T)										V			
Stock price outlook(T)											V		
House price outlook(T)												V	
World trade(T)	V												

Notes: Variables with a (T) mark indicate that they are computed from text data and (Q) mark indicates quarterly variables. In total, 83 variables are used with 69 monthly and 14 quarterly variables, consisting of 52 official statistics and 14 financial indices, and 17 text indices.

effectiveness of both linear and non-linear models for text indicators.

In addition to using TFM, which is a linear model, we also constructed a *Convolutional Recurrent Neural Network* (CRNN) as a non-linear model to compare its predictive accuracy using textual information. The CRNN consists of 64 convolutional filters applied in 3-month intervals, followed by a *Long Short-Term Memory* (LSTM) layer consisting of 32 units. The convolutional layer helps estimate direct relationships between observed variables, while the LSTM layer incorporates time-series characteristics into the model. Additionally, the CRNN model was designed to handle mixed-frequency data by aggregating monthly and quarterly data points into quarterly intervals using convolutional filters. Unlike TFM, which uses an unsupervised setup with latent factors and autoregressive structure to predict all observed variables simultaneously, CRNN was designed for a supervised setup specifically to predict GDP using all other variables.

For p-dimensional vector of input variables, X_t, and the target variable, y, CRNN consists of the following layers:

- Convolutional Layer:

$$Z_{t\in T_q} = \sigma(W_k X_{t\in T_q} + b_k). \tag{10}$$

- Long Short-Term Memory Layer

$$\begin{aligned} f_q &= \sigma(V_f Z_q + U_f h_{q-1} + e_f), & (11)\\ i_q &= \sigma(V_i Z_q + U_i h_{q-1} + e_i), & (12)\\ o_q &= \sigma(V_o Z_q + U_o h_{q-1} + e_o), & (13)\\ \tilde{c}_q &= \tilde{\sigma}(V_c Z_q + U_c h_{q-1} + e_c), & (14)\\ c_q &= f_q \circ c_{q-1} + i_q \circ \tilde{c}_q, & (15)\\ h_q &= o_q \circ \sigma(c_t). & (16) \end{aligned}$$

- Feed-forward Layer

$$\begin{aligned} Z'_q &= \sigma(W'_j c_q + b'_j), & (17)\\ y_q &= W'' Z'_q + b''. & (18) \end{aligned}$$

where $\sigma()$ is the ReLu function, i.e., element-wise $max(0,x)$, $\tilde{\sigma}$ is hyperbolic tangent function, and $\circ$ is the Hadamad product, i.e., element-wise product. $W_k \in \mathbb{R}^{p\times 3}$ and $b_k \in \mathbb{R}$ are the kernel weights and bias for convolutional layer for $k = 1,\cdots,\tilde{k}$ with the number of convolutional filters $\tilde{k} = 64$. $V \in \mathbb{R}^{h\times\tilde{k}}, U \in \mathbb{R}^{h\times h}$ and $e \in \mathbb{R}^h$ are the weights and bias for LSTM layer for $h = 32$ indicating the number of hidden units. $W' \in \mathbb{R}^{h\times d}, W'' \in \mathbb{R}^d, b' \in \mathbb{R}^d$, and $b'' \in \mathbb{R}$ are the weights and bias for the feed-forward (FF) layers with the number of hidden unit of FF layer $d = 4$. Within LSTM layer, f, i, and o represent the forget, update, and output gates respectively. c and h denote the cell and hidden state vectors.

$T_q = \{t_{i-2}, t_{i-1}, t_i\}$ is the index set of the quarter q that indicating 3 continuous months in quarter q. That is, for $\mathbb{X} \in \mathbb{R}^{p\times T}$, the kernel W_k is applied for every 3 month sliding through the input $\mathbb{X}$ and striding with 3 month time windows. These convolutional filters produce distinct bridge equations. In order to address the issue of ragged edges, we impute missing values that were not available during analysis with the most recently available data.

IV. Empirical Analysis

1. Validation of Theme Frequency Indices

One advantage of computing TFIs by sector is that their validity can be verified by comparing them to relevant official statistics. As shown in Table 5 and 6, TFIs have a high correlation with and tend to lead the corresponding official statistics. Some TFIs exhibit interesting characteristics in Figure 5 and Table 5, as outlined below:

- Most TFIs (excluding some) show that it is leading official statistics based on the cross-correlation and Granger causality test. Whether the text index leads or not varies by sector. A hypothesis is that the text sector that consists of expert opinions more than mere conveyance of released statistics tends to precede official statistics, which are the sectors with broader topics and having more attention from the public.
- Text-based TFIs tend to fluctuate in short periods indicating it includes noise, yet obvious outliers do not appear because it is an index quantifying

the attention of people. For house construction, the official statistic, which is based on surveying, shows more apparent outliers.

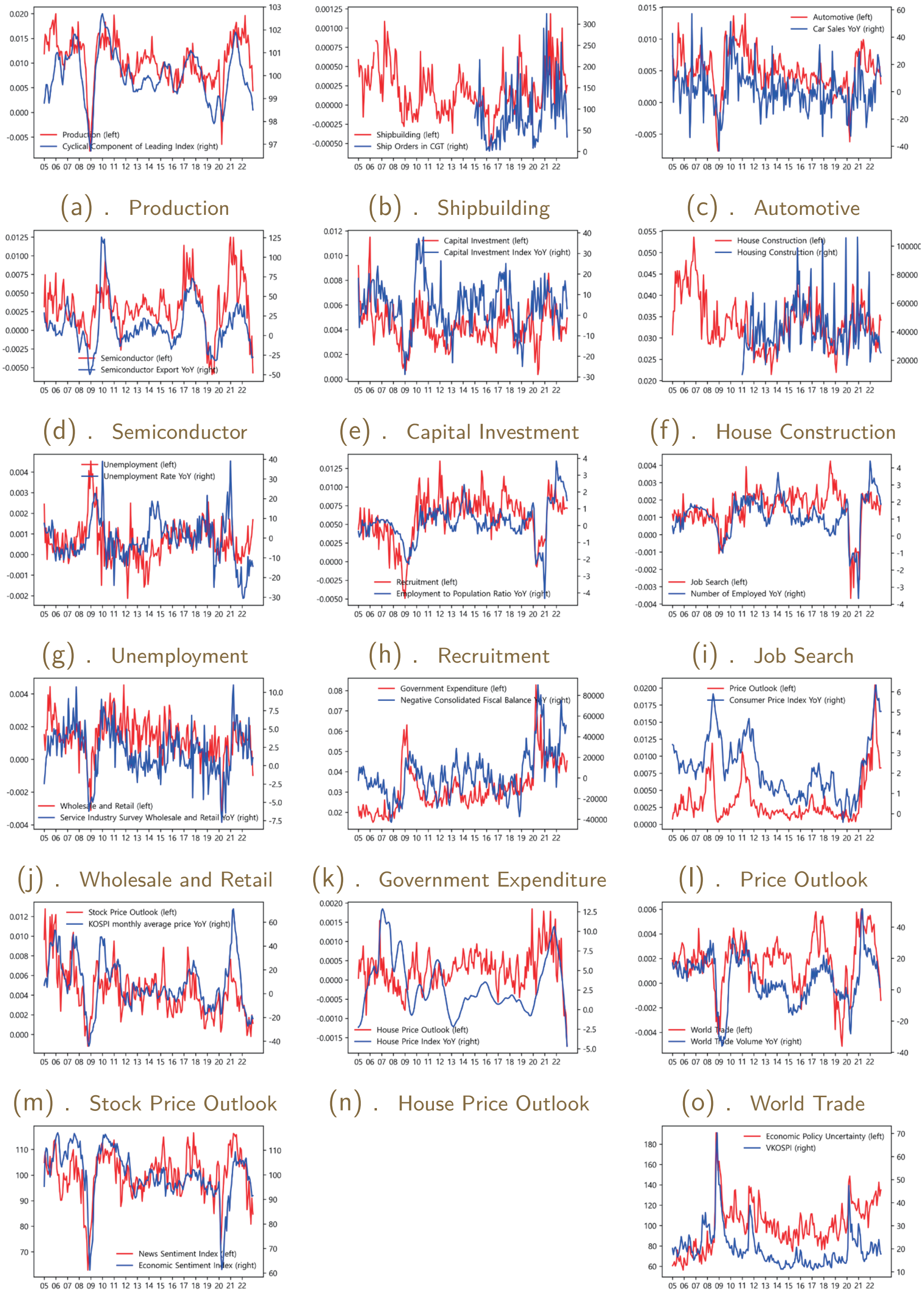
- For government expenditure TFI, the time lag that shows the maximum cross-correlation between the TFI and consolidated fiscal balance is 3 months, with the TFI leading. This is noteworthy because the official statistic is released with a 3-month delay.

- Regarding inflation outlook TFI, analysis of data from January 2006 to December 2022 shows that the maximum cross-correlation between TFI and CPI is 0.78, with TFI leading by 2 months. However, if we limit the data to January 2006 to December 2021, the maximum cross-correlation time lag increases to 5 months, with a value of 0.75. This may be due to the challenging environment of predicting inflation in 2022.
  When we compute the index without using words such as ’forecast 전망’, ’predict 예측’, or ’expect 예상’ both including and excluding the 2022 data, the time lag remains at 2 months with TFI leading, with a slight higher correlation of 0.82 and 0.77, respectively.

- To compute the inflation, stock price, and house price outlook TFIs, we included the words ’forecast 전망’ and ’predict 예측’. These TFIs have time lags with the maximum cross-correlation occurring at 2, 3, and 10 months, respectively. This finding aligns with our expectations given the forecasting horizons of interest, which are relatively short (around 1 to 3 months) for stock prices and inflation, and relatively long (around half to a year) for housing prices. This outcome is reasonable considering that TFIs reflect the opinions and forecasts of the field experts as expressed in news articles.

Table 5 and 6 demonstrate TFIs for each topic are useful as leading indicators for the corresponding sector and can be used on their own to identify trends in the sector. Therefore, one potential application of TFIs is to use them as instrumental variables (IV) to measure public interest of each topic. TFIs can be employed as IV in non-experimental research to address endogeneity issues. Figure 5 provides detailed time series comparisons between TFIs and official statistics, and Figure 7 shows the ± 6 month cross-correlations between them.

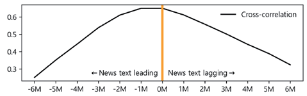
Figure 5. Time series plots of 15 sector text indices.



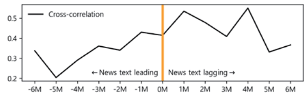
(a) . Production
(b) . Shipbuilding
(c) . Automotive
(d) . Semiconductor
(e) . Capital Investment
(f) . House Construction
(g) . Unemployment
(h) . Recruitment
(i) . Job Search
(j) . Wholesale and Retail
(k) . Government Expenditure
(l) . Price Outlook
(m) . Stock Price Outlook
(n) . House Price Outlook
(o) . World Trade
(p) . News Sentiment Index
(q) . Econ. Policy Uncertainty

Notes: Text-based economic indicators (red lines) and their relevant official statistics (blue lines) between January 2005 and December 2022 exhibit similar trends to each other.

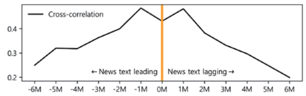
Figure 7. Cross-correlation plots between 15 sector text indices and their corresponding official statistics.



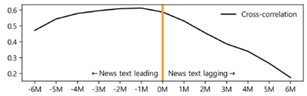
(a) . Production



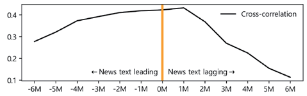
(b) . Shipbuilding



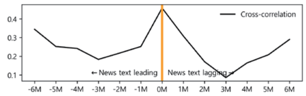
(c) . Automotive



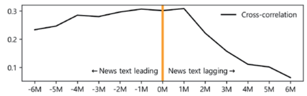
(d) . Semiconductor



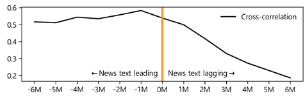
(e) . Capital Investment



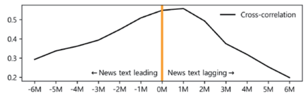
(f) . House Construction



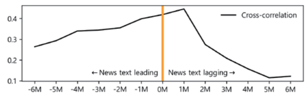
(g) . Unemployment



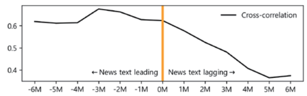
(h) . Recruitment



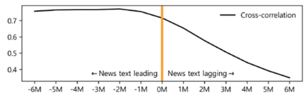
(i) . Job Search



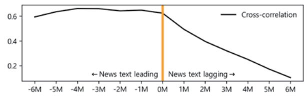
(j) . Wholesale and Retail



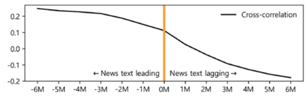
(k) . Government Expenditure



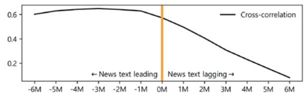
(l) . Price Outlook



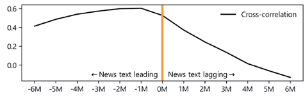
(m) . Stock Price Outlook



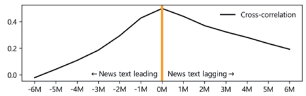
(n) . House Price Outlook



(o) . World Trade



(p) . News Sentiment Index



(q) . Econ. Policy Uncertainty

Table 5. Max leading lags and cross-correlation in the max lag between TFIs and official statistics.

Topic		Related Official Stat.	Max Leading Lag & Corr.
Industry	Product	Cyclical component of leading index (KOSTAT)	0 / 0.72
	Shipbuilding	Ship orders in CGT (Clarkson Research)	4 / 0.55
	Automotive	Car sales in units (KAA)	-1 / 0.51
	Semiconductor	Semiconductor exports (IITP)	-2 / 0.62
Facilities Investment		Equipment investment index (KOSTAT)	-1 / 0.47
House Construction		Housing construction (MLIT)	0 / 0.46
Employ-ment	Unemployment	Unemployment rate (KOSTAT)	1 / 0.32
	Recruitment	Employment to population ratio (KOSTAT)	-1 / 0.58
	Job Search	Number of employed (KOSTAT)	1 / 0.56
Wholesale and Retail		Service industry survey - W&R (KOSTAT)	1 / 0.46
Government Expenditure		Government fiscal balance (MOEF)	-3 / 0.69
Inflation Outlook		Consumer price index (KOSTAT)	-2 / 0.78
Stock Price Outlook		KOSPI closing price (KRX)	-6 / 0.64
House Price Outlook		House price index (Korea real estate board)	-10 / 0.34
World Trade		World merchandise trade volume (OECD)	-3 / 0.65
News Sentiment Index		Economic sentiment index (BOK)	-1 / 0.59
Econ. Policy Uncertainty		VKOSPI index (KRX)	0 / 0.54

Notes: The calculations are performed from January 2006 to December 2022. Max leading lags indicate the cross-correlation has the maximum value in the month between the text index and corresponding official statistics where the text index is leading.
KOSTAT: Korean Statistics, KAA: Korea Automobile Association, IITP: Institute of Information & communications Technology Planning & Evaluation, MLIT: Ministry of Land Infrastructure and Transport in Korea, MOEF: Ministry of Economy and Finance, KRX: Korea Exchange, OECD: Organization for Economic Cooperation and Development, BOK: Bank of Korea.

Table 6. Granger causality tests between TFIs and official statistics.

Topic		Offi. Stat.→Text		Text→Offi. Stat.		Direction
Industry	Product	3.428	(0.001)**	1.538	(0.138)	Stat.→Text
	Shipbuilding	2.884	(0.010)*	2.055	(0.059)	Stat.→Text
	Automotive	1.917	(0.052)	4.694	(0.000)**	Text→Stat.
	Semiconductor	2.414	(0.017)*	2.626	(0.010)*	Bidirectional
Facilities Investment		4.759	(0.010)*	4.323	(0.015)*	Bidirectional
House Construction		0.082	(0.775)	6.931	(0.009)**	Text→Stat.
Employ-ment	Unemployment	3.489	(0.017)*	0.631	(0.596)	Stat.→Text
	Recruitment	3.460	(0.033)*	9.165	(0.000)**	Text→Stat.
	Job Search	14.247	(0.000)**	0.521	(0.668)	Stat.→Text
Wholesale and Retail		6.626	(0.000)**	1.276	(0.284)	Stat.→Text
Government Expenditure		3.149	(0.006)**	6.663	(0.000)**	Bidirectional
Inflation Outlook		0.102	(0.750)	30.557	(0.000)**	Text→Stat.
Stock Price Outlook		0.391	(0.815)	1.829	(0.125)	Not sig.
House Price Outlook		1.609	(0.189)	1.346	(0.261)	Not sig.
World Trade		1.264	(0.281)	3.846	(0.002)**	Text→Stat.
News Sentiment Index		2.288	(0.080)	7.303	(0.000)**	Text→Stat.
Econ. Policy Uncertainty		1.103	(0.349)	4.498	(0.004)**	Text→Stat.

Notes: The calculations are performed from January 2006 to December 2022.
*, ** indicate that Granger statistics are significant with $\alpha = 0.05, 0.01$

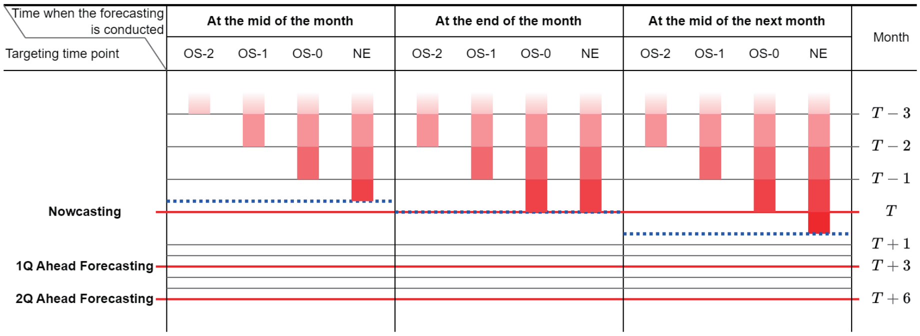
2. Forecasting Results with Text-Enhanced Factor Model

2.1. Validation of Predictive Accuracy Improvement through Textual Information

This subsection aims to examine how incorporating textual information in economic forecasting models affects their predictive accuracy. Figure 9 illustrates different forecasting scenarios. In this study, we focus on predicting GDP for the current and following quarters with predictions made at the end of each quarter.

To evaluate the testing error of predicting GDP growth, we computed both the mean absolute error (MAE) and the root mean square error (RMSE) for out-of-sample predictions from the first quarter of 2016 to the fourth quarter of 2022. Specifically, we made recursive predictions for the seasonally adjusted GDP quarter-over-quarter growth and non-seasonally adjusted GDP year-over-year growth rates for the current and following quarters using the available vintage data at the end of the current quarter. To distinguish the impact of adding textual information from the impact of model selection on accuracy improvement, we

Figure 9. Nowcasting and forecasting situations



Notes: The variables are categorized as follows based on their release delay.

- OS-2: official statistics released in two months - Retail sales index.
- OS-1: official statistics released in a month: Unemployment rate, Employment to population ratio, Number of employed people, Export price index, Consumer price index, Price index excluding agricultural product & oil, Price index excluding food & energy, Producer price index, Facility investment index, Manufacturing production index, Manufacturing industry shipment index, Manufacturing inventory index, Service industry production index, National accounts (GDP, Private consumption, Construction, Facility investment, Exports of goods and services).
- OS-0: official statistics available at the end of the month: All industries sales BSI, All industries performance BSI, Manufacturing export BSI, Manufacturing operation rate BSI, Manufacturing new orders BSI, Manufacturing domestic demand sales BSI, Manufacturing business performance BSI, Economic sentiment index, Current economic judgment CSI, Consumer sentiment index, Housing sales price index (Seoul, National), Housing lease price index (Seoul, National), Call rate, CD rate, 3-year Korean treasury bond rate, Exchange rates, KOSPI, KOSDAQ, VKOSPI, Dubai crude oil, WTI futures, gold futures.
- NE: 15 sector text indices, News sentiment index, and Economic uncertainty index.

computed the testing accuracy in a cross-tabulation of both data and models, as presented in Table 7. Additionally, we also explored smoothing techniques for text indicators to improve the forecasting accuracy in Table 8.

Textual information, which quantifies public opinions expressed in news articles, may contain noise and reflect various aspects of the economy. It can indicate the level or change of an economic variable over the past month, the past year, or a combination of both. Understanding these characteristics of textual information, we examined the impact of smoothing and seasonal adjustments on text indicators. To smooth the text indicators, we applied Hodrick and Prescott (1997) filters with a frequency of 0.5, equivalent to using the modifying multiplier $\lambda = 6.24 \cdot (0.5)^4 \approx 0.391$, and utilized the cycle components. We then employed X13 ARIMA from the United States Census Bureau to investigate the seasonal effect and transformed them into month-over-month and year-over-year growth rates. Table 8 shows that smoothing is crucial for forecasting GDP(SA) quarter-over-quarter growth rates, but has little effect on predicting GDP(NSA) year-over-year growth rates. For the remainder of this subsection, we focus on nowcasting using HP-filtered and seasonally adjusted month-over-month growth rates of the text indicators for predicting GDP(SA) and year-over-year transformed text indicators for predicting GDP(NSA).

Based on the findings in Table 7, the inclusion of textual information enhances the predictive accuracy of both linear and nonlinear models for both the current and the following quarters. The effect of textual information on GDP(SA) QoQ prediction of the current quarter is less obvious while GDP(NSA) YoY growth of both the current and the following quarters and GDP(SA) QoQ growth for the following quarter show the inclusion of textual information decreases the predictive errors. TFM estimates its factor by utilizing both the observed information including textual information and the autoregressive structure of the model. Hence, if the text indicators are suitable as the complementary variables for each factor, the model's fitness improves and the standard error of the predictions decreases.[4)]

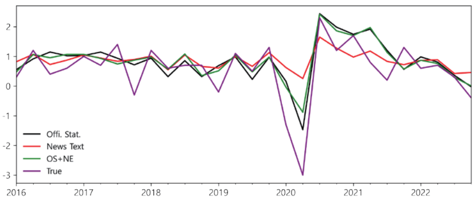
Additionally, in every case, TFM outperforms CRNN. One possible explanation for the superiority of the relatively simpler linear model, TFM, over the nonlinear model, CRNN, in predicting GDP could be attributed to the nature of GDP, which is merely a linear sum of sub-segments. Hence, the nonlinear structure of the CRNN model may not be necessary and could lead to overfitting errors, while the simpler linear model can provide more accurate predictions. In general, neural networks have benefits when the target variable is not expressed by a simple linear combination of inputs, and when new features synthesized from the network supplement the original inputs' insufficient information.

Furthermore, for CRNN, the model with official statistics shows inferior performances compared to that with only text indicators. This can be attributed to imputing missing values at ragged edges. CRNN extrapolates each missing observation using each individual time series separately. Hence, the CRNN with official statistics relies on outdated information to extrapolate. In contrast, CRNN with only text indicators and that with text-enhanced data can incorporate more recent information using text. On the other hand, the TFM and DFM models improve their predictions by imputing missing values with other observed variables such as text during the estimation process using the EM algorithm.

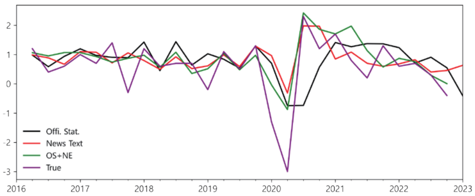
Our analysis demonstrated that when utilizing solely text indicators, the TFM's forecast accuracy for GDP(SA) nowcasting was slightly lower than that of DFM using solely official statistics. This outcome can be attributed to the limited number of text indicators, which only consist of 17 text indices. However, despite this limitation, the TFM's accuracy remains comparable to that of DFM, indicating that there is a vast amount of economic forecasting information available from news sources. Furthermore, the prediction accuracy of major monthly economic variables in TFM has been presented in Table 9. The improvement of forecasting accuracy varies across variables. For monthly variables that are rela-

4) In comparable studies of GDP nowcasting during the COVID-19 pandemic, Lee et al. (2022) achieved an MAE of 0.513 for GDP(SA) using XGBoost model with a fixed variable selection, and an MAE of 0.429 with dynamic optimization variable selection based on 145 variables. Additionally, Jung et al. (2022) achieved an RMSE of 0.99 and 0.87 for GDP(SA) using state-of-the-art LSTM and GRU neural network models with recursive training, and an RMSE of 0.92 and 0.98 using rolling window training with these models based on 36 variables.

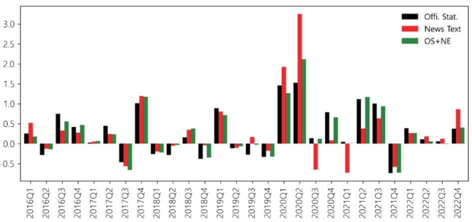
Figure 11. Predicted GDP(SA) growth rates and their error.



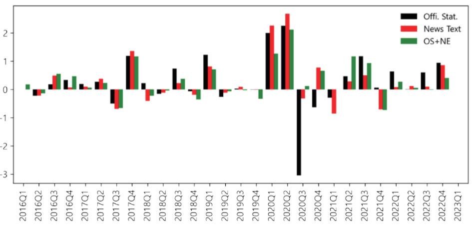
(a) . Nowcasting GDP



(b) . 1Q ahead forecasting GDP



(c) . Nowcasting error



(d) . 1Q ahead forecasting error

tively timely and have ample available information, incorporating text indicators have a lower impact on improving predictive power compared to quarterly GDP.

2.2. Exploring the Characteristics of the Text-Enhanced Factor Model

In this subsection, the forecast accuracy of TFM was examined for GDP year-over-year growth rate by the end of March 2020, when the COVID-19 outbreak began. Figure 12 demonstrates that TFM enhanced the prediction accuracy of GDP in March 2020 in comparison to using only official statistics. This highlights the value of utilizing textual information, particularly in scenarios where the economic landscape changes rapidly and inadequate quantitative data is available. Furthermore, the figure illustrates that incorporating textual information may alter the mid-term forecasting trajectory for GDP by up to one and a half years in the considered TFM.

Although, it is worth noting that evaluating machine learning models based on their long-term forecasting performance is unreasonable, as it is difficult to assume that the inputs contain information about the long-term forecast horizon, even if a particular model shows lower prediction errors. In Figure 12, the reason why the TFM forecast for the one-year horizon showed superior forecasting per-

Table 7. Out-of-sample testing error with its std. dev. for predicting gross domestic product (GDP(SA) and GDP(NSA))

	Data / Model	Nowcasting			1Q Ahead Forecasting		
		Offi. Stat.	News text	OS+NE	Offi. Stat.	News text	OS+NE
• Prediction on GDP seasonally adjusted (SA) QoQ growth							
MAE	AR	0.787	-	-	0.880	-	-
		(.876)	-	-	(.781)	-	-
	CRNN	0.677	0.605	0.589	-	-	-
		(.638)	(.508)	(.606)	-	-	-
	TFM†	0.495	0.529	**0.476**	0.656	**0.547**	0.598
		(.419)	(.674)	(.494)	(.749)	(.648)	(.497)
RMSE	AR	1.387	-	-	1.385	-	-
		(3.265)	-	-	(2.679)	-	-
	CRNN	0.931	0.790	0.845	-	-	-
		(1.801)	(1.096)	(2.097)	-	-	-
	TFM†	**0.643**	0.848	0.680	0.986	0.840	**0.772**
		(.606)	(2.029)	(.895)	(2.013)	(1.618)	(.940)
• Prediction on GDP non-seasonally adjusted (NSA) YoY growth							
MAE	AR	1.000	-	-	1.213	-	-
		(1.058)	-	-	(1.504)	-	-
	CRNN	0.953	0.950	0.902	-	-	-
		(.975)	(.775)	(.877)	-	-	-
	TFM†	0.790	0.838	**0.707**	1.217	1.085	**1.069**
		(.741)	(.655)	(.692)	(.826)	(.772)	(.737)
RMSE	AR	1.456	-	-	1.933	-	-
		(4.319)	-	-	(9.953)	-	-
	CRNN	1.364	1.226	1.258	-	-	-
		(3.140)	(2.356)	(3.085)	-	-	-
	TFM†	1.074	1.064	**0.989**	1.417	1.332	**1.299**
		(1.884)	(1.797)	(1.670)	(2.429)	(2.316)	(2.037)

Notes: The calculation are performed between 2016.1Q and 2022.4Q for mean absolute error (MAE) and root mean square error (RMSE).
† For the official-statistics-only model, the results have been obtained from DFM.

Table 8. Out-of-sample testing error with its std. dev. of TFM based on text index transformation for predicting GDP(SA) and GDP(NSA).

Trans. / Target	OS+NE HP SA MoM	OS+NE SA MoM	OS+NE HP YoY	OS+NE YoY	Offi. Stat.
GDP(SA)	**0.476**	0.519	0.503	0.503	0.495
(Nowcasting)	(.494)	(.477)	(.419)	(.420)	(.419)
GDP(NSA)	0.893	0.864	0.716	**0.707**	0.790
(Nowcasting)	(.784)	(.766)	(.688)	(.692)	(.741)

Notes: The calculation are performed between 2016.1Q and 2022.4Q, measured by mean absolute error (MAE).
HP: HP filtered. SA: seasonally adjusted. MoM: transformed with month-over-month growth. YoY: transformed with year-over-year growth.

Table 9. Out-of-sample testing error with its std. dev. of DFM and TFM for predicting monthly variables.

Variable \ Data	Nowcasting		1M ahead FC		2M ahead FC	
	OS	OS+NE	OS	OS+NE	OS	OS+NE
Unemployment rate	3.749 (2.926)	3.706 (2.883)	4.458 (2.583)	4.395 (2.694)	5.883 (4.645)	5.935 (4.596)
Consumer price index	0.179 (0.129)	0.178 (0.122)	0.239 (0.185)	0.234 (0.202)	0.232 (0.156)	0.224 (0.169)
Manufacturing prod. ind.(SA)	2.217 (2.354)	1.923 (2.026)	1.966 (1.303)	1.986 (1.248)	1.854 (1.592)	1.807 (1.334)
Service industry prod. ind.(SA)	0.850 (1.069)	0.764 (1.092)	0.641 (0.554)	0.664 (0.518)	0.835 (0.661)	0.840 (0.730)
Facility investment ind.(SA)	3.598 (2.468)	3.567 (2.536)	2.897 (1.795)	2.840 (1.965)	3.844 (2.692)	4.010 (2.610)
Construction completed(SA)	2.348 (2.278)	2.387 (2.38)	2.297 (2.121)	2.383 (2.246)	3.130 (2.138)	3.068 (2.213)
Housing sales price ind.-Nat.	0.178 (0.150)	0.150 (0.145)	0.109 (0.083)	0.125 (0.085)	0.188 (0.154)	0.156 (0.160)

Notes: The calculation are performed between January 2016 and December 2022, measured by mean absolute error (MAE).

formance compared to DFM could be attributed to the fact that the TFM model anticipated a deeper trough in the following quarter, June 2020. The subsequent forecast path was then determined by the autoregressive structure of the linear factor model.

In Figure 14, we can see the coefficient of determination, R^2, that each factor has for each variable, x, in the vintage data as of March 2020.

$$R_x^2 = \frac{\sum(x_t - \hat{x}_t)^2}{\sum(x_t - \bar{x})^2} \tag{19}$$

$$\hat{x}_t = \hat{\Lambda}\hat{f}_t \tag{20}$$

The figure allows us to analyze how each factor relates to the observed variables by computing the explanatory power of each factor for each observed variable. Overall-1 factor has the strongest relationship with most text indicators, and the coefficient of determination of the text indicators for individual factors varies depending on their respective factors.

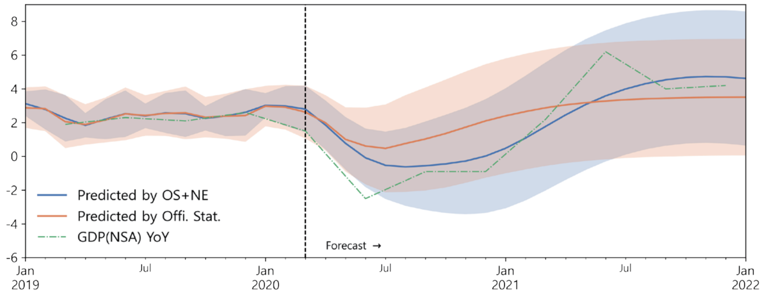
Tables 10 display the variables that had the greatest impact on the change in GDP forecast between December 2019 and March 2020. For the current quarter, the first quarter of 2020, the change in forecast is primarily attributed to survey

Table 10. The variables affecting the GDP(NSA) forecasts for the first and second quarters of 2020.

		Observed	Prev. forecast	News	Weight	Impact on GDP
• For forecasting the 1st quarter of 2020						
2020-02	Economic sentiment index	-10.00	1.21	-11.21	-0.02	0.18
	Manufact. export BSI	-10.78	1.29	-12.07	-0.01	0.18
	Production(T)	-0.36	0.21	-0.58	-0.20	0.12
2020-03	Government expenditure(T)	2.89	-0.10	2.99	0.04	0.11
	Automotive(T)	-0.35	0.22	-0.57	-0.18	0.10
2020-01	Manufact. ind. shipment(NSA)	-6.99	2.15	-9.14	0.01	-0.10
	Production(T)	0.03	0.25	-0.22	-0.45	0.10
2020-02	All industries sales BSI	-8.98	0.80	-9.79	-0.01	0.10
	All industries performance BSI	-11.82	1.06	-12.88	-0.01	0.10
2020-01	Manufact. ind. shipment(SA)	-4.44	0.29	-4.73	0.02	-0.08
• For forecasting the 2nd quarter of 2020						
2020-03	Government expenditure(T)	2.89	-0.10	2.99	-2.01	-6.01
2020-02	Government expenditure(T)	1.62	-0.08	1.69	2.98	5.05
	Production(T)	-0.36	0.21	-0.58	-4.29	2.47
2020-03	Production(T)	-0.72	0.17	-0.89	1.45	-1.29
2020-02	Automotive(T)	-0.05	0.26	-0.30	-2.74	0.83
	Wholesale & retail(T)	-1.19	0.53	-1.72	-0.46	0.79
	News sentiment index(T)	-0.10	0.08	-0.17	-3.56	0.62
2020-03	Economic policy uncert.(T)	0.46	-0.06	0.52	-1.17	-0.61
	Wholesale & retail(T)	-2.39	0.43	-2.83	0.22	-0.61
2020-02	Economic policy uncert.(T)	0.18	-0.08	0.26	2.34	0.60

Notes: At the end of March 2020, the variables affecting the forecasted GDP (NSA) YoY growth rate for the first and second quarters of 2020 were examined and compared to the values predicted in December 2019. The majority of the new information used for the current and next quarters forecast was derived from textual and survey variables.

Figure 12. The GDP forecasting comparison between TFM and DFM during the COVID-19 outbreak



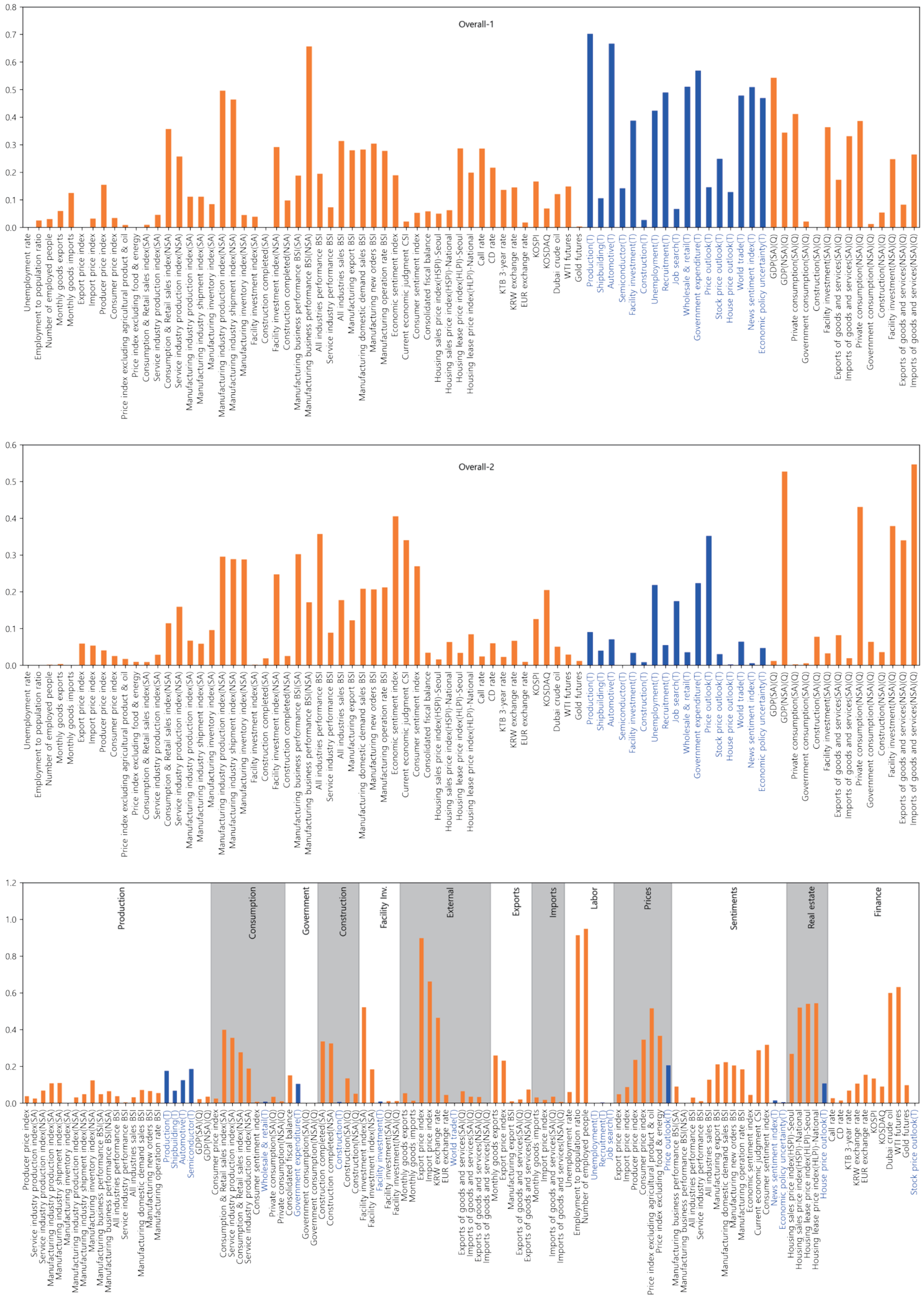
Notes: The comparison was made in March 2020, when the COVID-19 outbreak occurred. The DFM is based solely on official statistics, and the TFM is enhanced with textual information.

indices and textual information, while for the subsequent quarter forecasting, the second quarter of 2020, textual information has the greatest influence on the change.

V. Discussion

This paper proposes a new method to generate news-text-based economic indicators called *Theme Frequency Indices* (TFI), which reflect economic trends across sectors. TFIs are computed using simple subject-predicate patterns of word groups based on domain knowledge without any labeled data. By computing the text indicators across economic fields, the indices can be easily verified and used as timely alternatives to their corresponding official statistics. Empirical analysis demonstrates that TFIs computed using simple text-mining techniques, along with domain knowledge, offer insights highly correlated with and preceding official statistics. This is crucial because TFIs can help circumvent the costly training procedure. Furthermore, this paper investigates suitable econometric models to incorporate textual information into a forecasting model for the gross domestic product (GDP) in Korea. A model structure incorporating TFIs into separate factors has been proposed as *Text-enhanced Factor Model* (TFM) with empirical examinations of its predictive accuracy and characteristics. The empir-

Figure 14. The explanatory power, R^2, that each factor has for each observed variable.



Notes: The blue color in the visualization indicates that the variable is computed from textual data, while the orange color represents variables obtained from other sources.

ical study shows that TFM improves near-term economic forecasting accuracy for the gross domestic product (GDP) of Korea compared to the model relying on only official statistics.

The findings of this study can contribute to related research in two folds. First, the proposed text-mining technique using subject-predicate patterns can be applied to numerous topics to analyze textual information without expensive labeled data. The approach can complement traditional surveying methods with remarkably lower cost and less time by using ubiquitous news sources with automated algorithms. Moreover, this study illustrates the potential of textual information for economic forecasting, which is still in its infancy. This study emphasizes the information from news sources is valuable by providing big data-driven insight relevant to public perception of the economy.

References

Babii, A., Ghysels, E., Striaukas, J., 2022. Machine learning time series regressions with an application to nowcasting. Journal of Business & Economic Statistics 40, 1094–1106.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The quarterly journal of economics 131, 1593–1636.

Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. Handbook of economic forecasting 2, 195–237.

Bańbura, M., Giannone, D., Reichlin, L., 2010. Large bayesian vector auto regressions. Journal of applied Econometrics 25, 71–92.

Barbaglia, L., Consoli, S., Manzan, S., 2022. Forecasting with economic news. Journal of Business & Economic Statistics , 1–12.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A.M., Tambalotti, A., 2018. Macroeconomic nowcasting and forecasting with big data. Annual Review of Economics 10, 615–643.

Bybee, L., Kelly, B.T., Manela, A., Xiu, D., 2021. Business news and business cycles. National Bureau of Economic Research .

Caldara, D., Iacoviello, M., 2022. Measuring geopolitical risk. American Economic Review 112, 1194–1225.

Ferreira, L.N., et al., 2021. Forecasting with var-text and dfm-text models: exploring the predictive power of central bank communication. Banco Central do Brasil 2021.

Hájek, P., Olej, V., 2013. Evaluating sentiment in annual reports for financial distress prediction using neural networks and support vector machines. International Conference on Engineering Applications of Neural Networks , 1–10.

Hodrick, R.J., Prescott, E.C., 1997. Postwar us business cycles: an empirical investigation. Journal of Money, credit, and Banking , 1–16.

Hopp, D., 2022. Economic nowcasting with long short-term memory artificial neural networks (lstm). Journal of Official Statistics 38, 847–873.

Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K., Felix, W.F., 2011. Identification of fraudulent financial statements using linguistic credibility analysis. Decision Support Systems 50, 585–594.

Jung, S., Pyo, D.J., Heo, D.Y., 2022. 딥러닝 방법론을 활용한 gdp 예측모형 연구 [study of gdp prediction models using deep learning methods]. National Assembly Budget Office of Korea 2022-8.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., Kapadia, S., 2022. Making text count: economic forecasting using newspaper text. Journal of Applied Econometrics 37, 896–919.

Lee, G.H., Cho, J.H., Jo, J.G., 2020. New economic policy uncertainty indexes for south korea. The Korean Journal of Applied Statistics 33, 639–653.

Lee, Y., Kim, Y., You, T., 2022. 딥러닝 시계열 모형을 이용한 당분기 gdp 예측 성능 분석 [analysis of qoq gdp prediction performance using deep learning time series model]. Journal of KIISE (JOK) 49, 873–883.

Lee, Y.J., Kim, S., Park, K.Y., 2019. Measuring monetary policy surprises using text mining: The case of korea. BOK Working Paper 2019-11.

Lewis, C., Young, S., 2019. Fad or future? automated analysis of financial text and its implications for corporate reporting. Accounting and Business Research 49, 587–615.

Li, J., Seo, B., Lin, L., 2019. Optimal transport, mean partition, and uncertainty assessment in cluster analysis. Statistical Analysis and Data Mining: The ASA Data Science Journal 12, 359–377.

Li, N., Liang, X., Li, X., Wang, C., Wu, D., 2009. Network environment and financial risk using machine learning and sentiment analysis. Human and Ecological Risk Assessment 15, 227–252.

Li, X., Wu, P., Wang, W., 2020. Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. Information Processing & Management 57, 102212.

Mæhle, N., HlÃ, T., Pranovich, M., Selander, C., et al., 2021. Taking stock of imf capacity development on monetary policy forecasting and policy analysis systems. IMF: International Monetary Fund .

Mariano, R.S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real gdp. Oxford Bulletin of economics and statistics 72, 27–46.

Richardson, A., van Florenstein Mulder, T., Vehbi, T., 2021. Nowcasting gdp using machine-learning algorithms: A real-time assessment. International Journal of Forecasting 37, 941–948.

Sax, C., Eddelbuettel, D., 2018. Seasonal adjustment by x-13arima-seats in r. Journal of Statistical Software 87, 1–17.

Seabold, S., Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference.

Seo, B., 2022. 뉴스 텍스트를 이용한 경기 예측: 경제 부문별 텍스트 지표의 작성과 활용 [economic forecasting using news text: Development and utilization of text indicators by economic sector]. BOK Issue Note 2022-18. URL: `https://www.bok.or.kr/portal/bbs/P0002353/view.do?nttId=10070554&menuNo=200433&pageIndex=1`.

Seo, B., 2023. Ai알고리즘을 이용한 산업모니터링: 증권사 리포트 텍스트 분석 [industry analysis using ai algorithm: Analyst reports text mining]. BOK Issue Note 2023-5. URL: https://www.bok.or.kr/portal/bbs/P0002353/view.do?nttId=10075584&menuNo=200433&pageIndex=1.

Seo, B., Lee, Y., Cho, H., 2022. Machine-learning-based news sentiment index (nsi) of korea. BOK Working Paper 2022-15.

Shapiro, A.H., Sudhof, M., Wilson, D.J., 2022. Measuring news sentiment. Journal of econometrics 228, 221–243.

Shin, S.J., Seo, B., 2022. 빅데이터를 이용한 실시간 민간소비 예측 [real-time private consumption prediction using big data]. BOK Working Paper 2022-16.

Sun, A., Lachanski, M., Fabozzi, F.J., 2016. Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. International Review of Financial Analysis 48, 272–281.

Te Liew, W., Adhitya, A., Srinivasan, R., 2014. Sustainability trends in the process industries: A text mining-based analysis. Computers in Industry 65, 393–400.

Thorsrud, L.A., 2020. Words are the new numbers: A newsy coincident index of the business cycle. Journal of Business & Economic Statistics 38, 393–409.

Yap, B.W., Ong, S.H., Husain, N.H.M., 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Systems with Applications 38, 13274–13283.

⟨Abstract in Korean⟩

보편적 뉴스 텍스트를 이용한 계량적 경제전망: 텍스트 강화 인자모형

서범석*

뉴스 텍스트를 경제 예측에 활용하고자 하는 연구들이 주목받고 있다. 본 논문은 학습 데이터 없이 경제 부문별 서술형 정보를 효과적으로 정량화하여 경제 예측에 활용하는 방법을 검토하였다. 본 논문은 경제 도메인별로 나타나는 주어-술어 패턴을 이용하여 '테마별 빈도 지수(Theme Frequency Index, TFI)'를 추정하고 이를 통해 경제에 대한 대중의 인식을 측정하였다. 구체적으로 생산, 인플레이션, 고용, 자본 투자, 주식, 주택 가격 등 15개 부문의 TFI를 제시하고 은닉인자 구조에 기반한 '텍스트 강화 인자모형 (Text-enhanced Factor Model, TFM)'을 구축하여 서술형 정보를 경제 예측에 반영하였다. 약 1천8백만 건의 뉴스기사를 바탕으로 한 실증분석 결과 TFM은 단기 GDP 예측 정확도를 향상시키는 것으로 나타나며, 이는 도메인 지식을 반영한 텍스트 마이닝 기술이 학습 비용 없이도 정성적 정보를 효과적으로 처리할 수 있음을 보여준다. 본 논문에서 제시한 방법론은 신속하고 효율적이며, 서술형 정보를 이용하기 위한 다양한 경제 부문에서 활용이 가능할 것으로 기대된다.

핵심 주제어: 동적인자모형, 텍스트마이닝, 머신러닝, 경제전망, 나우캐스팅

JEL Classification: C45, C53, C55, C82

* 한국은행 경제모형실 거시모형팀
(전화: 02-759-4248, E-mail: bsseo@bok.or.kr, Web: https://seo-beomseok.github.io)

논고 작성에 많은 도움을 주신 한국은행 경제모형실 배병호 실장님, 거시모형팀 오형석 팀장님과 유익한 논평을 주신 모형전망팀 박경훈 팀장님, 지역경제조사팀 정민수 차장님 그리고 현안포럼 참석자분들께 감사의 말씀을 전합니다.

BOK 경제연구 발간목록

한국은행 경제연구원에서는 Working Paper인 『BOK 경제연구』를 수시로 발간하고 있습니다.
『BOK 경제연구』는 주요 경제 현상 및 정책 효과에 대한 직관적 설명 뿐 아니라 깊이 있는 이론 또는 실증 분석을 제공함으로써 엄밀한 논증에 초점을 두는 학술논문 형태의 연구이며 한국은행 직원 및 한국은행 연구용역사업의 연구 결과물이 수록되고 있습니다.
『BOK 경제연구』는 한국은행 경제연구원 홈페이지(http://imer.bok.or.kr)에서 다운로드하여 보실 수 있습니다.

제2019 –1	Deciphering Monetary Policy Board Minutes through Text Mining Approach: The Case of Korea	Ki Young Park · Youngjoon Lee · Soohyon Kim
2	The Impacts of Macroeconomic News Announcements on Intraday Implied Volatility	Jieun Lee · Doojin Ryu
3	Taking a Bigger Slice of the Global Value Chain Pie: An Industry–level Analysis	Chong–Sup Kim · Seungho Lee · Jihyun Eum
4	Trend Growth Shocks and Asset Prices	Nam Gang Lee
5	Uncertainty, Attention Allocation and Monetary Policy Asymmetry	Kwangyong Park
6	Central Bank Digital Currency and Financial Stability	Young Sik Kim · Ohik Kwon
7	은행의 수익 및 자산구조를 반영한 통화정책 위험선호경로	김의진 · 정호성
8	혁신기업에 대한 산업금융 지원: 이론모형 분석	강경훈 · 양준구
9	가계부채 제약하의 통화정책: 2주체 거시모형(TANK)에서의 정량적 분석	정용승 · 송승주
10	Alchemy of Financial Innovation: Securitization, Liquidity and Optimal Monetary Policy	Jungu Yang
11	Measuring Monetary Policy Surprises Using Text Mining: The Case of Korea	Youngjoon Lee · Soohyon Kim · Ki Young Park
12	Tracking Uncertainty through the Relative Sentiment Shift Series	Seohyun Lee · Rickard Nyman
13	Intra–firm and Arm’s Length Trade during the Global Financial Crisis: Evidence from Korean Manufacturing Firms	Moon Jung Choi · Ji Hyun Eum

14	특허자료를 이용한 우리나라 지식전파의 지역화 분석	이지홍 · 남윤미
15	Overhead Labour and Skill-Biased Technological Change: The Role of Product Diversification	Choong Hyun Nam
16	Does the Number of Countries in an International Business Cycle Model Matter?	Myunghyun Kim
17	High-Frequency Credit Spread Information and Macroeconomic Forecast Revision	Bruno Deschamps · Christos Ioannidis · Kook Ka
18	경제 분석을 위한 텍스트 마이닝	김수현 · 이영준 · 신진영 · 박기영
19	Takeover, Distress, and Equity Issuance: Evidence from Korea	Euna Cho
20	The Cash-Flow Channel of Monetary Policy: Evidence from Mortgage Borrowers	Sang-yoon Song
21	부의 효과의 분위 추정: 분위 정준 공적분회귀를 중심으로	김기호
22	Identifying Government Spending Shocks and Multipliers in Korea	Kwangyong Park · Eun Kyung Lee
23	Systemic Risk of the Consumer Credit Network across Financial Institutions	Hyun Hak Kim · Hosung Jung
24	Impact of Chinese Renminbi on Korean Exports: Does Quality Matter?	Jihyun Eum
25	Uncertainty, Credit and Investment: Evidence from Firm-Bank Matched Data	Youngju Kim · Seohyun Lee · Hyunjoon Lim
26	A Structural Change in the Trend and Cycle in Korea	Nam Gang Lee · Byoung Hoon Seok

제2020 –1	인구 고령화가 실질 금리에 미치는 영향	권오익·김명현
2	달러라이제이션이 확산된 북한경제에서 보유외화 감소가 물가 · 환율에 미치는 영향	문성민·김병기
3	상태공간 벡터오차수정모형을 이용한 월별 GDP 추정: 깁스표본추출 접근	김기호
4	우리나라 외환시장 오퍼레이션의 행태 및 환율변동성 완화 효과	박준서·최경욱
5	Common Factor Augmented Forecasting Models for the US Dollar–Korean Won Exchange Rate	Hyeongwoo Kim · Soohyon Kim
6	북한 「경제연구」로 분석한 경제정책 변화: 텍스트 마이닝 접근법	김수현·손 욱
7	북한의 광물 수출과 품목별 수입: 대중무역을 중심으로	김병연·김민정·김다울
8	Network–Based Measures of Systemic Risk in Korea	Jaewon Choi · Jieun Lee
9	Aggregate Productivity Growth and Firm Dynamics in Korean Manufacturing 2007–2017	Kyoo il Kim · Jin Ho Park
10	2001년 이후 한국의 노동생산성 성장과 인적자본: 교육의 질적 개선 효과를 중심으로	유혜미
11	House Prices and Household Consumption in Korea	Seungyoon Lee
12	글로벌 가치사슬 변화가 경제성장에 미치는 영향: 2008년 금융위기 전후 전 · 후방참여 효과의 국제비교를 중심으로	김세완·최문정
13	산업구조조정이 고용 및 성장에 미치는 영향	서병선·김태경
14	Cross–border Trade Credit and Trade Flows During the Global Financial Crisis	Moon Jung Choi · Sangyeon Hwang · Hyejoon Im

15	International Co-movements and Determinants of Public Debt	Hasan Isomitdinov · Vladimir Arčabić · Junsoo Lee · Youngjin Yun
16	북한 비공식금융 실태조사 및 분석 · 평가	이주영 · 문성민
17	북한의 장기 경제성장률 추정: 1956~1989년	조태형 · 김민정
18	Macroeconomic and Financial Market Analyses and Predictions through Deep Learning	Soohyon Kim
19	제조업의 수출과 생산성 간 관계 분석: 사업체 자료 이용	이윤수 · 김원혁 · 박진호
20	우리나라 제조업 수출기업의 내수전환 결정요인 분석	남윤미 · 최문정
21	A Model of Satisficing Behaviour	Rajiv Sarin · Hyun Chang Yi
22	Vulnerable Growth: A Revisit	Nam Gang Lee
23	Credit Market Frictions and Coessentiality of Money and Credit	Ohik Kwon · Manjong Lee
24	북한의 자본스톡 추정 및 시사점	표학길 · 조태형 · 김민정
25	The Economic Costs of Diplomatic Conflict	Hyejin Kim · Jungmin Lee
26	Central Bank Digital Currency, Tax Evasion, Inflation Tax, and Central Bank Independence	Ohik Kwon · Seungduck Lee · Jaevin Park
27	Consumption Dynamics and a Home Purchase	Dongjae Jung
28	자본유입과 물가상승률 간의 동태적 상관관계 분석: 아시아의 8개국 소규모 개방경제를 중심으로	최영준 · 손종칠

29	The Excess Sensitivity of Long-term Interest rates and Central Bank Credibility	Kwangyong Park
30	Wage and Employment Effects of Immigration: Evidence from Korea	Hyejin Kim
제**2021**-1	외국인력 생산성 제고 방안—직업훈련 프로그램의 노동시장 성과 분석을 중심으로	김혜진 · 이철희
2	한국경제의 추세 성장률 하락과 원인	석병훈 · 이남강
3	Financial Globalization: Effects on Banks' Information Acquisition and Credit Risk	Christopher Paik
4	The Effects of Monetary Policy on Consumption: Workers vs. Retirees	Myunghyun Kim · Sang-yoon Song
5	북한지역 토지자산 추정에 관한 연구: 프레임워크 개발 및 탐색적 적용	임송
6	김정은 시대 북한의 금융제도 변화 - 북한 문헌 분석을 중심으로 -	김민정 · 문성민
7	Chaebols and Firm Dynamics in Korea	Philippe Aghion · Sergei Guriev · Kangchul Jo
8	한국의 화폐환상에 관한 연구	권오익 · 김규식 · 황인도
9	재원조달 방법을 고려한 재정지출 효과 분석 : 미국의 사례를 중심으로	김소영 · 김용건
10	The Impact of Geopolitical Risk on Stock Returns: Evidence from Inter-Korea Geopolitics	Seungho Jung · Jongmin Lee · Seohyun Lee
11	Real Business Cycles in Emerging Countries: Are Asian Business Cycles Different from Latin American Business Cycles?	Seolwoong Hwang · Soyoung Kim
12	우리 수출의 글로벌 소득탄력성 하락 요인 분석	김경근
13	북한의 경제체제에 관한 연구: 실태와 평가	양문수 · 임송

14	Distribution-Dependent Value of Money: A Coalition-Proof Approach to Monetary Equilibrium	Byoung-Ki Kim · Ohik Kwon · Suk Won Lee
15	A Parametric Estimation of the Policy Stance from the Central Bank Minutes	Dong Jae Jung
16	The Immigrant Wage Gap and Assimilation in Korea	Hyejin Kim · Chulhee Lee
17	Monetary Non-Neutrality in a Multisector Economy: The Role of Risk-Sharing	Jae Won Lee · Seunghyeon Lee
18	International Transmission of Chinese Monetary Policy Shocks to Asian Countries	Yujeong Cho · Soyoung Kim
19	The Impact of Robots on Labor Demand: Evidence from Job Vacancy Data for South Korea	Hyejin Kim
20	전공 불일치가 불황기 대졸 취업자의 임금에 미치는 장기 효과 분석	최영준
21	Upstream Propagation of the U.S.-China Trade War	Minkyu Son
제**2022**-1	Immigration and Natives' Task Specialization: Evidence from Korea	Hyejin Kim · Jongkwan Lee
2	Transmission of Global Financial Shocks: Which Capital Flows Matter?	Bada Han
3	Measuring the Effects of LTV and DTI Limits: A Heterogeneous Panel VAR Approach with Sign Restrictions	Soyoung Kim · Seri Shim
4	A Counterfactual Method for Demographic Changes in Overlapping Generations Models	Byongju Lee
5	Housing Wealth, Labor Supply, and Retirement Behavior: Evidence from Korea	Jongwoo Chung

6	Demand Shocks vs. Supply Shocks: Which Shocks Matter More in Income and Price Inequality?	Seolwoong Hwang · Kwangwon Lee · Geunhyung Yim
7	Financial Literacy and Mutual Fund Retail Investing: Evidence from Korea During the 2008 Financial Crisis	Jongwoo Chung · Booyuel Kim
8	Exchange Rate Regime and Optimal Policy: The Case of China	Yujeong Cho · Yiping Huang · Changhua Yu
9	북한 수출입단가지수 추정: 북중무역 데이터를 중심으로	이종민 · 김민정
10	탄소배출을 감안한 국가별 녹색 총요소생산성 분석	안상기
11	북한 소비자 지급수단 조사 및 분석	이주영
12	Selection into Outsourcing versus Integration Strategies for Heterogeneous Multinationals	Sangho Shin
13	Central Bank Digital Currency and Privacy: A Randomized Survey Experiment	Syngjoo Choi · Bongseop Kim · Young Sik Kim · Ohik Kwon
14	Technological Change, Job Characteristics, and Employment of Elderly Workers: Evidence from Korea	Jongwoo Chung · Chulhee Lee
15	Machine-Learning-Based News Sentiment Index (NSI) of Korea	Beomseok Seo · Younghwan Lee · Hyungbae Cho
16	빅데이터를 이용한 실시간 민간소비 예측	신승준 · 서범석
17	Fixed Effects Quantile Estimations with Extended Within Transformation and their Application	Ki-Ho Kim
18	글로벌 금융위기 이후 가계소비행태 변화 분석: 세대별 소비행태를 중심으로	최영준

ISSN 2287-6200

m.bok.or.kr