

Measuring News Sentiment of Korea Using Transformer*

Beomseok Seo** · Younghwan Lee*** · Hyungbae Cho****

We have developed the Korean news sentiment index (NSI) to gauge the economic sentiment of Korea on a daily basis by analyzing news texts gathered from the Internet. Our framework utilizes cutting-edge natural language processing techniques to compute the NSI and examine keywords, offering insights into its fluctuations. We designed a sentiment classifier using transformer neural networks that effectively process extensive news samples to compute the NSI of Korea. We compute the NSI more frequently and immediately than official indices that rely on monthly surveys. Through this, we can identify changes in economic sentiment before official statistics are released. Moreover, the proposed framework offers keyword analysis and sector indices to clarify why economic sentiments fluctuate. Our comprehensive assessments demonstrate that the NSI is a valuable leading index and an essential tool for identifying inflection points in economic sentiment.

JEL Classification: C45, C82, E32

Keywords: News Text Data, Natural Language Processing for Economic Analysis, Sentiment Shocks

Received: July 5, 2022. Revised: April 28, 2023. Accepted: Aug. 31, 2023.

* This work is a substantially revised version of the same authors' working paper Seo et al. (2022). We thank Dr. Hyejung Moon at the Bank of Korea for providing invaluable advice and encouragement during the development of this work. Additionally, we express our appreciation to the anonymous referees for their valuable comments and suggestions. Notably, the viewpoints presented in this paper solely belong to the authors and do not necessarily mirror those of the Bank of Korea. Any errors that may be present are attributable to the authors.

** First Author, Corresponding Author: Economist, Office of Economic Modeling and Policy Analysis, Bank of Korea, 55, Namdaemun-ro, Jung-gu, Seoul 04532, Korea. Adjunct Professor, Department of Quantitative Applied Economics, Sungkyunkwan University. Adjunct Professor, Department of Statistics, Korea University. E-mail: bsseo@bok.or.kr

*** Co-Author, Economist, Department of Economic Statistics, Bank of Korea, E-mail: yhleec@bok.or.kr

**** Co-Author, Junior Economist, Office of Economic Modeling and Policy Analysis, Bank of Korea, E-mail: hyungbae.cho@bok.or.kr

I. Introduction

Economists and central bankers are interested in developing a frequently and immediately available economic indicator to gauge economic sentiment (Barsky and Sims, 2012). Although traditional methods rely on surveys to investigate economic sentiment, recent advances in natural language processing (NLP) allow the use of unstructured text data to predict economic cycles (Armah et al., 2013; Bennani and Neuenkirch, 2017; Gentzkow et al., 2019; Athey and Imbens, 2019). In this context, this study proposes a novel Korean news sentiment index (NSI) that measures the economic sentiment of the Korean domestic economy. News sentiment is the tone (positive, negative, or neutral) of the economy reflected in news articles. News sentiment can capture public and expert opinions about financial markets and socioeconomic events appearing in economic news articles. We measure the NSI daily by computing news sentiment using news texts scrapped from the Internet and processing it with a set of NLP techniques. Unlike official economic sentiment indices, which are based on monthly surveys, the proposed index is more frequently and immediately compiled. Moreover, this index can help identify inflection points in economic sentiments before the official statistics are released. This study explains the overall process of building the Korean NSI using machine learning approaches and provides in-depth assessments of the newly computed index to evaluate its validity and utility from an economic perspective.

News text data are gaining attention as a new source of information owing to its rich and diverse content, including vast amounts of economic information on various topics, and its ability to spread information rapidly (Gentzkow et al., 2019; Moon, 2019). Several studies have used news texts in academia and public organizations, including central banks, to extract economic information, such as measuring economic uncertainty (Baker et al., 2016), improving the accuracy of economic cycle forecasting (Bybee et al., 2021; Seki et al., 2022), and predicting inflation expectations (Larsen et al., 2021).

Particularly, NSI is a widely studied topic in the realm of news text data (Shapiro et al., 2020). NSI is an economic index computed by counting the number of positive and negative sentences in news articles and compiling the difference between the two. Unlike traditional economic sentiment indices, which rely on surveys and are time-consuming and cost-intensive, NSI is based on readily available news articles from the Internet. Hence, this index provides the advantage of identifying economic sentiment more frequently and immediately with significantly lower costs. However, the process of computing NSI is not straightforward. Unlike survey-based sentiment statistics that undergo strict quality control processes to ensure index stability, extracting economic sentiments from news texts involves various complex problems that must be resolved to use NSI as a

reliable economic indicator.

The first challenge of using news text as a source of economic sentiment is that unstructured text data may contain more noise than structured survey data. Hence, large samples are preferred to compute a stable index to mitigate noise in text data. Otherwise, researchers must carefully select journalists who represent public opinion, which is another difficult issue to address. The second hurdle is consistently deciding the economic sentiment of news articles considering that the interpretation of news articles may differ according to the evaluators. These challenges become significant when a few evaluators must evaluate large samples of news articles. Hence, classifying all news articles to evaluate the economic sentiments of news texts is impractical. Conversely, using NLP and text mining technologies makes computing NSI highly feasible by providing a way to build a consistent classifier that can efficiently handle large samples of news texts.

The use of NLP and text mining techniques to develop NSI has primarily been investigated by researchers in central banks and international organizations who require quick assessments of the economic situation (Shapiro et al., 2020; Huang et al., 2019; Nguyen et al., 2020; Thorsrud, 2020, 2016; Babii et al., 2021; Jeon et al., 2020; Won et al., 2017; Kim et al., 2021). Central bank researchers have led the development of NSI, aiming to introduce a timely sentiment index as economic sentiments significantly impact the effectiveness of monetary policy (Shapiro et al., 2020). For instance, the Federal Reserve Bank of San Francisco utilizes a lexical approach to analyze news texts and release daily NSI (Shapiro et al., 2020). Similarly, the International Monetary Fund and the Reserve Bank of Australia have published works on NSI of different regions based on the lexical approach (Huang et al., 2019; Nguyen et al., 2020). By contrast, the Central Bank of Norway has employed unsupervised learning techniques to determine news indices by topics (Thorsrud, 2020). Academic researchers also actively investigate news texts to improve the accuracy of economic forecasting (Thorsrud, 2016; Babii et al., 2021). For instance, the Bank of Korea has studied text data to develop economic indices using a lexical dictionary for economics, topic modeling, and keyword analysis in news articles (Jeon et al., 2020; Won et al., 2017; Kim et al., 2021, 2019). Attempts have been made to develop a supplementary sentiment index of Korea using online news data. However, their work did not employ state-of-the-art NLP models to analyze Korean texts and conduct in-depth analyses of their correlation with macroeconomic variables and other sentiment indices.

Our proposed NSI differs from other works that used a lexical approach for English text data, as it is computed directly using a machine learning approach. For the rule-based lexical approach, the rules of classifying the sentiments of sentences are expressed by relatively simple patterns that can be explained by words appearing in sentences. The lexical dictionary can also be learned through machine learning as in Lee et al. (2019a,b) by learning the sentiments of each word with a statistical

model. However, in the lexical approach, the rule is always expressed by words, which is too restrictive to classify the sentiments of complicated sentences. From another aspect, using the machine learning approach directly on sentences allows us to build complicated patterns for classification rules. Moreover, this approach is likely to increase the prediction accuracy, although the patterns are projected on the new feature space and hence become highly difficult to interpret with a simple explanation. Moreover, considering that no known economic sentiment dictionary in Korean exists, such as the LM dictionary (Loughran and McDonald, 2011) for English, the machine learning approach can provide an alternative method to build a sentiment classifier.

To use a machine learning approach to Korean news text data, we begin by constructing large training samples consisting of randomly chosen news sentences and corresponding labels classified by humans. We then use these training samples to develop a precise classifier for the NSI of Korea. To achieve this, we design a new model based on the transformer neural network classifier, using its encoder structure for classification. We also compute the NSI by counting the positive and negative sentiments predicted by the new model.

We assess the validity and utility of the proposed index from multiple perspectives. In terms of validity assessments, we not only evaluate the classification accuracy of the new classifier but also investigate whether the newly proposed index can reflect the true economic cycle, even prior to the official statistics. For this purpose, we conduct a comparison analysis of the computed monthly NSI with other economic sentiment indices and real economic indices. Moreover, we examine the impulse response of the sentiment shocks of the NSI on macroeconomic variables based on a VAR model. We also address the utility of the proposed index. One of the greatest benefits of computing the NSI from news texts is the ability to investigate why the index fluctuates via its keyword analysis and sector indices. In addition, we investigate the temporal priority of the daily NSI by reviewing cases when the NSI reacts prior to official statistics.

This study provides contributions in two aspects. First, we provide the Korean NSI as a regular statistic, accompanied by an in-depth analysis of its validity and utility. The computed daily and monthly NSIs are now publicly available through the Economic Statistics System in the Bank of Korea (ecos.bok.or.kr). The NSI has been registered as an experimental statistic in Korea (No. 2022-001) and can be used for future research, such as developing forecasting models using textual data. Second, we present a practical framework for analyzing Korean text data with machine learning models to compile an economic index. The proposed framework is executed using an automated process facilitated by a Python-based program, demonstrating the potential to increase the efficiency of public work for compiling statistics. This framework could also be beneficial for developing new indices in other sectors.

The rest of the paper is organized as follows. Section 2 provides a detailed explanation of how we compute the Korean NSI using machine learning techniques and proposes a newly designed sentiment classifier model for Korean texts. Section 3 evaluates the validity of the proposed index by comparing it to other economic statistics. Section 4 examines the various utilities of the NSI, and finally, Section 5 summarizes and discusses the future usage of the NSI.

II. Compiling the Korean NSI

2.1. Concept of NSI

The NSI is computed by counting the number of positive and negative sentences in daily news articles. Therefore, efficient and accurate classification of sentiment in a vast amount of news articles is crucial for computing a precise NSI. We employ a machine learning approach to achieve this goal. The approach involves analyzing patterns in news sentences classified by humans to train a statistical model. The model is then applied to a new sentence to classify its sentiment. The key to this machine learning approach is building a sentiment classifier that fits the data pairs $(s, l)_{i=1}^N$ well, where $s_i \in S$ is a news sentence and $l_i = (p_i^{(0)}, p_i^{(1)}, p_i^{(2)})$ is its sentiment label classified by humans, with $p_i^{(0)} = P(s_i \text{ is positive})$, $p_i^{(1)} = P(s_i \text{ is negative})$, and $p_i^{(2)} = P(s_i \text{ is neutral})$. That is, given a model parameter vector θ , the sentiment classifier f is defined as follows:

$$f_\theta : S \rightarrow \{(1,0,0), (0,1,0), (0,0,1)\}. \quad (1)$$

2.2. News Text Data

The text data used as input in model (1) consists of news articles collected through web scraping. Web scraping is a method used to directly download publicly available data from the Internet. In this study, we collected news articles from an Internet news portal that classifies articles into six categories based on authors' choices: politics, economy, society, life/culture, IT/technology, and the world. We collected news articles only from the economy section to restrict the scope of the news sentiment we computed to economic sentiment. We collected all news articles released from 2005 on the news portal. The dataset comprises over 80 publishers and contains 18.3 million articles, with approximately 4,000 articles on normal business days and 1,000 articles on holidays.¹

Data quality can be a concern when collecting news articles through web scraping as the scraper may collect duplicate articles. To address this issue, we

exclude newly collected articles that are exact duplicates of any news in the database within the last 30 days. This process helps to ensure that only unique and relevant news articles are included in the database.

2.3. Preprocessing for News Text

The data should be preprocessed and converted into a numerical format that can be interpreted by the sentiment classifier in Eq. (1) to utilize machine learning on news text data. The preprocessing steps include establishing a data structure, tokenizing the text data using part-of-speech (POS) units (Webster and Kit, 1992), and converting the tokenized text data into a numeric format. These processes are achieved through sequential applications of text-mining techniques.

The first step in sentiment analysis is to determine the input structure, that is, whether to classify sentiments based on articles or sentences. Considering that news articles often contain positive and negative sentiments, classifying the sentiment of an entire article based on its text is not effective. To address this issue, the NSI is computed based on randomly sampled sentences from the news database, rather than using entire articles. In other words, the input data consist of each sentence $(s, l)_{i=1}^N$ that is randomly selected from the news database.

The second step in making a sentiment classifier understand text data is to tokenize the text in a sentence into POS units. POS is the smallest unit of a word that is assigned according to its syntactic functions in Korean. This process is known as POS tokenization. For instance, the Korean sentence “뉴스심리지수를 작성하였다.” is tokenized as follows:

‘뉴스심리지수’(proper noun) + ‘-를’(object case marker) +
 ‘작성’(general noun) + ‘하’(verb derivative suffix) +
 ‘-었’(pre-final ending) + ‘-다’(sentence-closing ending) + ‘.’(period).

Tokenizing sentences into POS allows distinguishing corpus tokens in different conjugated forms by transforming the tokens into the root form. Therefore, POS tokenization is an essential step to use text data for any statistical model. The reason is that POS tokenization is an essential step in using text data for any statistical model. For the collection of sequences of tokens, T , the POS tokenizer is a function, g , assigning a sentence, $s_i \in S$, to a sequence of tokens, $\tau_i \in T$, as

¹ Provider	Num. of articles	Examples
Economic newspapers	10.8m (59.0%)	Money Today, Asisa Economics, EDAILY, etc.
News agencies	4.8m (26.1%)	Yonhan News, Newsis, News1, etc.
Regular news	2.5m (13.8%)	Segye Ilbo, Donga Ilbo, Kyughyang Shinmun, etc.
Others	0.2m (1.1%)	Shindonga, Jungang Sunday, etc.

follows:

$$g : S \rightarrow T. \quad (2)$$

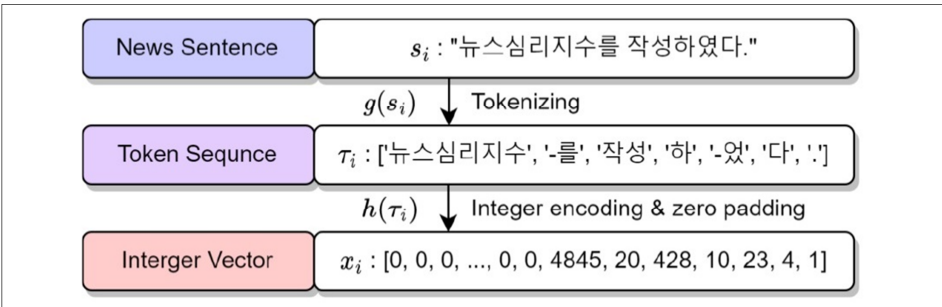
The choice of tokenizer is language-specific and depends on the field in which the tokenizer is used. The tokenizer determines whether a compound word should be split into multiple root words with separate meanings or interpreted as one word. The results of tokenization have a significant impact on the final classification performance. In our work, we use a pre-trained tokenizer called *Open Korea Text (OKT)* (Ryu, 2014) for g , which is available as an open-source software. Our text data for measuring economic sentiments include various fields and general expressions and have a large volume. Therefore, we choose the OKT tokenizer, which is widely used to analyze Korean texts for general purposes and shows high efficiency in terms of computational time.

Last, the tokenized data are transformed into numerical digits through integer encoding. The integer encoding assigns an integer to each distinct corpus token and represents an input sentence as a numeric vector. The maximum length of the numeric vector is limited to 80 for each sentence to maintain a uniform input data structure. For sentences shorter than 80 tokens, zeros are padded in front of the numeric vector, and for sentences longer than 80 tokens, the last tokens are truncated. The maximum length, $m = 80$, is chosen based on the 99th percentile of sentence lengths in the news database. The integer encoding assigns a sequence of tokens, $\tau_i \in T$, to an integer vector, $x_i \in \mathbb{R}^m$, through the following bijection function h .

$$h : T \rightarrow \mathbb{R}^m \quad (3)$$

The above-mentioned preprocessing steps are applied to any sentence input in the training phase for building a sentiment classifier and testing phase for computing daily NSI using the trained model.

[Figure 1] Preprocessing steps of Korean news texts to compute the Korean NSI



2.4. Construction of Korean News Sentiment (KoNS) Classifier

We use the state-of-the-art *transformer neural network* to build the sentiment classifier (1), which has been widely adopted in various NLP tasks. The transformer model is composed of a multi-head attention and a feed-forward network and is known to be capable of understanding the context of a sentence (Vaswani et al., 2017) owing to its multiple-head attention structure. This structure assigns high weights to the inputs that require intensive learning, which enhances the model's ability to capture meaningful relationships among words. We use the encoder structure of the transformer model to build the sentiment classifier for the NSI of Korea, which we call the *KoNS classifier*. Figure 2 shows a schematic diagram of the KoNS classifier.

The first step in the KoNS model is to convert the input integer vector x_i of size m into a matrix of real numbers $z_i \in \mathbb{R}^{m \times d_m}$ using the word2vec (Mikolov et al., 2013) mechanism. This mechanism represents each unique word with a d_m -dimensional real-valued vector so that semantically similar words are located close to each other in the embedded space. The embedded input matrix z_i is then fed into the self-attention layers as the query, key, and values.

The multi-head attention is computed using the following equations:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O, \quad (4)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (5)$$

$$\text{Attention}(Q', K', V') = \text{softmax}_m \left(\frac{Q'K'^T}{\sqrt{d_k}} \right) V', \quad (6)$$

where d_k is the output dimension of queries, Q' , and keys K' ; d_v is the output dimension of values, V' ; and h is the number of heads. For the output dimension of the embedding, d_m , $W_i^Q \in \mathbb{R}^{d_m \times d_k}$, $W_i^K \in \mathbb{R}^{d_m \times d_k}$, $W_i^V \in \mathbb{R}^{d_m \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_m}$ are the model parameters. As KoNS is a classification model, we used a self-attention mechanism, that is, the embedded input sequence, z_i , is used for all Q , K , and V . We set $d_k = d_v = d_m = 32$ and $h = 2$. Furthermore, Concat is a function that concatenates the input matrices along the second axis of the dimensions such that $\text{Concat}(x_1, \dots, x_h) = [x_1^T \dots x_h^T]^T$. softmax is a real-valued function that normalizes the input vector such that the output is a probability distribution over the elements and $\text{softmax}_m(x_j) = \frac{e^{x_j}}{\sum_{l=1}^m e^{x_l}}$, $\text{softmax}_m: \mathbb{R} \rightarrow (0,1)$ for $j = 1, \dots, m$ where m is the length of the input sequence.

The output of the multi-head attention is attached to the original embedded input and followed by layer normalization (Ba et al., 2016) to reduce computation time. The output is then fed into a feed-forward network with its output dimension

equal to $d_f = 32$ in the transformer block. The equation for the feed-forward network is as follows:

$$\text{FeedForward}(x) = \sigma(W^F x + b^F), \quad (7)$$

Here, $W^F \in \mathbb{R}^{d_m \times d_f}$, $b^F \in \mathbb{R}^{d_f}$, and $\sigma(\cdot)$ are the rectified linear unit activation function. The output of the previous layer is denoted by x in Eq. (7).

The output of the transformer block is then averaged along the first axis, which corresponds to the m number of words, and passed through two fully connected dense layers. The output dimensions of these layers are set to 20 and 3 to map the output to the three classes: positive, negative, and neutral. Finally, a softmax function is applied over the three classes to obtain the probability distribution for each class.

The configuration of the transformer neural network, including its hidden unit dimensions (d_k, d_v, d_m, d_f) , is important for the performance of the sentiment classifier. In our work, we follow the conventions used for the transformer model for similar text classification tasks. Although several hidden layers and units are required to express a complex function (LeCun et al., 2015), there is no theoretical reason to use more than two layers (Heaton, 2008). Therefore, we choose to use a comparable configuration for similar tasks and data (Chollet et al., 2015). For detailed aspects of the transformer neural network used to build KoNS, we refer the readers to the original paper by Vaswani et al. (2017).

Now, we express KoNS in Eq. (1) as the model \hat{f}_θ taking a news sentence $s_i \in S$ as input and predicting the probability of the sentiment, \hat{l}_i , through the sequential procedures of preprocessing and conducting the sentiment classification. That is,

$$\hat{l}_i = \hat{f}_\theta(s_i) \quad (8)$$

$$= \text{TransformerBasedClassifier} \circ h \circ g(s_i), \quad (9)$$

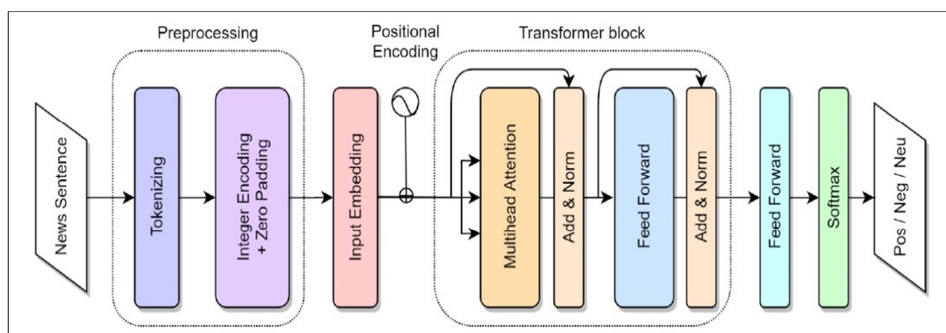
where \hat{l}_i is the predicted value such that $\hat{l}_i \in \{(\hat{p}_i^{(0)}, \hat{p}_i^{(1)}, \hat{p}_i^{(2)})\}$, and $0 \leq \hat{p}_i^{(c)} \leq 1$ for $c=0,1,2$, and $\sum_{c=0}^2 \hat{p}_i^{(c)} = 1$, and h and g indicate the integer embedding process and tokenizer, respectively.

The model parameters, θ , are estimated by minimizing the cross-entropy loss function so that the difference between the observed label, $\{(p_i^{(0)}, p_i^{(1)}, p_i^{(2)})\}$, and the predicted probability, $\{(\hat{p}_i^{(0)}, \hat{p}_i^{(1)}, \hat{p}_i^{(2)})\}$, becomes small for each sentence, $s_i, i=1, \dots, N$.

$$L(\theta | \{(s_i, l_i)\}_{i=1}^N) = - \sum_{i=1}^N \sum_{c=0}^2 p_i^{(c)} \log \hat{p}_i^{(c)}. \quad (10)$$

In addition to the KoNS sentiment classifier, we build another model with the same structure of (1), but this time, to classify the geographical scope of the sentences into one of the three categories based on their contents: domestic, foreign, and both combined. We use only the sentences classified as domestic to compute the Korean NSI. This process excludes foreign news articles from the compilation of the Korean NSI as they often show different patterns from the domestic economic sentiment. For simplicity, we assume hereafter that any data inputs for KoNS are cleaned with the geographical scope classifier in advance and consist of only domestic sentences.

[Figure 2] Transformer-based sentiment classifier for the Korean NSI



2.5. Training Data and Model Training

A training dataset with correct labels is crucial to estimating KoNS. We combined data from three separate projects to create a training dataset, which comprised a total of 446,478 sentences. In the first project, conducted in 2019, we labeled 230,583 news sentences sampled from publications between January 2008 and December 2018. In 2020, we added 84,069 sentences to the dataset based on news published between January 2008 and May 2020. In our most recent effort, we generated an additional 131,826 labeled sentences using news published between January 2005 and June 2021. We ensured that no duplicate sentences existed in the training data.

The training sentences were subjected to sentiment classification by 16 trained individuals who were hired explicitly for the labeling task. Each sentence was classified into one of three categories, namely, positive, negative, or neutral sentiment. We enlisted additional reviewers, economists from the Bank of Korea, to review the sentiment labels after the initial classification to minimize measurement errors.

However, given the complexities of economic subjects and contexts, certain sentences presented challenges in their sentiment classification. For example, inflation can have positive and negative impacts on the economy, rendering it

difficult to classify as either positive or negative. In response to this issue, we established minimal guidelines for frequently occurring economic subjects, which we outline in Table 1. These guidelines were designed to provide a framework for the sentiment classification of economic subjects that present challenges, thereby promoting consistency in the labeling process.

[Table 1] Minimal guidelines for sentiment classification of economic news sentences to build the training data

Subjects	Guidelines
Stock price, interest rates, exchange rates, and other asset prices	<ul style="list-style-type: none">• A sentence simply reading the change of the index is neutral.• If the change is mentioned with specific reasons, the sentence is classified as positive or negative depending on its context.
Industry and companies	<ul style="list-style-type: none">• Advertisements are neutral.• A simple statement on historical facts is neutral.• If past events are mentioned with precise implications for the present situation or future expectations, the sentence is classified as positive or negative.• Despite a brief remark, if a sentence mentions the change in financial statements, it is classified as positive or negative.
Real estate and construction	<ul style="list-style-type: none">• Advertisements for apartment sales are neutral.• Statements on the unsold apartments are classified as positive or negative.
Government and public sectors	<ul style="list-style-type: none">• Government supports are classified as positive or negative.• Government regulations are classified as positive or negative when specific expected impacts are indicated.

Note: The subjectivity in sentiment interpretation is unavoidable because the economic sentiment is indeed abstract.

Table 1 provides guidelines that limit the interpretation of sentiment and result in a cautious classification by highlighting the clear tones found in the training data. Consequently, approximately 80% of the sentences in the training data are classified as neutral, whereas the remaining 20% are divided equally between positive and negative classifications. Economic sentiment is an abstract concept, and therefore, subjectivity in sentiment interpretation is inevitable when constructing training data for machine learning approaches. Even with traditional lexical approach and surveying methods, subjectivity is present as individuals use different criteria to evaluate their economic situation.

Notably, the imbalanced class weights of the training data can affect the final prediction of a classifier. However, as we do not possess the labels of the daily news data, we cannot guarantee that the class weight of the daily news is consistent with the training dataset. We modify the loss function in Eq. (10) by incorporating class weights, $w^{(c)} = \frac{N}{3N^{(c)}}$, where $N^{(c)}$ denotes the number of sentences in class c of the training data, to address this issue and account for the fluctuation of the computed

NSI.

$$L(\theta | \{(s_i, l_i)\}_{i=1}^N) = - \sum_{i=1}^N \sum_{c=0}^2 w^{(c)} p_i^{(c)} \log \hat{p}_i^{(c)}. \quad (11)$$

Table 2 shows the comparison between using the loss with and without the weight adjustment.

2.6. Prediction for Daily News Data

Finally, we apply the KoNS classifier to daily news data to compile the NSI. We collect approximately 4,000 news articles from the Internet per weekday as of 2021, which is equivalent to approximately 70,000 sentences. The number of sentences decreases to roughly 10,000 per day on weekends. We randomly sample 10,000 sentences every day and use only the sampled sentences to compute the NSI to reduce computing costs and stabilize the NSI across days. This sampling process prevents certain days from dominating the economic sentiment of the NSI when many news articles are released on those days.

2.7. Computing the Korean NSI

The daily NSI is calculated by first predicting the sentiment of daily news sentences using KoNS. We then count the number of positive and negative sentences from the sample collected during the previous seven days. The daily NSI is computed as the ratio of the difference and sum of the positive and negative counts. Using a seven-day period of data ensures a smooth and stable index. Similarly, the monthly NSI is computed by the same procedure but based on news sentences collected during the corresponding month.

Daily and monthly NSIs are standardized using their long-term averages to facilitate comparison with past economic sentiment. The NSIs are standardized such that their averages and standard deviations are equal to 100 and 10, respectively. The standardization interval starts from 2005 and extends to the end of the previous year at the beginning of each new year. The computation process for the composite consumer sentiment index (CCSI) and economic sentiment index (ESI) is also standardized in the same manner.

Specifically, the daily NSI of a particular day t , denoted as $NSI_t^{(\text{daily})}$, is calculated as follows:

$$NSI_t^{(\text{daily})} = \left(\frac{X_t - \bar{X}}{S} \right) \times 10 + 100, \quad (12)$$

$$\text{where } X_t = \frac{\sum_{u=1}^7 P_{t-u} - \sum_{u=1}^7 N_{t-u}}{\sum_{u=1}^7 P_{t-u} + \sum_{u=1}^7 N_{t-u}}, \quad (13)$$

$$\bar{X} = \frac{1}{|U|} \sum_{u \in U} X_u, \quad S = \sqrt{\frac{1}{|U|-1} \sum_{u \in U} (X_u - \bar{X})^2}. \quad (14)$$

Here, P_t and N_t respectively denote the count of positive and negative sentences gathered during period t . U represents the index set of days for the standardized interval, which corresponds to the days between January 1, 2005, and the last day of the previous year. Without loss of generality, the monthly NSI, $NSI_v^{(\text{monthly})}$, for a particular month, v , is compiled in the same way but by replacing U with the index set of months V , which includes all months between January 2005 and December of the previous year, and X_t with X_v as follows.

$$X_v = \frac{\sum_{v \in M_v} P_v - \sum_{v \in M_v} N_v}{\sum_{v \in M_v} P_v + \sum_{v \in M_v} N_v}, \quad (15)$$

where M_v is the index set of days belonging to the particular month v . That is, the monthly NSI is calculated by analyzing the sentences published during the corresponding month, whereas the daily NSI is calculated using news data collected during the previous seven days. The reason is that the daily NSI can be highly volatile, and using a seven-day period of data helps to generate a stable index.

A NSI greater than 100 indicates that the economic sentiment in the analyzed news articles is more optimistic than the past average, whereas a NSI less than 100 indicates a more pessimistic sentiment.

2.8. Automated Compilation

NSI has been designed to be compiled automatically every week without the need for human labor. The entire process, from web-scraping of daily news articles from the Internet to text preprocessing, sentiment prediction, and compilation of the daily and monthly NSIs, is carried out using a Python script that runs on an automated batch. The script collects news articles from the Internet every day, from the last point of the database to the previous day at 5 AM, and then uses KoNS to compute the daily and monthly NSIs. This automation of the compiling process improves efficiency, saves labor, and reduces costs compared with the survey approach.

III. Validity Assessments of NSI

In this section, we evaluate the newly computed Korean NSI from multiple perspectives. The validation of NSI is focused not only on the accuracy of the sentiment classification but also on whether it implies the true economic sentiment and even indicates the truth prior to other official statistics. Therefore, we validate NSI by sentiment classification accuracy, comparative analysis with other economic indicators based on the cross-correlation, Granger causality tests, and impulse response analysis using a VAR macroeconomic model.

3.1. Sentiment Classification Accuracy

We compared the accuracy of the KoNS classifier against the lexical approach and other statistical models using 10-fold cross-validation.² We translated the sentiment of 50,754 English keywords from Shapiro et al. (2020) into their corresponding Korean words to implement the lexical approach. The reason is that no sentiment dictionary exists for economic keywords in Korean. Table 2 shows the classification accuracy of KoNS and other standard methods. We compare only the positive and negative classes because the neutral class does not affect the computation of NSI. Table 2 indicates that KoNS outperforms the other methods that were evaluated. The lexical approach demonstrates inferior accuracy compared

² The performance metrics have been defined as follows. Denote the true sentiment label and predicted sentiment label of i th sentence by l_i and \hat{l}_i respectively. Since we are concentrating on the case where the sentiment of the sentence is either positive or negative, \hat{l}_i, l_i take values from the set {positive, negative}. Our performance metrics include accuracy, sensitivity, specificity, and precision, which we define as follows:

$$\begin{aligned}\text{Accuracy} &= \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \\ \text{Sensitivity} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \\ \text{Specificity} &= \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \\ \text{Precision} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}\end{aligned}$$

where the term “True Positives” refers to the number of samples where the predicted label (\hat{l}_i) and the true label (l_i) are positive. Conversely, “False Positives” denotes the number of samples where the predicted label is positive (\hat{l}_i), but the true label is negative (l_i). Similarly, the terms “True Negatives” and “False Negatives” are defined based on the number of samples where the predicted and true labels are negative or where the predicted label is negative, but the true label is positive, respectively. The F1 score is a balanced performance metric of classification, defined as a harmonic mean of sensitivity and precision that is $[\frac{1}{2}(\frac{1}{\text{Sensitivity}} + \frac{1}{\text{Precision}})]^{-1}$.

with the machine learning approach, even when using simple linear regression. One interesting finding is that the simple linear regression achieves a competitive accuracy to the sophisticated NLP models. The main reason is that the news text sentences are relatively straightforward, and their sentiment is easily revealed based on the specific words in sentences rather than being influenced by complicated sentence structures or contextual implications.

[Table 2] Ten-fold cross-validation accuracy for the binary classification of the positive and negative classes

	LEX	LR	FFN	SVM		KoNS	
		SL	SL	SL	WAL	SL	WAL
Accuracy	0.82	0.85	0.85	0.86	0.81	0.97	0.92
	(0.027)	(0.016)	(0.015)	(0.026)	(0.017)	(0.009)	(0.022)
Sensitivity	0.81	0.84	0.90	0.81	0.91	0.96	0.91
	(0.055)	(0.043)	(0.042)	(0.038)	(0.051)	(0.019)	(0.035)
Specificity	0.83	0.86	0.76	0.86	0.80	0.98	0.93
	(0.024)	(0.032)	(0.061)	(0.044)	(0.032)	(0.008)	(0.014)
Precision	0.78	0.85	0.83	0.85	0.80	0.98	0.93
	(0.067)	(0.032)	(0.034)	(0.044)	(0.043)	(0.012)	(0.029)
F-1 score	0.80	0.85	0.79	0.86	0.80	0.98	0.93
	(0.036)	(0.020)	(0.026)	(0.037)	(0.021)	(0.009)	(0.020)

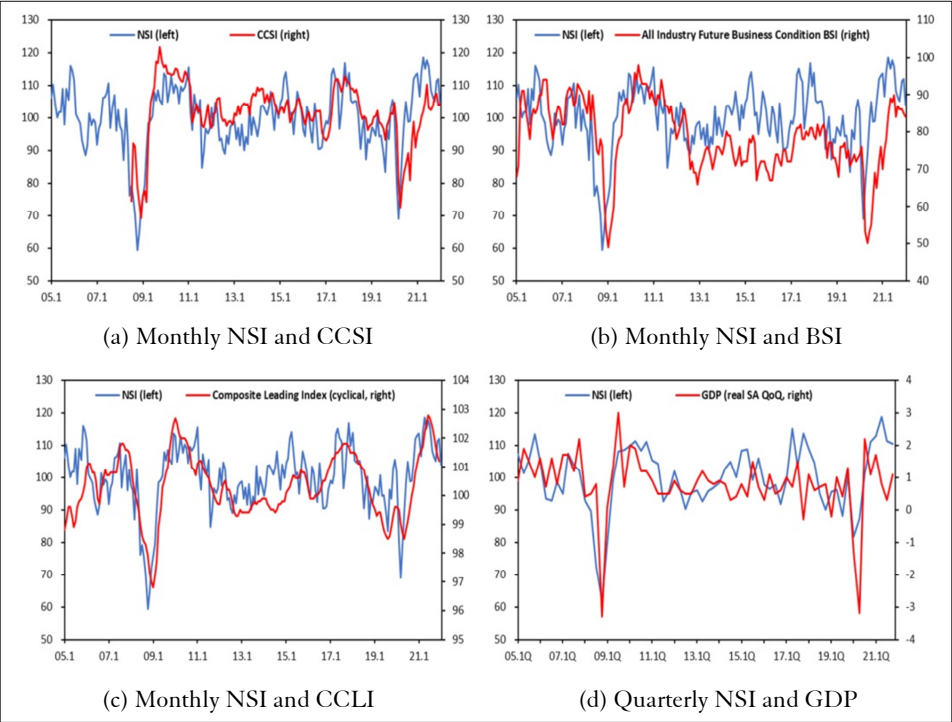
Note: The neutral class has been ignored in predicted and true labels. LEX: lexical approach, LR: logistic regression, FFN: feed-forward network,³ SVM: support vector machine, SL: standard loss used, WAL: weight-adjusted loss used.

3.2. Comparative Analysis with Economic Indicators

The computed NSI is compared with various official statistics. First, the following economic sentiment indicators are compared with NSI: consumer survey index (CSI), CCSI, business survey index (BSI) representing the economic sentiment of entrepreneurs, and ESI, which is a composite index of CSI and BSI. As these economic sentiment indicators are compiled based on monthly surveys, we compare them to monthly NSI. The comparative analysis is conducted based on the available data until December 2021. Notably, BSI and ESI are available since January 2005 and CCSI and CSI since July 2008. In Table 3, the monthly NSI is leading most of the economic sentiment indicators by one to two months with high correlation according to the cross-correlation analysis. Particularly, the monthly NSI shows a high correlation of 0.75 with CCSI leading it by one month. NSI is also compared with real economic indicators. The monthly NSI is leading the

³ Feed-forward network (FFN) is a specific structure of neural networks that consist of layers connected by an affine transformation and an activation function without any recurrent connection.

[Figure 3] Comparisons between NSI and various economic indicators



[Table 3] Cross-correlation analysis results between the NSI and major economic indicators

Economic indicators		Max. Corr. (Lag)	
		KoNS	LEX
CCSI		0.75 (-1)	0.62 (-1)
ESI		0.61 (-2)	0.27 (-1)
CSI	Living Standards of Household	0.74 (-1)	0.80 (-1)
	Domestic Economic Situation	0.73 (-1)	0.54 (-1)
	Expectations of Living Standard	0.73 (-1)	0.63 (-1)
	Expectations of Domestic Economic Situation	0.70 (-1)	0.45 (0)
	Expectation of Household Income	0.68 (-1)	0.64 (-1)
	Spending Plan	0.57 (-1)	0.42 (-1)
	Expectation of Employment Situation	0.72 (-1)	0.54 (-1)
BSI	All Industries Business Condition	0.64 (-1)	0.23 (-1)
	All Industries Profitability	0.68 (-1)	0.49 (-1)
	All Industries Financial Situation	0.64 (-1)	0.33 (-1)
	All Industries Future Business Condition	0.61 (-2)	0.20 (-2)
	All Industries Future Profitability	0.65 (-2)	0.44 (-2)
	All Industries Future Financial Situation	0.61 (-2)	0.29 (-2)
Real indices	KOSPI (monthly closing price YoY%)	0.68 (-1)	0.26 (-1)
	Cycle of Composite Leading Index (CCLI)	0.76 (-2)	0.36 (-1)
	GDP (real SA QoQ%)	0.53 (0)	0.32 (0)

cyclical composite leading index (CCLI) by two months with its correlation equal to 0.76. The quarterly NSI is coincident with the gross domestic product (GDP), which is measured by the quarter-on-quarter (QoQ) change in the seasonally adjusted real GDP, with its correlation equal to 0.53.

Figure 3 shows that the monthly NSI has a similar trend to CCSI and all industry future business condition BSI. In addition, the NSI hits the lowest point one to two months earlier than the official statistics during the global financial crisis in 2008 and the COVID-19 crisis in 2020.

3.3. Granger Causality Tests

Furthermore, we computed NSI using the LEX approach and compared it to NSI calculated using the KoNS classifier. As shown in Table 3, the correlations of NSI to other sentiment indicators and macroeconomic variables were significantly low for the lexical approach. Although the lexical approach exhibited a relatively high correlation with the CCSI, it displayed weak correlations with the BSI and real indices. This limitation is due to the lexical approach's use of a restricted set of keywords to describe business conditions.

We also performed the Granger causality test on those economic indicators that appear to be led by NSI from the cross-correlation analysis to test if the NSI leads economic indicators. For a fair comparison, we only consider data between July 2008 and December 2021, as CCSI and CSI data are not available before July 2008. To determine whether variable x Granger-cause variable y , we use the following model:

$$y_t = \alpha_0 + \sum_{l=1}^L \alpha_l y_{t-l} + \sum_{l=1}^L \beta_l x_{t-l} + \varepsilon_t.$$

The model is estimated using OLS, and causality is tested under the null hypothesis $\beta_l = 0$ for all l . The decision rule is if the null hypothesis is rejected or not. That is, if the null hypothesis is not rejected, then x does not seem to cause y . The test is performed with the Wald test procedure, and L is set to 2.

Table 4 summarizes the test results, providing statistically significant evidence for the Granger causality of the NSI to economic indicators, except for a few (KOSPI and CCLI). For most economic indicators, particularly for the economic sentiment-related indicators, the null hypothesis that “the NSI does not cause the economic indicator” is rejected at the 0.1% significance level. By contrast, the null hypothesis that “the NSI does not cause the economic indicator” is not rejected at the 1% significance level. This result is robust to alternative specifications, where L is set to 1 or 3.

However, notably, the Granger causality test only measures the statistical

predictability of one variable over another and does not imply that news articles are the direct cause of variations in economic sentiment. Identifying the channels through which news influences the beliefs of economic agents is beyond the scope of this study. Nevertheless, the results support the view that the NSI is useful for predicting economic sentiment.

[Table 4] Summary of Granger causality test results

Y	N of obs.	Null hypothesis	
		NSI \nRightarrow Y	Y \nRightarrow NSI
CCSI	160	45.5*** (< 0.1%)	1.6 (0.46)
ESI	160	36.8*** (< 0.1%)	1.4 (0.49)
Living Standards of Household	160	45.3*** (< 0.1%)	0.6 (0.75)
Domestic Economic Situation	160	34.2*** (< 0.1%)	0 (0.98)
Expectations of Living Standard	160	35.1*** (< 0.1%)	1 (0.59)
Expectations of Domestic Economic Situation	160	23.8*** (< 0.1%)	2.4 (0.3)
Expectations of Household Income	160	48.1*** (< 0.1%)	6.8** (0.03)
Spending Plan	160	36.2*** (< 0.1%)	4.1 (0.13)
Expectations of Employment Situation	160	26.6*** (< 0.1%)	0.9 (0.65)
All Industries Business Condition	160	9.1** (0.011)	2.5 (0.28)
All Industries Profitability	160	18.2*** (< 0.1%)	1.2 (0.55)
All Industries Financial Situation	160	15.3*** (< 0.1%)	0.9 (0.63)
All Industries Future Business Condition	160	46.1*** (< 0.1%)	1.3 (0.53)
All Industries Future Profitability	160	41.2*** (< 0.1%)	1.9 (0.39)
All Industries Future Financial Situation	160	53.6*** (< 0.1%)	0.3 (0.84)
KOSPI (monthly closing price YoY%)	160	3.1 (0.22)	3.8 (0.15)
Cycle of Composite Leading Index (CCLI)	160	16.3*** (< 0.1%)	10.2*** (0.01)

Note: The elements of the third and fourth columns are the value of χ^2 statistics under the given null hypothesis. Here, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, and the values in parenthesis are the corresponding p-values.

3.4. Impulse Response Analysis

We validate the impact of economic sentiment shocks measured by NSI using a macroeconomic model and comparing it to ESI shocks. We build a standard VAR system following the procedure of van Aarle and Kappler (2012), in which they study the interaction between confidence indicators and macroeconomic adjustments using four variables: unemployment (UNE), industrial production (IND), retail sales (RET), and ESI. In our study, we use the monthly unemployment rate, industrial production index, and retail business service index for the variables. All the variables are seasonally differenced by using year-on-year (YoY) growth rates. Let t denote time. Then, the structural representation of the considered model is as follows:

$$Cy_t = \alpha + \sum_{i=1}^k C_i y_{t-i} + \varepsilon_t, \quad (16)$$

where $y_t = \begin{pmatrix} IND_t \\ UNE_t \\ RET_t \\ NSI_t \end{pmatrix}$ and $\varepsilon_t = \begin{pmatrix} \varepsilon_t^{IND} \\ \varepsilon_t^{UNE} \\ \varepsilon_t^{RET} \\ \varepsilon_t^{NSI} \end{pmatrix}$.

Here, y_t denotes the vector of endogenous variables, and ε_t denotes the vector of residuals. We assume C^{-1} has the recursive structure and its reduced form errors $\varepsilon_t = C^{-1} \varepsilon_t$, which indicates that C^{-1} has the lower triangular structure.

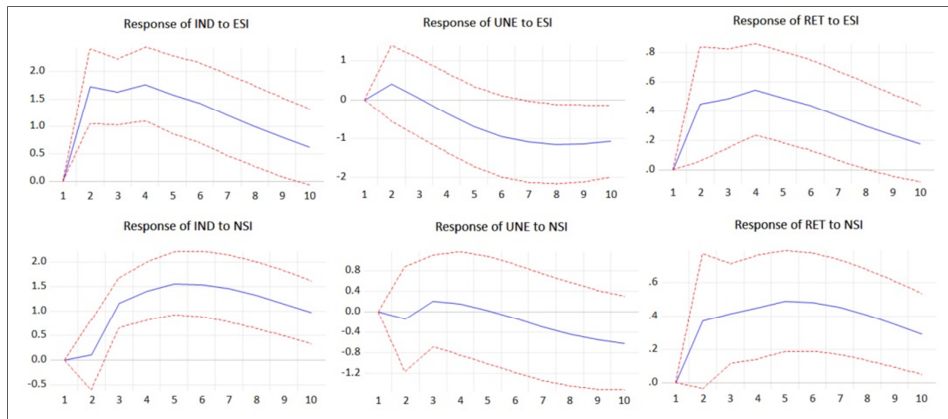
[Table 5] Estimation results of the VAR model for macroeconomic variables with sentiment indicators (SI): NSI and ESI

	ESI model			NSI model		
	IND	UNE	RET	IND	UNE	RET
SI_{t-1}	0.314* (0.060)	0.075* (0.089)	0.082** (0.035)	0.012** (0.044)	-0.018* (0.063)	0.046** (0.025)
SI_{t-2}	-0.181* (0.060)	-0.114* (0.090)	-0.043** (0.036)	0.115** (0.044)	0.056* (0.064)	-0.004** (0.025)
Adj. R^2	0.586	0.526	0.321	0.557	0.524	0.324

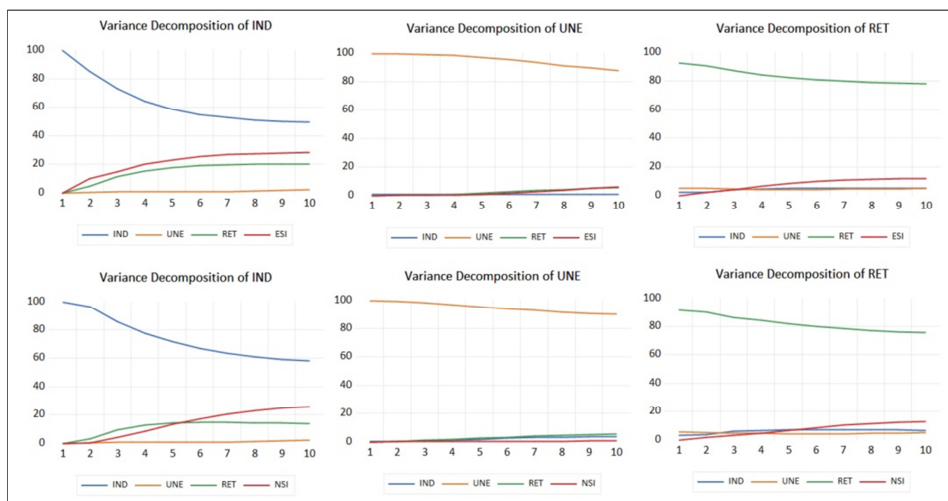
Note: * and ** indicate that the statistics are significant at 0.10 and 0.05, respectively.

We recover the orthogonalized shocks with $\varepsilon = Ce_t$, where $C = chol(\Sigma)$, which is the Choleski decomposition of the covariance matrix of residuals. We checked the robustness of the model by changing the ordering of the variables, but the estimated impulse responses had little change. Table 5 shows the estimation results of the models with NSI and ESI. The table demonstrates that NSI has very similar explanatory power to ESI and has the adjusted R^2 close to that of the ESI model. Figure 4 shows the impulse responses of the two models. The impulse response of the industrial production (IND) to NSI implies that the production is boosted by the increase in economic confidence with its hike appearing in three months. By contrast, the ESI shocks on industrial production show the largest hike in two months, which supports the same result that NSI is leading ESI. The variance error decompositions, displayed in Figure 5, also indicate that the sentiment shocks measured by NSI largely contribute to the variance in the business cycle variable. Moreover, for the other two economic variables, the impact of NSI is less clear, which coincides with the findings of van Aarle and Kappler (2012), who had similar conclusion using ESI in the euro area.

[Figure 4] Impulse responses of macroeconomic variables on the economic sentiment shocks measured by ESI and NSI



[Figure 5] Variance error decomposition for the VAR models with ESI and NSI



Note: The sentiment shocks measured by ESI or NSI largely contribute to the variance in the business cycle variable, IND.

IV. Utility Assessments of NSI

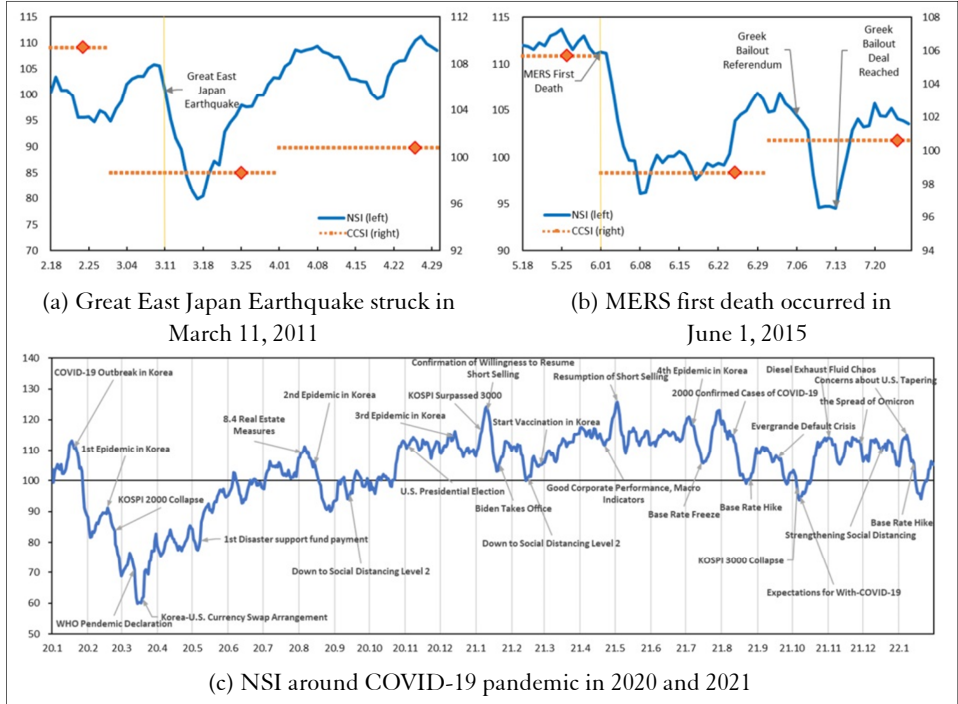
In this section, we present utility assessments of NSI. One of the most significant advantages of NSI is its timeliness as a daily economic indicator. We compare its ability to identify inflection points in economic sentiment to that of other official statistics to test the temporal priority of daily NSI. Additionally, NSI provides great explainability by providing detailed information about probable factors of fluctuation through keywords and metadata. We analyze NSI using keyword

analysis and sectoral NSIs computed for three economic sectors.

4.1. Temporal Priority

NSI has the advantage of being compiled daily, allowing it to quickly identify changes in economic sentiment before monthly official statistics based on surveys are released. As shown in Figure 6, NSI can immediately quantify the impact of important economic events and effectively detect inflection points in economic sentiment. For example, in 2011, when the Great East Japan Earthquake struck in early March, NSI dropped sharply and then rebounded soon after the event. By contrast, the drop in the Composite Consumer Sentiment Index (CCSI) was observed at the end of March when the statistics were released after the monthly survey was conducted. Similarly, in 2015, NSI dropped immediately after the first death from MERS was known to the public on June 1. However, CCSI could not indicate the event until the end of June. Thus, NSI provides immediate information earlier than survey-based official statistics and can be used as a supplementary indicator to quickly detect inflection points in economic sentiment.

[Figure 6] NSI with major economic events

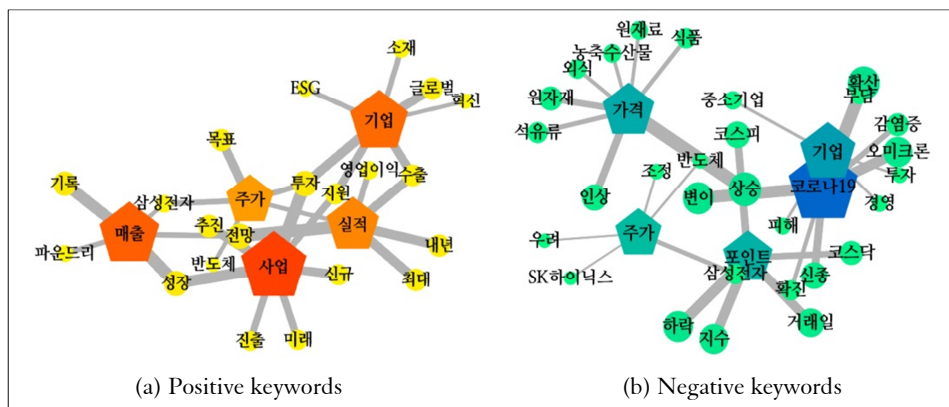


Note: In panels a and b, the red bullets indicate the day the monthly CCSI is released. These figures demonstrate that NSI detects the inflection points of the economy effectively and even prior to official monthly statistics.

4.2. Explainability and Keyword Analysis

Another advantage of NSI is that the analysis of keywords reveals the factors behind NSI fluctuations. Figure 7 displays the keyword networks in positive and negative news articles, which were classified by KoNS in the last week of November 2021. The positive keywords of the period mainly consisted of company earnings, whereas the negative keywords were related to the price hike of raw materials and the COVID-19 Omicron mutation. The keywords can also disclose the main economic issues at a given time. For instance, in the last week of November 2021,

[Figure 7] Keyword networks of positive and negative sentences classified by KoNS



Note: The pentagon nodes indicate the five top main keywords appearing in the sentences in the last week of November 2021, and the circle nodes indicate the related keywords appearing together with the main keywords.

[Table 6] The most appeared keywords for sector NSIs

Sec.	Positive sentences		Negative sentences	
	Related words	Keywords	Keywords	Related words
Mac.	(증가, 11 월, 최고, 실적) (분기, 증가, 가락, 성장) (투자, 자원, 글로벌, 성장)	수출 매출 사업	코로나19 가격 생산	(확산, 변이, 오미크론, 위기) (상승, 석유류, 외식, 원자재) (감소, 자동차, 광공업, 반도체)
Fin.	(상승, 전망, 실적, 삼성전자) (자수, 외국인, 상승, 코스닥) (성장, 투자, 확대, 미래)	주가 코스피 기업	코스피 주가 거래	(하락, 지수, 오미크론, 최저) (하락, 삼성전자, 업황, 우려) (포인트, 하락, 코스피, 지수)
Ind.	(투자, 추진, 소재, 친환경) (분기, 증가, 성장, 파운드리) (증가, 반도체, 11 월, 실적)	사업 매출 수출	코로나19 기업 업계	(확산, 신종, 장기, 변이) (피해, 중소기업, 과징금) (항공, 반도체, 조선, 자동차)

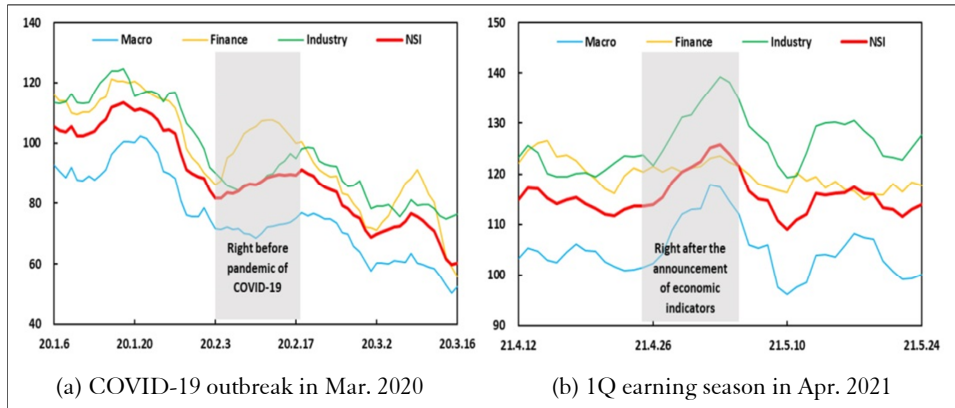
Note: The most appeared keywords and their related words in news sentences were grouped into one of the positive and negative sentiments in three sectors: macro, finance, and industry. The bar plots indicate the counts of keywords appearing in the groups of sentences. The data are as of the last week of November 2021.

“Omicron” mutation, “KOSPI,” and the government’s deliberation on the “budget” for supporting small business owners were mentioned more than in the previous week, indicating the main economic issues of the week. Table 6 presents the keywords of the same week in positive and negative news articles divided by sub-sectors: macro, finance, and industry. The table illustrates that the impact of stock market fluctuations was mentioned in positive and negative articles in the finance sector. In the macro sector, the rise of exports in November was largely responsible for the rise of NSI, whereas the worsening industrial production statistics were attributed as a factor for the decline in NSI. Such detailed information on economic sentiments is challenging to obtain through surveys.

4.3. Sector NSIs

NSI has the advantage of computing sector-wise sentiment easily by dividing news sources into sectors. This is a benefit of using news articles to compute economic sentiment considering that sector information is obtained as metadata when collecting the news articles. Internet portals categorize news articles, and hence, we can use these categories to divide news sources into three sub-categories: macro, finance, and industry, and compute sector NSIs. As shown in Figure 8, although the sector NSIs are highly correlated with the aggregated NSI, they exhibit different patterns at particular time points when specific events occur, such as the COVID-19 outbreak and company earnings announcements. Sector NSIs provide additional

[Figure 8] Sector NSIs with economic events



Note: The sector NSIs provide additional detailed information about the economic sentiments. Right before the COVID-19 pandemic in March 2020, the finance sector NSI highly increased owing to the earnings surprises of financial holdings companies in Korea; Nevertheless, other sector NSIs remained flat because of the concern for the new virus. In April 2021, macro and industry sector NSIs rose after 1Q GDP of Korea was announced higher than anticipated.

[Table 7] The correlation coefficients between the sector NSIs and the aggregated NSI.

Sector	After 2015			After 2018			After 2020		
	Mac.	Fin.	Ind.	Mac.	Fin.	Ind.	Mac.	Fin.	Ind.
Macro	1.00	0.75	0.76	1.00	0.85	0.84	1.00	0.88	0.92
Finance	-	1.00	0.67	-	1.00	0.71	-	1.00	0.79
Industry	-	-	1.00	-	-	1.00	-	-	1.00
Aggregated	0.94	0.89	0.88	0.97	0.91	0.91	0.98	0.92	0.95

Note: The correlations are computed based on the daily indices.

detailed information on economic sentiment in different sectors. Table 7 shows that the sector NSIs became highly homogeneous after the COVID-19 outbreak. This result indicates that the COVID-19 pandemic dominated the factors in economic sentiment fluctuations after its outbreak. Obtaining such information through surveys would significantly increase costs by adding questions.

V. Summary and Discussion

In this study, we present the development and evaluation of a NSI for Korea, which is computed based on news articles collected from the Internet using web-scraping techniques. We develop a KoNS classifier using a state-of-the-art NLP model to compute NSI. The encoder structure of the transformer model is utilized, and the model is trained with a modified loss function to control imbalanced data. Over 450,000 training sentences labeled by 16 trained personnel are used to train KoNS. Daily and monthly NSIs are computed based on the counts of the positive and negative sentences classified by KoNS for daily news sentence samples.

The validity of the computed NSI is evaluated in multiple perspectives. The cross-correlation analysis shows that NSI leads the CCSI by one month with a correlation of 0.75 and the CCLI by two months with a correlation of 0.76. Additionally, the impulse response analysis shows that a rise in NSI stimulates industrial production, with the hike appearing in three months.

Furthermore, we evaluate the various utilities of NSI. Daily NSI is found to correctly detect inflection points in economic sentiments before official statistics based on monthly surveys. Keyword analysis reveals the factors that cause NSI fluctuations and provides additional information that can be quantified through survey-based statistics. Last, sector NSIs are easily computed by dividing news articles into sectors, and they provide further detailed information on economic sentiments in different sectors.

Our contributions are twofold. First, we propose a new economic index, the Korean NSI, and comprehensively evaluate its validity and utility. Second, we provide a practical framework for analyzing Korean text data with machine learning

approaches to compile an economic index. Our framework addresses various issues related to using text data as a source of a new economic indicator, such as building a Korean text classification model, adjusting imbalanced data, and standardizing stability control of the index. Furthermore, we provide a framework to improve the efficiency of public work for compiling public economic statistics automatically without human intervention. As an experimental attempt to compile regular economic statistics without surveying, NSI demonstrates that efficient observational studies can complement traditional surveying methods. The proposed process of automatic compilation of NSI can be applied to invent similar economic indices in different fields.

For future work, there are several interesting directions to explore using NSI. One potential avenue is to investigate the predictive power of the text-based index by incorporating it into existing economic forecasting or nowcasting models. Additionally, efforts to construct comparable text-based indices in different sectors and fields, such as production, employment, and inflation, would be beneficial. We hope that our work can inspire the use of machine learning techniques as innovative tools for economic research.

References

- Armah, N. A. et al. (2013) "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, 2013(Summer), 32–39.
- Athey, S., and G. W. Imbens (2019), "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 11(1), 685–725.
- Ba, J. L., J. R. Kiros, and G. E. Hinton (2016), "Layer Normalization," *arXiv preprint arXiv:1607.06450*.
- Babii, A., E. Ghysels, and J. Striaukas (2021), "Machine Learning Time Series Regressions with an Application to Nowcasting," *Journal of Business & Economic Statistics*, pages 1–23.
- Baker, S. R., N. Bloom, and S. J. Davis (2016), "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131(4), 1593–1636.
- Barsky, R. B., and E. R. Sims (2012), "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence," *American Economic Review*, 102(4), 1343–1377.
- Bennani, H., and M. Neuenkirch (2017), "The (home) Bias of European Central Bankers: New Evidence Based on Speeches," *Applied Economics*, 49(11), 1114–1131.
- Bybee, L., B. T. Kelly, A. Manela, and D. Xiu (2021), "Business News and Business Cycles," Technical Report, National Bureau of Economic Research.
- Chollet, F. et al. (2015), Keras. <https://keras.io>.
- Gentzkow, M., B. Kelly, and M. Taddy (2019), "Text as Data," *Journal of Economic Literature*, 57(3), 535–574.
- Heaton, J. (2008), *Introduction to Neural Networks with Java*, Heaton Research, Inc.
- Huang, C., S. Simpson, D. Ulybina, and A. Roitman (2019), *News-based Sentiment Indicators*, International Monetary Fund.
- Jeon, J.-J., S. An, M. Lee, and H. Hwang (2020), "경제용어 감성사전 구축방안연구 [Study on the Construction Methods for the Sentiment Vocabulary of Economic Terminology]," *BOK 지역경제리뷰[BOK Regional Economics Review]*.
- Kim, H., J. Im, H. Lee, and S. Lee (2019), "온라인 뉴스 기사를 활용한 경제심리보조지수 개발 [Development of Economic Sentiment Supplementary Index using Online News Articles]," *BOK 국민계정리뷰 [BOK National Accounts Review]*, 2019(2).
- Kim, H.-j., S.-r. Jeo, and D. Kim (2021), "경제 텍스트 데이터를 활용한 키워드 분석방안 연구 [Study on Keyword Analysis Methods using Economic Text Data]," *BOK 국민계정리뷰 [BOK National Accounts Review]*, 2021(1).
- Larsen, V. H., L. A. Thorsrud, and J. Zhulanova (2021), "News-driven Inflation Expectations and Information Rigidities," *Journal of Monetary Economics*, 117, 507–520.
- LeCun, Y., Y. Bengio, and G. Hinton (2015), "Deep Learning," *Nature*, 521(7553), 436–444.
- Lee, Y. J., S. Kim, and K. Y. Park (2019a), "Deciphering Monetary Policy Board Minutes with Text Mining: The Case of South Korea," *The Korean Economic Review*, 35(2), 471–511.
- _____ (2019b), "Measuring Monetary Policy Surprises using

- Text Mining: The Case of Korea,” *Bank of Korea WP*, 11.
- Loughran, T., and B. McDonald (2011), “When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-ks,” *The Journal of Finance*, 66(1), 35–65.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013), “Efficient Estimation of Word Representations in Vector Space,” *arXiv preprint arXiv, 1301.3781*.
- Moon, H. (2019), “빅데이터의 경제통계 활용 현황 및 시사점 [Current Status and Implication of the Use of Big Data in Economic Statistics],” *한국경제포럼[Korean Economic Forum]*, 11(4), 89–105.
- Nguyen, K., and G. La Cava et al. (2020), “Start Spreading the News: News Sentiment and Economic Activity in Australia,” *Sydney: Reserve Bank of Australia*, 33.
- Ryu, W. H. (2014), Open Korean Text (okt). <https://github.com/open-korean-text/open-korean-text>.
- Seki, K., Ikuta, Y., and Y. Matsubayashi (2022), “News-based Business Sentiment and its Properties as an Economic Index,” *Information Processing & Management*, 59(2), 102795.
- Seo, B., Y. Lee, and H. Cho (2022), “기계학습을 이용한 뉴스심리지수(nsi)의 작성과 활용 [Compilation and Utilization of News Sentiment Index (nsi) using Machine Learning],” *BOK 국민계정리뷰 [BOK National Accounts Review]*, 2022(1).
- Shapiro, A. H., M. Sudhof, and D. J. Wilson (2020), “Measuring News Sentiment,” *Journal of Econometrics*.
- Thorsrud, L. A. (2016), “Nowcasting using News Topics. Big Data Versus Big Bank,” *Working Papers, Centre for Applied Macro and Petroleum Economics (CAMP), BI Norwegian Business School*, 2016(6).
- _____ (2020), “Words are the New Numbers: A Newsy Coincident Index of the Business Cycle,” *Journal of Business & Economic Statistics*, 38(2), 393–409.
- van Aarle, B., and M. Kappler (2012), “Economic Sentiment Shocks and Fluctuations in Economic Activity in the Euro Area and the usa,” *Intereconomics*, 47(1), 44–51.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017), “Attention is All You Need,” *Advances in Neural Information Processing Systems*, 30.
- Webster, J. J., and C. Kit (1992), “Tokenization as the Initial Phase in nlp,” In *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*.
- Won, J.-H., W. Son, and H. Moon (2017), “텍스트 마이닝 기법을 이용한 경제심리 관련 문서 분류 [Economic-sentiment-related Article Classification using Text Mining Techniques],” *BOK 국민계정리뷰 [BOK National Accounts Review]*, 2017(4).

트랜스포머 모델을 이용한 한국의 뉴스 심리 측정*

서 범 석** · 이 영 현*** · 조 형 배****

초 록 본 연구는 인터넷 상에서 수집한 뉴스 기사를 바탕으로 한국 경제에 대한 경제 심리를 측정하는 뉴스심리지수(news sentiment index, NSI)의 작성 방법을 제시하고 검증하였다. 이를 위해 최신 자연어처리 기법을 이용하여 NSI를 작성하고 키워드분석을 통해 지수의 변동 요인을 제시하였다. 구체적으로 본 연구는 트랜스포머 인공지능경망을 기반으로 대규모 뉴스 표본을 처리할 수 있는 감성분석기를 구축한 뒤 이를 이용하여 NSI를 계산하였다. NSI는 서베이 기반 공식 통계지표인 월별 심리지수 보다 신속하게 수시로 작성하는 것이 가능하다. 또한 NSI는 키워드 분석과 부문별 지수 작성을 통해 경제 심리의 변화 요인을 분석하는 것이 가능하다. 본 연구에서 검증한 바에 따르면 NSI는 경기선행지수로서 유용하며 경제 심리의 변곡점을 포착하는 데 이점이 있는 것으로 나타난다.

핵심 주제어: 뉴스 텍스트 데이터, 경제분석을 위한 자연어처리, 심리충격

경제학문헌목록 주제분류: C45, C82, E32

투고 일자: 2022. 7. 5. 심사 및 수정 일자: 2023. 4. 28. 게재 확정 일자: 2023. 8. 31.

* 본 연구과정에서 많은 지원을 아끼지 않은 한국은행 문혜정 팀장님과 논문에 유익한 제안을 해주신 익명의 두 심사자에게 감사드립니다. 본 연구는 동일한 저자의 위킹페이퍼(Seo et al., 2022)를 교정한 연구임을 밝힙니다. 또한 본 논문의 결과는 한국은행의 공식견해가 아닌 집필자 개인의 견해이며, 남아 있는 모든 오류는 저자들의 책임임을 밝힙니다.

** 제1저자 겸 교신저자, 한국은행 경제모형실 과장, 성균관대 쿼트응용경제학과 겸임교수, 고려대 통계학과 겸임교수, e-mail: bsseo@bok.or.kr

*** 공동저자, 한국은행 경제통계국 과장, e-mail: yhlee@bok.or.kr

**** 공동저자, 한국은행 경제모형실 조사역, e-mail: hyung-bae.cho@bok.or.kr