

News sentiment and bond risk premia

Chang-Mo Kang , Donghyun Kim & Hoyoung Park

To cite this article: Chang-Mo Kang , Donghyun Kim & Hoyoung Park (2026) News sentiment and bond risk premia, Cogent Economics & Finance, 14:1, 2602340, DOI: [10.1080/23322039.2025.2602340](https://doi.org/10.1080/23322039.2025.2602340)

To link to this article: <https://doi.org/10.1080/23322039.2025.2602340>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 21 Dec 2025.



[Submit your article to this journal](#)



Article views: 107



[View related articles](#)



[View Crossmark data](#)

News sentiment and bond risk premia

Chang-Mo Kang^a, Donghyun Kim^b and Hoyoung Park^a

^aHanyang University, Seoul, South Korea; ^bChung-Ang University, Seoul, South Korea

ABSTRACT

This study examines whether and how news sentiment regarding interest rate changes predicts bond risk premia in Korea, a leading emerging market. Using machine learning techniques, we construct the Bond News Sentiment Index (BNSI) from news articles on interest rates. Our analysis reveals that the BNSI has a hump-shaped relationship with bond risk premia in the following month. This result suggests that a stronger consensus of news articles on interest rate changes signals decreases in uncertainty, leading to lower bond risk premia. We propose a BNSI-based forecasting factor, which exhibits predictive power beyond existing factors both in- and out-of-sample.

IMPACT STATEMENT

This paper shows that news sentiment related to interest rate changes plays an important role in predicting bond risk premia. By applying machine learning methods to news text, the study introduces a novel sentiment index that captures information beyond conventional risk factors. The findings provide meaningful implications for understanding bond market behavior and the role of public information in shaping investor expectations.

ARTICLE HISTORY

Received 1 September 2025
Revised 27 October 2025
Accepted 7 December 2025

KEYWORDS

News sentiment; bond risk premia; interest rates; factor analysis; machine learning; word-to-vector

JEL CLASSIFICATION CODES

G12; G14; G15

SUBJECTS

Quantitative Finance;
Economics; Statistics for
Business, Finance &
Economics

1. Introduction

Textual analysis has become an increasingly valuable tool in finance, with news articles serving as key sources of investor sentiment and information. Tetlock (2007) first demonstrated that the tone of the Wall Street Journal's daily columns predicts short-term stock returns, suggesting that sentiment influences markets beyond fundamental information. Subsequent studies extend news sentiment analysis to various financial news contexts, finding return predictability at both market and stock levels (Dougal et al., 2012; García, 2013; Manela & Moreira, 2017; Tetlock, 2007). Heston and Sinha (2017) further show that textual analysis captures both investor sentiment and fundamental information.

This study examines whether a Bond News Sentiment Index (BNSI), constructed using machine learning-based textual analysis, predicts bond risk premia—excess returns on long-term sovereign bonds—in emerging markets. Unlike a monotonic relationship, we hypothesize a hump-shaped (inverted U-shaped) relationship between BNSI and bond risk premia. When sentiment is moderate, uncertainty about interest rate changes is high, leading investors to demand greater risk compensation (Baker et al., 2016; Pástor & Veronesi, 2012). In contrast, when sentiment is extreme, either strongly hawkish or dovish, uncertainty declines, lowering risk premia.

Liquidity could further amplify this effect. At moderate BNSI levels, uncertainty discourages trading, reducing bond market liquidity and raising risk premia (Amihud & Mendelson, 1991; Pástor & Stambaugh, 2003). When sentiment reaches extreme values, investors adjust portfolios with greater confidence, improving liquidity and lowering risk premia. Together, these mechanisms imply the non-monotonic relationship between news sentiment and future bond risk premia.

CONTACT Hoyoung Park  stox74@hanyang.ac.kr  Hanyang University, Seoul, South Korea

 Supplemental data for this article is available online at <https://doi.org/10.1080/23322039.2025.2602340>

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

We test this hypothesis using news and sovereign bond market data from Korea, a leading emerging country. The investment demand for Korean Treasury Bonds (KTB) has been increasing in global financial markets. The scale of foreign investments more than doubled from \$87 billion in January 2016 to \$186 billion in 2021.¹ As of August 2020, Korea ranked 10th in the size of supranational, sovereign, and agency bond markets.² Thus, our study is not specific to small and isolated markets, where the participation of sophisticated investors is largely limited.

We first construct a new dictionary to identify textual tones regarding interest rate changes. We do not use an existing lexicon that classifies negative and positive words because the relationship between a news article's tone and its sentiment towards interest rates relies on the economic variables covered by the article.³ For instance, a positive tone about retail sales may imply positive sentiment about an increase in interest rates, whereas a positive tone about monetary supply can be interpreted in the opposite manner. Furthermore, news sentiments can be retrieved more precisely if multiple economic variables are considered jointly.

Specifically, we collect 103,472 sentences from bond analyst reports and interest rate-related online news articles and convert each sentence into a 500-dimensional vector using the word-to-vector (Word2Vec) method. For each sentence, we define a positive tone indicator that assigns one if the sentence predicts an increase in interest rate, and zero otherwise. We then estimate a logistic model using the positive-tone indicator as the dependent variable and the word vector as the independent variable. Coefficient estimates can be regarded as a lexicon because we can measure the tone of a new sentence by converting it into a word vector and calculating the predicted probability of the word vector based on the coefficient estimates.

Next, we measure news sentiments regarding interest rate increases. We collect 13,255 news articles (191,290 sentences) containing 'interest rates' in their titles from the archives of three major Korean economic newspapers between January 2005 and November 2021. These articles do not overlap with the online articles used in the previous learning procedure. We then convert each sentence from the collected articles into a word vector and calculate the predicted probability of an interest rate increase using the coefficient estimates from the logistic model described above. Finally, we construct the monthly bond news sentiment index (BNSI) by computing the monthly averages of predicted probabilities and standardizing them using the sample mean and standard deviation. In line with economic theories, the BNSI is positively associated with short-term rate variables and indicators of economic expansion.

Using this index, we conduct predictive regressions for bond risk premia, computed as the excess returns of long-term KTB compared to one-year KTB over the 12 months. Following Laborda and Olmo (2014), we consider the level, square, and year-on-year change of the BNSI as predictors. Our results confirm the hump-shaped relationship. The squared term of BNSI is the only statistically significant predictor. A one-standard-deviation increase in squared BNSI predicts declines in excess returns two-, three-, four-, and five-year KTBs by 0.12, 0.28, 0.40, and 0.56%, respectively.

A potential concern is that the return predictability of BNSI may overlap with broader monetary policy uncertainty measures (Ioannidis & Ka, 2021). However, our hypothesis suggests that BNSI captures a distinct channel through which uncertainty and liquidity dynamics influence bond risk premia. While monetary policy is a key driver of interest rate uncertainty, BNSI reflects a broader dimension that includes macroeconomic and market-based factors. To ensure that the BNSI provides unique predictive power, we control for both monetary policy and broader uncertainty measures. Specifically, we include the Korean and U.S. Monetary Policy Uncertainty Index (MPUI) proposed by Baker et al. (2016), the changes in the VIX (ΔVIX) to capture global risk aversion Bloom (2009), the U.S. high-yield option-adjusted spread (ΔHY_OAS) to reflect credit-market conditions Gertler and Lown (1999), the term spread between long- and short-maturity Treasury yields (ΔTS) to capture macro-monetary expectations Wright (2006), and the yearly change in the Korean Political Stability Index (ΔPS_t) obtained from the World Bank's Worldwide Governance Indicators. These results indicate that the BNSI captures unique information about market uncertainty and investor sentiment that is not fully reflected in conventional monetary, financial, or political uncertainty measures. Even after accounting for these variables, the squared BNSI and $\Delta TermSpread$ remain a significant predictor, indicating that BNSI captures aspects of market uncertainty and sentiment that are not fully reflected in existing monetary policy uncertainty measures.

We further propose a single return forecasting factor, the News Sentiment (NS) factor, by applying the two-stage estimation method of Cochrane and Piazzesi (2005) to the level, square, and changes in the BNSI. The NS factor significantly predicts bond excess returns over the subsequent 12 months, even after controlling for well-established bond return predictors such as the Cochrane and Piazzesi (2005) forward rate factor (CP factor), the Ludvigson and Ng (2009) macroeconomic predictor (LN factor), and the Laborda and Olmo (2014) investor sentiment factor (BW factor). Out-of-sample tests show that including the NS factor improves forecasting accuracy, reducing mean squared errors relative to models based only on conventional bond pricing factors.

This study contributes to literature on bond risk premia. Several predictors of bond risk premiums have been proposed since the seminal work of Fama and Bliss (1987). Cochrane and Piazzesi (2005) showed that a linear combination of forward rates has significant predictive power over US bond risk premia. Ludvigson and Ng (2009) proposed a bond return predictor that aggregates the information contained in numerous macroeconomic variables, which have been studied previously (Buraschi & Jiltsov, 2005; Cooper & Priestley, 2009; Eriksen, 2017; Ioannidis & Ka, 2021; Wright, 2011). Laborda and Olmo (2014) developed an investor sentiment factor based on the measures proposed by Baker and Wurgler (2007).⁴ We established a new predictor of bond risk premia by applying the news sentiment analysis technique. We provided novel evidence that the news sentiment dispersion captures the uncertainty perceived by bond market investors, which influences the bond risk premia.

This study also contributes to the literature on news sentiment analysis in finance. Existing literature retrieved the return predictors of various asset classes by applying news sentiment analysis techniques (Dougal et al., 2012; García, 2013; Manela & Moreira, 2017; Smales, 2014; Tetlock, 2007). We constructed a valid predictor of bond risk premia by extending textual analysis to sovereign bond market news, which has been underexplored previously. Our work is also distinct from previous studies that examine the relationship between news sentiment and the yield curve (Bauer, 2015; Gotthelf & Uhl, 2019). While these studies focused on yield curve changes over a short horizon, we examine the bond risk premia predictability of news sentiment measures over a long horizon.

The remainder of this paper is organized as follows: Section 2 describes data and empirical methods. Section 3 examines the bond excess return predictability of the BNSI and its relation to monetary policy uncertainty. Section 4 constructs a single forecasting factor based on the BNSI and compares it with the prevailing forecasting factors. Section 5 concludes.

2. Data and methodology

2.1. Data

Two sets of text data are collected from January 2006 to November 2020 to measure bond market news sentiment using machine learning: (i) interest-rate-related news articles published by three major economic newspapers (Maeil Business, Seoul Economic Daily, and Korea Economic Daily) and (ii) bond market analyst reports of a Korean Securities company and interest-rate-related news articles published by two leading online economic news media (Money Today and E-daily). As described in Section 2.2, we measure the sentiments of the former news articles by training machine learning models using the latter text data.

We also use macroeconomic and financial market data to analyze the KTB excess returns. The Korea Fund Investment Association provides the monthly prices of one, two, three, four, and five-year KTBs reported by three major bond rating agencies in Korea (NICE Investors Service, Korea Investors Service, and Korea Asset Pricing). We use the average price of each tenor. Macroeconomic variables are collected from the Bank of Korea and National Statistical Office databases. Among the 132 macroeconomic variables considered in Ludvigson and Ng (2009) US treasury bond market analysis, we select 64 that are available for the entire sample period. For the selected variables, we address non-stationarity by employing the differencing method. The selected macroeconomic variables and applied differencing methods are listed in the Supplementary Appendix (Table A1).

Finally, we construct the Korean investor sentiment index proposed by Kim and Byun (2010), modifying Baker and Wurgler (2007) methodology to make it suitable for Korean capital market data. This index

Table 1. Examples of positive sentiment sentences.

Positive sentiment sentences
The Bank of Korea will raise rates much more aggressively than they have claimed.
Why U.S. government is struggling to address inflation.
Decline in stock markets accelerates on interest rate fears.
Bonds are terrible deals right now.
Interest rates expected to rise this week.
The Bank of Korea is likely to maintain its base rate hike stance until the end of this year, because of the persisting high growth and inflation.
The sharp increase in interest rates has a negative impact on the real estate market.
The Bank of Korea no longer plans to cut interest rates any further.
This table presents examples of training data sentences classified as indicating positive sentiments in bond markets. Positive sentiment implies an increase in bond yields or interest rates. In the original articles, all the sentences were published in Korean.

comprises six sentiment proxies and controls for six business cycle macroeconomic variables. The sentiment proxies include the monthly trading imbalance of individual investors, the monthly stock fund flow ratio, customer deposits for stock investment, the turnover ratio, relative equity issuance, and customer expectation index for the business cycle.⁵ The business cycle macroeconomic variables include the growth of industrial production, durable sales, semi-durable sales index, non-durable sales index, service production index, and coincident composite index. We collect stock and fund market data from the Korea Exchange and Korea Financial Investment Association, respectively. Kim and Byun (2010) provide a detailed description of the variables and index construction methods.

2.2. Bond news sentiment index

We measure the sentiments of news articles by applying Word2Vec, developed by Google in 2013. Word2Vec transforms the words in a sentence into N-dimensional vectors.⁶ When words are transformed into vectors, the similarity between words can be measured using the Euclidean distance or cosine similarity in a vector space. Compared to TF-IDF (Term Frequency-Inverse Document Frequency), another widely used natural language analysis method, Word2Vec needs significantly less computations.

The procedure for measuring news sentiment is as follows: First, we decompose bond analysts' reports and interest-rate-related news articles published by the two online economic news media into sentences. After excluding short sentences of 20 characters or fewer, we obtain 103,472 sample sentences. We then identify sentences that indicate positive sentiments about increases in interest rate. Approximately half of the sample sentences are classified as positive sentiment sentences. Table 1 presents examples of positive-sentiment sentences.

Next, we extract nouns, verbs, adjectives, and adverbs from each sentence using KoNLPy and Gensim (Korean analysis libraries in Python), and convert them into 500-dimensional vectors using the Word2Vec method. We then combine these vectors with a dummy variable, the Sentiment Indicator, which assigns one to a positive sentiment sentence and zero to others. By aggregating all word vectors and Sentiment Indicator values of the 103,472 sentences, we construct a matrix to be trained using machine learning techniques.

We determine the most suitable machine learning classification method by evaluating the performance of four widely used classifiers: logit regression, random forest, AdaBoost, and XGB. Additionally, we perform hyperparameter tuning by examining three Word2Vec vector sizes (100, 300, and 500) and context window sizes (1, 3, and 5), respectively. In sum, 36 ($4 \times 3 \times 3$) combinations of classifiers and hyperparameters are considered. As the performance evaluation metric, we select accuracy given the low likelihood of sentiment imbalance in our sample sentences. In the exercises, we divide the 103,472 sentences into the training (80%) and testing (20%) datasets.

Table 2 lists the classification accuracies for different hyperparameters and models. The logit regression and XGB classifiers achieve ~65% accuracy, whereas the random forest and AdaBoost methods exhibit about 60% accuracy. Despite the similar accuracy levels of the logit regression and the XGB classifier in the test data, the XGB classifier displays a larger accuracy gap between the training and test data, suggesting the possibility of overfitting. As a result, we choose the logit regression method with a context window size of 5 and a vector dimension of 500, which demonstrates the highest accuracy of

Table 2. Text data classification accuracy.

Method	Data	(1) 1 window	(2) 3 windows	(3) 5 windows	(4) Average
(A) 100 vectors					
Logistic regression	Train (A)	62.82	63.66	63.94	63.47
	Test (B)	62.07	63.22	63.24	62.84
XGB classifier	A	65.55	66.94	66.74	66.41
	B	62.13	63.84	64.18	63.38
Random forest classifier	A	59.12	60.08	60.22	59.81
	B	57.74	58.55	58.88	58.39
Ada Boost classifier	A	60.02	60.64	61.35	60.67
	B	58.56	59.07	60.09	59.24
(B) 300 vectors					
Logistic regression	A	64.95	65.58	65.65	65.39
	B	63.97	65.02	64.92	64.64
XGB classifier	A	66.74	68.49	67.67	67.63
	B	63.23	64.79	64.71	64.24
Random forest classifier	A	59.38	60.48	60.18	60.01
	B	58.29	58.88	58.41	58.53
Ada Boost classifier	A	60.73	61.86	61.62	61.40
	B	59.85	60.24	60.08	60.06
(C) 500 vectors					
Logistic regression	A	65.04	65.21	65.51	65.25
	B	64.25	64.82	65.17	64.75
XGB classifier	A	68.10	68.60	68.24	68.31
	B	63.54	65.06	64.89	64.50
Random forest classifier	A	60.02	59.71	60.34	60.02
	B	58.37	58.20	58.68	58.42
Ada Boost classifier	A	61.08	61.83	62.04	61.65
	B	59.70	60.45	60.80	60.32

This table shows the classification accuracies of the four machine learning methods: logistic regression, XGB classifier, random forest classifier, and AdaBoost classifier. Accuracy denotes the accuracy of the model. It is measured using the total number of predictions in the train or test data set as the denominator and the number of accurately predicting the meaning of interest rate cuts or interest rate increases as the numerator. We divide 103,472 sentences into train and test data with a ratio of 4:1. (A–C) Show the cases in which the word-vector dimensions were set to 100, 300, and 500, respectively. In each panel, columns (1)–(3) present the accuracies of each method and data when setting the context window size to 1, 3, and 5, respectively, whereas column (4) presents the average of the previous three columns.

65.17%. This performance is comparable to that of Mikolov et al. (2013) for word-similarity classification scores.

Notably, Rushin et al. (2017) compare the performance of logit regression, gradient boost, early XGB, and deep learning for detecting credit card fraud and find that logit regression has a slightly lower performance than the other two classification models. However, in our study, logit regression demonstrates a comparable performance to the XGB classifier on the test data, prompting us to select it as a more appropriate classification model because of its lower risk of overfitting.

Using the word matrix trained by the logit regression method, we measure the sentiments of 13,255 interest rate-related news articles published by the three major economic news media between January 2006 and November 2021. Specifically, we first convert each sentence in the articles into a 500-dimensional vector, as with the training data, and then measure the sentiment score. We then compute the average sentiment scores for all the sentences in each month's published articles. Finally, we standardize the monthly average scores as in Equation (1) and referred to it as the Bond News Sentiment Index (BNSI).

$$BNSI = \frac{\text{Monthly Average Sentiment Score (MAS)} - \text{Sample Mean of MAS}}{\text{Sample Standard Deviation of MAS}}. \quad (1)$$

We assess the validity of the BNSI by investigating its relationship with macroeconomic variables. We estimate ordinary least squares models in which the BNSI is regressed on each of the 64 macroeconomic variables.

Table 3 presents the coefficient estimates, *t*-statistics, and coefficients of determination (R^2) of the 15 regressions with the highest R^2 . The BNSI is positively associated with the economic sentiment index, short-term rates (certificate of deposit rates, one-year KTB yield, and repurchase agreement rates), price

Table 3. BNSI and macroeconomic variables.

Macroeconomic variables (regressors)	Coeff	t-Value	R ²
Economy sentiment index	0.47	6.91	0.21
Retail sales total (%)	0.40	5.72	0.16
CD rate (91 days)	0.39	5.49	0.15
Total equipment investment (%)	0.38	5.33	0.14
KTb 1Y	0.36	5.07	0.13
RP rate (%)	0.36	5.05	0.13
Index of producer shipment	0.35	4.87	0.12
House price index	0.31	4.33	0.10
PPI service	0.27	3.74	0.07
PPI industrial goods	0.27	3.63	0.07
M1 monetary supply	−0.26	−3.57	0.07
CPI transportation	0.22	2.92	0.05
KTb 3Y	0.22	2.91	0.05
Index of producer inventory	0.21	2.83	0.04
Construction completed (%)	0.21	2.75	0.04

This table shows the estimation results of 15 simple ordinary least squares (OLS) models with the highest coefficient of determinants (R^2) among those that regress the *BNSI* on each of the 64 macroeconomic variables. The four columns present macroeconomic variables (regressors), coefficient estimates, t -statistics, and R^2 , respectively. All economic indicators were obtained from the Bank of Korea. The sample period spans from January 2006 to November 2020.

level (house price index, producer price index, and consumer price index), and real economic indicators (retail sales, total equipment investment, and index of producer shipment). It is also negatively related to the monetary supply. Overall, a higher BNSI captures the economic and market conditions in which bond yields rise or, equivalently, bond prices decline.

To enhance interpretability, we further conduct a qualitative analysis to open the ‘black box’ of the Word2Vec-to-Logit classifier. For each word in the trained vocabulary, we compute the dot product between its embedding vector and the estimated Logit coefficient vector, which measures the directional contribution of the word to the probability of a hawkish (rate-hike) classification. We then rank all tokens by this score and identify those that most strongly influence the model’s predictions. The analysis shows that words such as rate increase, tightening, Federal Reserve, consumer, and dollar contribute positively to the probability of a rate hike, whereas rate cut, stimulus, financial crisis, support, and loan contribute negatively. These results are consistent with monetary-policy narratives and provide an intuitive mapping between the model’s parameters and economically meaningful vocabulary, thereby improving the transparency of our text-based sentiment measure.

2.3. Bond excess return prediction model and forecasting factors

We compute excess bond returns following Cochrane and Piazzesi (2005). Let $p_t^{(n)}$ be the log price of n -year (i.e. maturing in n years) discount bond for month, t . We use parentheses to distinguish maturity from exponentiation in superscripts. The log yield is defined as follows:

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)}. \quad (2)$$

The one-year log-holding period return of a n -year bond in month t is computed as follows:

$$r_t^{(n)} = p_{t+12}^{(n-1)} - p_t^{(n)}. \quad (3)$$

We compute the log excess return (or risk premium) of a n -year bond in month t as

$$rx_t^{(n)} \equiv r_t^{(n)} - y_t^{(1)} = p_{t+12}^{(n-1)} - p_t^{(n)} - y_t^{(1)}. \quad (4)$$

Next, we examine the BNSI’s bond excess return predictability by estimating Laborda and Olmo (2014) predictive regression model. For a n -year bond, the regression model is as follows:

$$rx_{t+1}^{(n)} = \gamma_0 + \delta_1 BNSI_t + \delta_2 \Delta BNSI_t + \delta_3 BNSI_t^2 + u_{t+1}, \quad (5)$$

where $BNSI_t$ is the BNSI level in month t , $\Delta BNSI_t = BNSI_t - BNSI_{t-12}$ is the year-on-year change in BNSI, $BNSI_t^2$ is the square of $BNSI_t$, and u_{t+1} is the residual term.

In this prediction model, we hypothesize that $\delta_3 < 0$, indicating a hump-shaped relationship between BNSI and future bond risk premia. When sentiment is moderate (i.e. the BNSI is close to zero), uncertainty about interest rate changes is high, leading investors to demand greater risk compensation (Baker et al., 2016; Pástor & Veronesi, 2012). In contrast, when sentiment is extreme (i.e. the BNSI is highly positive or negative), uncertainty declines, reducing risk premia.⁷

We further propose a bond return predicting factor for news sentiment (NS factor) following the method of Laborda and Olmo (2014) and test its predictive power relative to the prevailing forecasting factors. Using one to five years of KTB bond prices, the NS factor is defined as the predicted value of the following regression:

$$\overline{rx}_{t+1} = \gamma_0 + \gamma_1 BNSI_t + \gamma_2 \Delta BNSI_t + \gamma_3 BNSI_t^2 + \zeta_{t+1}, \quad (6)$$

where $\overline{rx}_{t+1} = \frac{1}{4} \sum_{n=2}^5 rx_{t+1}^{(n)}$ is the average bond excess return of different maturities in month $t+1$, $BNSI_t$ is the level of BNSI in month t , $\Delta BNSI_t = BNSI_t - BNSI_{t-12}$ is the year-on-year change in BNSI, $BNSI_t^2$ is the square value of $BNSI_t$, and ζ_{t+1} is the residual term. Then, for a n -year bond, the NS factor's excess return predictability is estimated by running the following regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 NS_t + \Lambda X_t + \varepsilon_{t+1} \quad (7)$$

where NS_t is the NS factor, X_t is a vector of control variables, and $\varepsilon_{t+1}^{(n)}$ is the residual term.

As controls, we included three bond risk premia forecasting factors proposed in previous studies. First, we controlled for the forward rate factor (CP factor) developed by Cochrane and Piazzesi (2005). We constructed the CP factor using the KTB data. Under no-arbitrage conditions, the log-forward rate between n and $n+1$ year is defined as follows:

$$f_t^{n \rightarrow n+1} = p_t^{(n)} - p_t^{(n+1)}. \quad (8)$$

Using KTB bond prices for one to five years, the CP factor was estimated as the predicted value of the following regression:

$$\overline{rx}_{t+1} = \lambda_0 + \lambda_1 y_t^1 + \lambda_2 f_t^{1 \rightarrow 2} + \lambda_3 f_t^{2 \rightarrow 3} + \lambda_4 f_t^{3 \rightarrow 4} + \lambda_5 f_t^{4 \rightarrow 5} + \mu_{t+1}, \quad (9)$$

where μ_{t+1} is a residual term.

Second, we controlled for the macroeconomic variable factor (LN factor) proposed by Ludvigson and Ng (2009). To estimate this factor, we began by running a principal component analysis of 64 Korean macroeconomic variables.⁸ The number of $PCA(i)_t$ factors is determined using information criteria proposed by Bai and Ng (2002). The criteria show that the factor structure is well explained by the six common factors. Using these six principal components, which explained 42% of the total variance, we construct a single forecasting factor by running the following regression:

$$\overline{rx}_{t+1} = \kappa_0 + \sum_{i=1}^6 \kappa_i PCA(i)_t + \nu_{t+1}, \quad (10)$$

where $PCA(i)_t$ is the i -th principal component and ν_{t+1} is a residual term.

Finally, we added the investor sentiment factor (BW factor) proposed by Laborda and Olmo (2014), who use the US investor sentiment index developed by Baker and Wurgler (2007). Instead, we used the Korean investor sentiment index constructed by Kim and Byun (2010), modifying Baker and Wurgler (2007) original method to make it more suitable for Korean economic and market data. The BW factor was estimated as the predicted value of the following regression:

$$\overline{rx}_{t+1} = \theta_0 + \theta_1 S_t + \theta_2 \Delta S_t + \theta_3 S_t^2 + \tau_{t+1}, \quad (11)$$

where S_t is the investor sentiment index in month t , ΔS_t is the year-on-year change in investor sentiment, S_t^2 is the squared value of the index, and τ_{t+1} is the residual term. The estimation results for each factor are provided in the [Supplementary Appendix](#). In all estimations, we adjusted the standard error using the Newey-West method, where the lag is set to 18 months, following the method of Cochrane and Piazzesi (2005).

Figure 1 shows the long-term trends of the BNSI and the CP factor, which is known to be the most significant forecasting factor. The CP factor and news sentiment diverged during the financial crisis

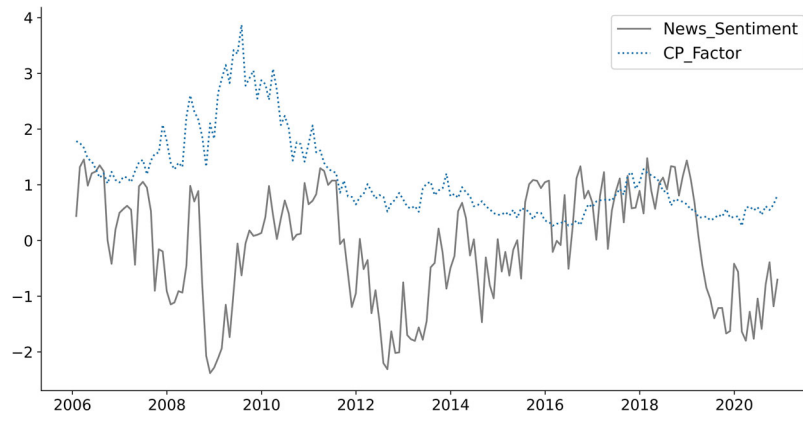


Figure 1. BNSI and CP factor.

This figure shows the time series of the BNSI (solid line) and CP factor (dotted line) during the sample period from January 2006 to November 2020. The construction methods for these two variables are described in [Section 2](#).

Table 4. Summary statistics and correlations.

	$rx_{t+1}^{(2)}$	$rx_{t+1}^{(3)}$	$rx_{t+1}^{(4)}$	$rx_{t+1}^{(5)}$	CP_t	LN_t	BW_t	$BNSI_t$
(A) Summary statistics								
Mean	0.59	0.83	1.59	1.51	1.13	1.13	1.13	0.00
Std. Dev.	0.77	1.35	2.04	2.53	0.76	0.51	0.20	1.00
Median	0.49	0.89	1.84	2.05	0.86	1.12	1.16	0.01
(B) Correlation table								
$rx_{t+1}^{(2)}$	1.00	0.94	0.89	0.77	0.51	0.34	0.11	-0.14
$rx_{t+1}^{(3)}$		1.00	0.97	0.92	0.44	0.31	0.15	-0.07
$rx_{t+1}^{(4)}$			1.00	0.96	0.52	0.31	0.11	-0.01
$rx_{t+1}^{(5)}$				1.00	0.39	0.28	0.11	0.09
CP_t					1.00	0.15	-0.29	0.05
LN_t						1.00	0.09	0.29
BW_t							1.00	-0.18
$BNSI_t$								1.00

This table shows the summary statistics (A) and correlations (B) of the KTB excess returns, CP, LN, BW factors, and BNSI. (A) Presents the means, standard deviations, and medians of the variables. The 12-month excess log return from the n -year KTB, denoted by $rx_{t+1}^{(n)}$, is one-month leading relative to the other variables. The subscripts represent the lead-lag relationship. CP_t is Cochrane and Piazzesi (2005) forward rate factor. LN_t is Ludvigson and Ng (2009) macroeconomic factor. BW_t represents Laborda and Olmo (2014) investor sentiment factors. $BNSI_t$ is a standardized news sentiment index for the KTB markets. The variable construction methods are described in [Section 2](#). The sample period spans from January 2006 to November 2020.

period between 2007 and 2009 and then synchronized again after the recovery. Furthermore, as the low interest rate persisted after the mid-2010s, the volatility of the CP factor decreased rapidly. Overall, the figure suggests that BNSI can contain information orthogonal to the CP factor.

[Table 4](#) displays means, standard deviations, and medians of the bond excess returns, BNSI, NS factor, and other forecasting factors (Panel A) and their correlations (Panel B). Panel A shows that the excess returns are on average higher for longer tenor bonds. The mean excess returns are highest for four-year KTBs (1.59%) and lowest for two-year KTBs (0.59%). The longer term KTBs also exhibit higher excess return volatilities. The standard deviation of excess returns is highest for five-year KTBs. Regarding the forecasting factors, Panel A shows that CP, LN, and BW factors exhibit similar means (1.13%) but a large divergence in standard deviation (0.76, 0.51, and 0.2%, respectively).

Panel B shows that the BNSI is negatively correlated with the subsequent 12-month excess returns of two, three, and four-year KTBs but positively associated with those of five-year KTBs. Furthermore, the correlation coefficient increases monotonically with the tenor. This observation suggests the possibility of a non-monotonic relationship between the BNSI and future bond risk premia. Regarding the forecasting factors, the BNSI is positively correlated with the CP (0.05) and LN (0.29) factors but negatively associated with the BW factor (-0.18). The negative relationship with the BW factor suggests that the return predictability of the BNSI is orthogonal to investor sentiment.

3. Bond risk premia predictability of BNSI

3.1. Predictive regressions

Table 5 presents the estimation results of the predictive regressions in Equation (5). Among the explanatory variables, only the square value of the BNSI is negatively and significantly related to the excess return in the subsequent 12 months. The coefficient estimates of the square term are -0.12 , -0.28 , -0.40 , and -0.56 in the regressions of two, three, four, and five-year KTB excess returns, respectively. The estimates are statistically significant at the 1% level for the five-year bond and at the 5% level for other tenors. In contrast, the coefficient estimates of the level and change in BNSI are statistically insignificant. Furthermore, the regression model exhibits stronger return predictability for longer bond tenors. The coefficient of determinants (R^2) increases with tenors, from 0.03 (two-year bonds) to 0.08 (five-year bonds).

Our estimation results confirm a hump-shaped relationship between BNSI and future bond risk premia, given that BNSI is standardized to have a zero mean. This pattern suggests that bond risk premia decline when news sentiment about interest rate changes is strong and unified, as captured by the squared value of BNSI. These findings support the hypothesis that extremely strong news sentiment reflects lower uncertainty about interest rate movements, reducing the risk compensation required by bond market investors.

3.2. BNSI, monetary policy uncertainty, and other uncertainty controls

In this subsection, we examine whether BNSI serves as a novel return predictor relative to Korean and US monetary policy uncertainty measures. There may be a concern that the bond excess return predictability based on news sentiment dispersion could substantially overlap with that of monetary policy uncertainty (Ioannidis & Ka, 2021). To address this concern, we first add the Korean and US Monetary Policy Uncertainty Index (MPUI)—a widely used measure proposed by Baker et al. (2016)—as control variables. We further extend the analysis to incorporate broader sources of global and domestic uncertainty, including the changes in the VIX (ΔVIX) to represent global risk aversion, following Bloom (2009), the U.S. high-yield option-adjusted spread (ΔHY_OAS) to capture credit-market conditions following Gertler and Lown (1999), the term spread between long- and short-maturity Treasury yields (ΔTS) to reflect macro-monetary expectations following Wright (2006), and the yearly change in the Korean Political Stability Index (ΔPS_t) to account for domestic political uncertainty.

Table 6 presents the estimation results. The square value of BNSI exhibits a significant predictive power, even after controlling for global and domestic uncertainty measures. The coefficient estimate is negative and statistically significant at the 1% level for five-year bonds, at the 5% level for three- and four-year bonds, and at the 10% level for two-year bonds. The results confirm the robustness of the non-linear relationship between news sentiment dispersion and future bond risk premia. Contrastingly, the Korean and U.S. MPUI, ΔVIX , and ΔHY_OAS are not significantly associated with future bond excess

Table 5. Bond excess return predictability of the BNSI.

	Dependent var.	β_0	β_1	β_2	β_3	R^2
(1)	$rx_{t+1}^{(2)}$	0.70*** (0.19)	-0.19 (0.13)	0.02 (0.11)	-0.12** (0.06)	0.03
(2)	$rx_{t+1}^{(3)}$	1.10*** (0.35)	-0.35 (0.27)	0.13 (0.20)	-0.28** (0.12)	0.04
(3)	$rx_{t+1}^{(4)}$	2.00*** (0.52)	-0.49 (0.45)	0.30 (0.31)	-0.40** (0.16)	0.05
(4)	$rx_{t+1}^{(5)}$	2.11*** (0.64)	-0.45 (0.57)	0.45 (0.40)	-0.56*** (0.21)	0.08

This table reports estimates from the following predictive regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 BNSI_t + \beta_2 \Delta BNSI_t + \beta_3 BNSI_t^2 + \varepsilon_{t+1}^{(n)}$$

The dependent variable $rx_{t+1}^{(n)}$ is the 12-month excess log return from the n -year KTB ($n = 2, 3, 4, 5$) in month $t + 1$. The model includes three independent variables: $BNSI_t$ is standardized news sentiment index on KTB markets in month t ; $\Delta BNSI_t$ is the year-on-year change of $BNSI_t$; $BNSI_t^2$ is the square value of $BNSI_t$. Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. We use (Newey and West, 1987) correction with 18-month lags in the estimation of the standard errors reported in parentheses. The asterisks ***, **, and * indicate that the coefficient estimates are significant at the 1, 5, and 10% levels, respectively. The sample period spans from January 2006 to November 2020.

Table 6. BNSI with global and domestic uncertainty variables.

Dependent var	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	R^2
(1) $rx_{t+1}^{(2)}$	0.44 (0.35)	0.08 (0.15)	-0.08 (0.09)	-0.09* (0.05)	0.00 (0.00)	-0.00 (0.00)	0.09 (0.18)	0.03 (0.03)	0.67*** (0.22)	-1.14* (0.86)	0.32
(2) $rx_{t+1}^{(3)}$	1.06* (0.63)	-0.04 (0.28)	0.02 (0.19)	-0.24** (0.11)	0.00 (0.01)	-0.00 (0.00)	0.08 (0.33)	0.03 (0.05)	0.78** (0.36)	-2.10 (1.55)	0.16
(3) $rx_{t+1}^{(4)}$	1.92** (0.89)	-0.03 (0.46)	0.14 (0.29)	-0.35** (0.15)	0.00 (0.01)	-0.00 (0.01)	0.18 (0.56)	0.03 (0.09)	1.09** (0.47)	-3.37 (2.24)	0.16
(4) $rx_{t+1}^{(5)}$	2.56** (1.08)	-0.12 (0.57)	0.38 (0.40)	-0.53** (0.21)	0.00 (0.01)	0.00 (0.01)	0.07 (0.64)	0.03 (0.08)	0.60 (0.51)	-3.66 (2.81)	0.11

This table reports estimates from the following predictive regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 BNSI_t + \beta_2 \Delta BNSI_t + \beta_3 BNSI_t^2 + \beta_4 EPU_1 + \beta_5 EPU_2 + \beta_6 \Delta VIX_t + \beta_7 \Delta HY_OAS_t + \beta_8 \Delta TS_t + \beta_9 \Delta PS_t + \varepsilon_{t+1}^{(n)}.$$

In (A), the dependent variable $rx_{t+1}^{(n)}$ is the 12-month excess log return from the n -year KTB ($n = 2, 3, 4, 5$) in month $t + 1$. The model includes three independent variables: $BNSI_t$ is standardized news sentiment index on KTB markets in month t ; $\Delta BNSI_t$ is the year-over-year change of $BNSI_t$; $BNSI_t^2$ is the square value of $BNSI_t$. EPU_1 is the Monetary Policy Uncertainty of South Korea. EPU_2 is the Monetary Policy Uncertainty of US. In addition, we include three market control variables: ΔVIX_t denotes monthly changes in the implied volatility index (VIX) capturing global risk aversion; ΔHY_OAS_t denotes monthly changes in the U.S. high-yield option-adjusted spread, reflecting corporate credit risk; and ΔTS_t denotes monthly changes in the term spread between long- and short-maturity U.S. Treasury yields, representing the slope of the yield curve. ΔPS_t represents the yearly change in the Korean Political Stability Index from the World Bank's Worldwide Governance Indicators (WGI). Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. We use the correction of Newey and West (1987) with 18-month lags in the estimation of standard errors, reported in the parentheses. The asterisks ***, **, and * indicate that the coefficient estimates are significant at the 1, 5, and 10% levels, respectively. The sample period spans from January 2006 to November 2020.

returns. Only ΔTS exhibits a significant predictive power, implying that shifts in U.S. macro-monetary expectations incorporates material information on Korean bond risk premia. Overall, these results confirm that, even after accounting for a wide range of political and economic uncertainty variables, the square value of BNSI remains a strong and significant predictor of bond excess returns.

4. News sentiment forecasting factor of bond risk premia

4.1. Bond risk premia predictability of news sentiment factor

Table 7 presents the two-step regression estimates. Panel A reports the results of the first-step regression in Equation (6) in which the NS factor is constructed. In line with the estimates of the predictive regression model reported in Table 5, the square value of the BNSI is negatively and significantly associated with the average excess returns of 2–5 years of bonds in the subsequent 12 months. The NS factor was set to the predicted value of the regression model. Panel B reports the results of the second-step regression in Equation (7), in which the NS factor's bond excess return predictability is estimated. We ran a univariate regression as a benchmark. For all the tenors of bonds other than the two-year one, the NS factor exhibits excess return predictability that is statistically significant at the 5% level. The coefficient estimates of the NS factor increase with bond maturity, suggesting that the factor exhibits stronger return predictability for longer-term bonds.

Next, we tested whether the NS factor provides novel information about future bond risk premia beyond the prevailing predictors, namely, CP, LN, and BW factors. Table 8 presents the estimation results. After controlling for the three factors, we found that the NS factor exhibits significant excess return predictability for all tenors of bonds. The coefficient estimates of the NS factor are statistically significant at the 10% level for two-year bonds and at the 5% level for all other maturities. The economic magnitude of the return predictability is also substantial. A one-standard-deviation increase in the NS factor (i.e. a decrease in the positivity of news tone) predicts 0.09, 0.29, 0.47, and 0.73% increases in the two-, three-, four-, and five-year KTB excess returns in the subsequent 12 months, respectively. Importantly, these magnitudes are economically meaningful when compared with realistic investment costs. According to Fund Doctor,⁹ the annual total expense ratio of Korean bond mutual funds—including both management fees and transaction costs—typically ranges between 0.1 and 0.3% of net asset value. Hence, the predicted excess returns for the 3-, 4-, and 5-year maturities clearly exceed the representative cost range, while the 2-year maturity return (0.09%) is close to the lower bound but remains economically non-trivial. These results imply that the predictability of the NS factor not only passes statistical significance but also delivers economically meaningful performance that can cover realistic

Table 7. Bond excess return predictability of the NS factor.

(A) First-stage regression (NS factor)						
	Dependent var.	γ_0	γ_1	γ_2	γ_3	R^2
(1)	rx_{t+1}	1.48*** (0.41)	−0.37 (0.35)	0.22 (0.25)	−0.34** (0.13)	0.05
(B) Second-stage regressions (bond return predictability)						
	Dependent var.	β_0	β_1			R^2
(1)	$rx_{t+1}^{(2)}$	0.30 (0.27)	0.26 (0.16)			0.02
(2)	$rx_{t+1}^{(3)}$	0.00 (0.44)	0.74** (0.34)			0.05
(3)	$rx_{t+1}^{(4)}$	0.20 (0.71)	1.23** (0.56)			0.06
(4)	$rx_{t+1}^{(5)}$	−0.50 (0.92)	1.77** (0.75)			0.09

This table reports the bond risk premium predictability of the NS factor. (A) Presents the estimates of the following predictive regressions:

$$rx_{t+1} = \gamma_0 + \gamma_1 BNSI_t + \gamma_2 \Delta BNSI_t + \gamma_3 BNSI_t^2 + \varepsilon_{t+1}.$$

The dependent variable rx_{t+1} is the average of the 12-month excess log returns from the two, three, four, and five-year KTBs for month $t + 1$. The model includes three independent variables: $BNSI_t$ is standardized news sentiment index on KTB markets in month t ; $\Delta BNSI_t$ is the year-on-year change of $BNSI_t$; $BNSI_t^2$ is the square value of $BNSI_t$. The NS factor is defined as the predicted value of the above regression. (B) Presents the estimates of the following predictive regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 NS_t + \varepsilon_{t+1}.$$

The dependent variable $rx_{t+1}^{(n)}$ is the 12-month excess log return from the n -year KTB ($n = 2, 3, 4, 5$) in month $t + 1$. The independent variable NS_t is the NS factor in month t . Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. In all estimations, we use Newey and West's (1987) correction with 18-month lags in the estimation of the standard errors reported in parentheses. The asterisks ***, **, and * indicate that the coefficient estimates are significant at the 1, 5, and 10% levels, respectively. The sample period spans from January 2006 to November 2020.

Table 8. Bond excess return predictability of NS factor and other factors.

	Dependent var.	CP	LN	BW	NS	R^2
(1)	$rx_{t+1}^{(2)}$	0.54*** 0.13	0.37** 0.18	0.98*** 0.3	0.20* 0.11	0.38
(2)	$rx_{t+1}^{(3)}$	0.83*** 0.17	0.62*** 0.24	1.89*** 0.55	0.67** 0.26	0.36
(3)	$rx_{t+1}^{(4)}$	1.43*** 0.21	0.89*** 0.24	2.61*** 0.95	1.08** 0.42	0.42
(4)	$rx_{t+1}^{(5)}$	1.32*** 0.27	1.12*** 0.27	2.84* 1.48	1.67** 0.67	0.31

This table reports estimates from the following predictive regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 NS_t + \lambda X_t + \varepsilon_{t+1}.$$

The dependent variable $rx_{t+1}^{(n)}$ is the 12-month excess log return from the n -year KTB ($n = 2, 3, 4, 5$) in month $t + 1$. The independent variable NS_t is the NS factor for month t . The control variable X_t includes three variables: CP_t is Cochrane and Piazzesi (2005) forward rate factor; LN_t is Ludvigson and Ng (2009) macroeconomic factor; BW_t denotes an investor sentiment factor proposed by Laborda and Olmo (2014). The variable construction methods are described in Section 2. Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. In all estimations, we use Newey and West's (1987) correction with 18-month lags in the estimation of the standard errors reported in parentheses. The asterisks ***, **, and * indicate that the coefficient estimates are significant at the 1, 5, and 10% levels, respectively. The sample period spans from January 2006 to November 2020.

transaction costs. We also confirmed that, by adding the NS factor, the R^2 increases by 0.04, 0.05, and 0.07 for three-, four-, and five-year bonds, respectively, which suggests that the NS factor has predictive power beyond the other three factors. Overall, we conclude that the NS factor not only provides information uncaptured by the CP and LN factors but also offers economically significant predictive value that remains relevant for practical investment decisions.

4.2. Bond risk premia predictability during normal period

Our sample includes two exceptional events, the Global Financial Crisis and the COVID-19 pandemic. To check the robustness of the bond excess return predictability during the normal period, we estimate our baseline predictive regressions after excluding the NBER-defined recession periods, December 2007–June 2009 (the Global Financial Crisis), and February 2020–April 2020 (the COVID-19 pandemic).

Table 9 presents the estimates. Overall, the results are largely consistent with the full sample estimates in Table 5. In the subsample analysis, the coefficients on the squared value of BNSI ($BNSI_t^2$) are

Table 9. Bond excess return predictability of the BNSI.

	Dependent var.	β_0	β_1	β_2	β_3	R^2
(1)	$rx_{t+1}^{(2)}$	0.54 *** (0.13)	-0.16 ** (0.12)	0.10 (0.11)	-0.12 ** (0.06)	0.06
(2)	$rx_{t+1}^{(3)}$	0.83 *** (0.28)	-0.32 (0.28)	0.25 (0.23)	-0.21 * (0.11)	0.08
(3)	$rx_{t+1}^{(4)}$	1.71 *** (0.48)	-0.41 (0.46)	0.43 (0.37)	-0.41 ** (0.18)	0.10
(4)	$rx_{t+1}^{(5)}$	1.81 *** (0.63)	-0.44 (0.61)	0.58 (0.47)	-0.50 ** (0.24)	0.11

This table reports estimates from the following predictive regression:

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 BNSI_t + \beta_2 \Delta BNSI_t + \beta_3 BNSI_t^2 + \varepsilon_{t+1}^{(n)}.$$

The dependent variable $rx_{t+1}^{(n)}$ is the 12-month excess log return from the n -year KTB ($n = 2, 3, 4, 5$) in month $t + 1$. The model includes three independent variables: $BNSI_t$ is standardized news sentiment index on KTB markets in month t ; $\Delta BNSI_t$ is the year-on-year change of $BNSI_t$; $BNSI_t^2$ is the square value of $BNSI_t$. Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. We use Newey and West's (1987) correction with 18-month lags in the estimation of the standard errors reported in parentheses. The asterisks ***, **, and * indicate that the coefficient estimates are significant at the 1, 5, and 10% levels, respectively. The sample period spans from January 2006 to November 2020, excluding the global financial crisis period (December 2007–June 2009) and the COVID-19 pandemic period (February 2020–April 2020).

Table 10. Out-of-sample bond excess return predictability of NS factor.

	Dependent var.	MSE restricted	MSE unrestricted	p -Value
(1)	$rx_{t+1}^{(2)}$	0.19	0.18	0.21
(2)	$rx_{t+1}^{(3)}$	0.86	0.75	0.00
(3)	$rx_{t+1}^{(4)}$	2.13	1.83	0.00
(4)	$rx_{t+1}^{(5)}$	4.39	3.57	0.00

This table reports the out-of-sample bond risk premium predictability of the NS factor. We present the mean squared error (MSE) of the following restricted and unrestricted predictive regression models:

$$(\text{Restricted}) \quad rx_{t+1}^{(n)} = \beta_0 + \Lambda X_t + \varepsilon_{t+1}^{(n)}.$$

$$(\text{Unrestricted}) \quad rx_{t+1}^{(n)} = \beta_0 + \beta_1 NS_t + \Lambda X_t + \varepsilon_{t+1}^{(n)}.$$

In both regression models, the dependent variable $rx_{t+1}^{(n)}$ is the excess log return on the n -year Korean Treasury bonds. The independent variable NS_t is the NS factor for month t . The control variable X_t includes three variables: CP_t is Cochrane and Piazzesi (2005) forward rate factor; LN_t is Ludvigson and Ng (2009) macroeconomic factor; BW_t denotes an investor sentiment factor proposed by Laborda and Olmo (2014). Rows (1)–(4) report the estimates from the regressions of the two, three, four, and five-year KTB excess log returns. The last column presents ' p -value' reports the result of Diebold and Mariano test (2002) for a null hypothesis that the restricted and unrestricted model produce the same MSE. The sample period spans from January 2006 to November 2020.

statistically significant at 5% level for all maturities of bond excess returns but the 3-year maturity, which is significant at 10% level. In untabulated tests, we confirm significant predictive power of the NS factor for 3-, 4-, and 5-year bond excess returns during this subsample period.

These tests demonstrate that the predictability of BNSI square values and NS factor are not driven by the extraordinary volatility of the financial crisis or pandemic episodes. Even under normal market conditions, the bond-return predictability associated with bond market news sentiment remains statistically robust and economically meaningful.

4.3. Out-of-sample test

Finally, we assessed the effectiveness of the NS factor with the power to forecast excess bond returns on the KTB. To evaluate the predictive ability of the NS factor, we calculated the mean square errors (MSEs) of the restricted and unrestricted models. The restricted model regresses excess bond returns over the subsequent 12 months on the CP, LN, and BW factors, while the unrestricted model includes the NS factor as an additional regressor. We tested the difference between the MSEs of the two models, following Diebold and Mariano (2002). To estimate the out-of-sample predictive power of the models, we ran rolling regressions for over 60 months. The initial estimation period spans January 2006 to December 2010, and the fitted parameters are used to forecast the 12-month excess bond returns from January 2011 to December 2011. This procedure was repeated until November 2020.

The test results are presented in Table 10. For all the tenors of the bonds, the unrestricted model exhibited lower MSEs than the restricted model. The differences are statistically significant at the 1% level for all tenors of bonds except the two-year one, which is consistent with the in-sample analysis results reported in Table 8. Overall, our findings suggest that the NS factor provides predictive power over bond risk premia beyond those contained in previously documented forecasting factors.

5. Conclusion

In this study, we provide evidence that news sentiment in bond markets, measured by a machine learning technique, has predictive power over bond risk premia in a leading emerging Korean market. The results suggest a hump-shaped relationship between news sentiment and future bond risk premia. That is, strong news sentiment on interest rate changes, either rises or cuts, predicts lower bond risk premia. This supports the uncertainty hypothesis, positing that strong (moderate) news sentiment signals reduced (heightened) uncertainty about interest rate changes, reducing (increasing) risk compensation required by investors. We also show that the return predictability of bond market news sentiment is not subsumed by the prevailing predictors of bond risk premia, both in- and out-of-sample.

This study has novel implications for bond market investors and regulators. First, the return predictability of the BNSI, beyond the prevailing forecasting factors, suggests that news articles contain novel information about risk pricing in sovereign bond markets. Second, the consensus of news articles on bond markets conveys signals not only about market prospects or policy rate changes, but also investors' perceptions of uncertainty. Investors and policymakers should consider this aspect during the decision-making process.

Finally, while our empirical analysis focuses on the Korean market, the methodological framework can be applied globally. The Word2Vec-to-Logit approach provides a flexible and portable way to extract economic sentiment from unstructured news text, which can be applied to other countries with appropriate language-specific training data. Nevertheless, we acknowledge that certain components of the empirical design—such as the construction of the BW factor and the selection of macro variables for the LN factor—are tailored to the institutional features of the Korean market. Extending this framework to other economies would require country-specific news datasets and model calibration, which we view as a promising direction for future research.

Notes

1. Korea Financial Supervisory Service Statistics, <https://fisis.fss.or.kr/fss/fsiview/indexw.html>.
2. <https://www.icmagroup.org/market-practice-and-regulatory-policy/secondary-markets/bond-market-size/>.
3. Depending on the target for news sentiment, previous studies measure news tones by using the existing lexicon (e.g. García, 2013), adopting the news sentiment indices provided by specialized agencies (Heston & Sinha, 2017), or creating a new lexicon like us (e.g. Loughran & McDonald, 2016).
4. Li (2021) extended the analysis to emerging market sovereign bonds.
5. Since 2019, the Bank of Korea replaced the customer expectation index with the economic sentiment index.
6. Leading models of word vectorization include Word2Vec (Mikolov et al., 2013) developed by Google in 2013, Glove (Pennington et al., 2014) proposed by Stanford University, and FastText (Joulin et al., 2016) published by Facebook.
7. Previous studies have proposed economic uncertainty as one of the key economic determinants of risk premia (e.g. Bollerslev et al., 2009; Breach et al., 2020; Brogaard & Detzel, 2015; Grishchenko et al., 2017; Bech & Malkhozov 2016; Wright, 2011).
8. In the [Supplementary Appendix](#), we provide the macroeconomic variables that are highly correlated with each principal component. The key variable is industrial production for the first principal component, yield spread between Baa bond and 10-year KTB for the second principal component, and manufacturing production capacity for the third principal component.
9. Fund Doctor is a leading fund evaluation and data analytics company in Korea, providing comprehensive return and fee data for domestic mutual funds.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

CRedit: **Chang-Mo Kang**: Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing; **Donghyun Kim**: Investigation, Methodology, Validation, Writing – review & editing; **Hoyoung Park**: Data curation, Resources, Software.

Funding

Chang-Mo Kang acknowledges that this work was supported by the research fund of Hanyang University (HY-20230000003542). No other funding was received for this study.

About the authors

Chang-Mo Kang is an Associate Professor of Finance at the School of Business, Hanyang University, Seoul, Republic of Korea. He received his Ph.D. in Finance from the University of Texas at Austin, and his research interests include corporate finance, corporate governance, executive compensation, and market efficiency.

Donghyun Kim is an Associate Professor of Finance at the College of Business and Economics, Chung-Ang University, Seoul, Republic of Korea. He earned his Ph.D. in Finance from the University of Texas at Austin, and his research interests include investments, asset pricing, macro finance, and behavioral finance.

Hoyoung Park is a Ph.D. candidate in Finance at Hanyang University, Seoul, Republic of Korea. His research focuses on asset pricing and behavioral finance.

Data availability statement

The authors agree to share the data and materials supporting the findings of this study upon reasonable request to the corresponding author, except for components restricted by licensing agreements and data-provider terms. Detailed code used for analysis can be provided upon reasonable request.

References

- Amihud, Y., & Mendelson, H. (1991). Liquidity, Maturity, and the Yields on U.S. Treasury Securities. *The Journal of Finance*, 46(4), 1411–1425. <https://doi.org/10.1111/j.1540-6261.1991.tb04623.x>
- Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221. <https://doi.org/10.1111/1468-0262.00273>
- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129–151. <https://doi.org/10.1257/jep.21.2.129>
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/10.1093/qje/qjw024>
- Bauer, M. D. (2015). Nominal interest rates and the news. *Journal of Money, Credit and Banking*, 47(2-3), 295–332. <https://doi.org/10.1111/jmcb.12177>
- Bech, M. L., & Malkhozov, A. (2016, March). *How have central banks implemented negative policy rates?* BIS Quarterly Review, March, 31–44.
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623–685.
- Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492. <https://doi.org/10.1093/rfs/hhp008>
- Breach, T., D'Amico, S., & Orphanides, A. (2020). The term structure and inflation uncertainty. *Journal of Financial Economics*, 138(2), 388–414. <https://doi.org/10.1016/j.jfineco.2020.04.013>
- Brogaard, J., & Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, 61(1), 3–18. <https://doi.org/10.1287/mnsc.2014.2044>
- Buraschi, A., & Jiltsov, A. (2005). Inflation risk premia and the expectations hypothesis. *Journal of Financial Economics*, 75(2), 429–490. <https://doi.org/10.1016/j.jfineco.2004.07.003>
- Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1), 138–160. <https://doi.org/10.1257/0002828053828581>
- Cooper, I., & Priestley, R. (2009). Time-varying risk premia and the output gap. *Review of Financial Studies*, 22(7), 2801–2833. <https://doi.org/10.1093/rfs/hhn087>
- Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. <https://doi.org/10.1198/073500102753410444>
- Dougal, C., Engelberg, J., García, D., & Parsons, C. A. (2012). Journalists and the stock market. *Review of Financial Studies*, 25(3), 639–679. <https://doi.org/10.1093/rfs/hhr133>
- Eriksen, J. N. (2017). Expected business conditions and bond risk premia. *Journal of Financial and Quantitative Analysis*, 52(4), 1667–1703. <https://doi.org/10.1017/S0022109017000369>
- Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4), 680–692.
- García, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3), 1267–1300. <https://doi.org/10.1111/jofi.12027>

- Gertler, M., & Lown, C. S. (1999). The information in the high-yield bond spread for the business cycle: Evidence and some implications. *Oxford Review of Economic Policy*, 15(3), 132–150. <https://doi.org/10.1093/oxrep/15.3.132>
- Gotthelf, N., & Uhl, M. W. (2019). News sentiment: A new yield curve factor. *Journal of Behavioral Finance*, 20(1), 31–41. <https://doi.org/10.1080/15427560.2018.1432620>
- Grishchenko, O. V., Mouabbi, S., & Renne, J.-P. (2017). The joint dynamics of U.S. and euro-area inflation rates: Expectations and time-varying uncertainty. Banque de France Working Paper No. 622. Banque de France, Paris.
- Heston, S. L., & Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3), 67–83. <https://doi.org/10.2469/faj.v73.n3.3>
- Ioannidis, C., & Ka, K. (2021). Economic policy uncertainty and bond risk premia. *Journal of Money, Credit and Banking*, 53(6), 1479–1522. <https://doi.org/10.1111/jmcb.12748>
- Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext.zip: Compressing text classification models. *arXiv Preprint*, arXiv:1612.03651.
- Kim, K., & Byun, J. (2010). Effect of investor sentiment on market response to stock split announcement. *Asia-Pacific Journal of Financial Studies*, 39(6), 687–719. <https://doi.org/10.1111/j.2041-6156.2010.01029.x>
- Laborda, R., & Olmo, J. (2014). Investor sentiment and bond risk premia. *Journal of Financial Markets*, 18, 206–233. <https://doi.org/10.1016/j.finmar.2013.05.008>
- Li, Y. (2021). Investor sentiment and sovereign bonds. *Journal of International Money and Finance*, 115, 102388. <https://doi.org/10.1016/j.jimonfin.2021.102388>
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230. <https://doi.org/10.1111/1475-679X.12123>
- Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22(12), 5027–5067. <https://doi.org/10.1093/rfs/hhp081>
- Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1), 137–162. <https://doi.org/10.1016/j.jfineco.2016.01.032>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv Preprint*, arXiv:1301.3781.
- Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703. <https://doi.org/10.2307/1913610>
- Pástor, L., & Stambaugh, R. (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, 111(3), 642–685. <https://doi.org/10.1086/374184>
- Pástor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, 67(4), 1219–1264. <https://doi.org/10.1111/j.1540-6261.2012.01746.x>
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). <https://doi.org/10.3115/v1/D14-1162>
- Rushin, G., Stancil, C., Sun, M., Adams, S., & Beling, P. (2017). Horse race analysis in credit card fraud—Deep learning, logistic regression, and Gradient Boosted Tree. In *2017 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 117–121). IEEE. <https://doi.org/10.1109/SIEDS.2017.7937700>
- Smales, L. A. (2014). News sentiment and the investor fear gauge. *Finance Research Letters*, 11(2), 122–130. <https://doi.org/10.1016/j.frl.2013.07.003>
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Wright, J. H. (2006). *The yield curve and predicting recessions* (Finance and Economics Discussion Series 2006-07). Board of Governors of the Federal Reserve System.
- Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, 101(4), 1514–1534. <https://doi.org/10.1257/aer.101.4.1514>