News-Implied Sovereign Default Risk*

Chukwuma Dim[†] Kevin Koerner[‡] Marcin Wolski[‡] Sanne Zwart[‡]

This version: March 2022

Abstract

We develop a sovereign default risk index using natural language processing techniques and 10 million news articles covering over 100 countries. The index is a high-frequency measure of countries' default risk, particularly for those lacking market-based measures: it correlates with sovereign CDS spreads, predicts rating downgrades, and reflects default risk information not fully captured by CDS spreads. We assess the influence of sovereign default concerns on equity markets and find that spikes in the index are negatively associated with same-week market returns, which reverses over the next week, indicating that investors might overreact to default concerns. Equity markets' reaction to default concerns is more pronounced and persistent for countries with tight fiscal constraints. The response to global, compared to country-specific, default concerns is much stronger, underlining the relevance of global "push" factors for local asset prices.

Keywords: Sovereign default, Credit risk, Equity returns, Machine learning, Natural language processing, Early warning indicators

JEL: F30, G12, G15

*We are grateful to Cristian Condrea, Jochen Schanz and Grigory Vilkov for helpful comments and suggestions. We also thank participants at the EBRD/EIB Big Data seminar 2021 for their helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the European Investment Bank.

[†]Corresponding author. Chukwuma Dim is affiliated with Frankfurt School of Finance & Management, Germany, c.dim@fs.de.

[‡]European Investment Bank, Luxembourg.

1 Introduction

Global sovereign debt soared to unprecedented levels recently as governments provided fiscal stimulus in response to the Covid-19 pandemic. Government debt reached $83.4 trillion in Q1 2021, compared to $70.1 trillion at end-2019 (IIF, 2021b,a). Due to the economic contraction, global public debt ratios also rose significantly when measured relative to the size of the economy, reaching 98.6% of GDP in 2020 from 83.6% in 2019 (IMF, 2021). The IMF (2021) notes that global government debt has now stabilized at just below 100% of GDP, with substantial variation in fiscal and economic developments across countries. The rapid rise in public debt has prompted warnings about the dangers of excessive debt levels, the drag they will place on economic growth and the burden they represent for future generations (see, Eichengreen et al., 2021; Mitchener and Trebesch, 2021).[1]

The literature on the evolution and impacts of sovereign default risk has traditionally relied on credit default swap (CDS) spreads, i.e., the price of insurance against sovereign default, as the daily proxy for default risk (see Augustin (2014), for a review). Despite some concerns that CDS spreads are partly driven by speculation, they are often preferred to alternative measures of sovereign default risk.[2] However, CDS spreads narrow any analysis to the subset of countries with (liquid) CDS contracts, excluding many emerging markets and virtually all less developed countries that pose even higher default concerns to investors. In this study, we address this shortcoming by developing a novel text-based measure of sovereign default risk—the **N**ews-implied **S**overeign **R**isk **I**ndex (NSRI)—that can be computed for any country by applying Natural Language Processing (NLP) techniques to socioeconomic, political and financial news.

Our use of news text to quantify sovereign risk is predicated on the idea that news media is observed at a relatively high frequency and contains timely information about countries' socioeconomic and political developments that should be predictive of and salient to agents' concerns, including governments' creditworthiness. This view is consistent with Shiller (2017) who ad-

[1]High debt levels can constrain governments' fiscal leeway to respond to future crises, such as financial crises, natural disasters, or wars (e.g., Romer and Romer, 2018; Battaglini and Coate, 2016), and can result in default through explicit debt repudiation or inflation (Yared, 2019).

[2]For example, sovereign bond prices are influenced by differences in currency denominations, tenor, covenants and legal jurisdictions, and they are also often less liquid than CDS contracts. Similarly, sovereign credit ratings are only adjusted at irregular intervals and do not reflect real-time dynamics of sovereign default concerns.

vocates for the use of textual analysis to quantify the dynamics of popular narratives in the news media to better understand economic fluctuations. Moreover, competition among media outlets ensures that news caters to market participants' fundamental risk concerns regarding any country at any point in time. Hence, we expect news text to provide timely and informative signals specifically about the future dynamics of countries' riskiness, just as it reflects concerns about future economic growth and consumption (e.g., Liu and Matthies, 2021).

We exploit the rich information in news text with three main contributions to the literature. First, we construct the NSRI which is computable for *any* country in *real-time* and particularly relevant for developing and emerging markets without liquid market proxies of default risk. Second, we show that the NSRI is strongly associated with sovereign CDS spreads both contemporaneously and in predictive terms, predicts future sovereign rating downgrades, and captures default risk information not fully reflected in CDS spreads. Hence, the NSRI can complement the use of CDS spreads to track default risk for countries for which CDS data is available and can serve as a stand-alone high-frequency proxy of default risk for countries that lack market-based measures of sovereign risk. Third, motivated by the evidence that NSRI is a valuable proxy for sovereign default risk, we apply the index to study how default concerns influence equity markets. We show that equity markets tend to overreact to sovereign default risk increase, but less so for countries facing tight fiscal constraints for which the decline in equity returns is more persistent. The overreaction is more pronounced for global default concerns than country-specific concerns, signaling, in general, the relevance of global (push) factors for local asset prices.

Our analysis begins with a large text corpus of 10 million news articles covering over 100 countries from the PressReader News API.[3] We apply two relatively simple yet powerful Machine Learning (ML) techniques—topic modelling and text similarity—to quantify the degree of sovereign default concerns reflected in news articles. Topic modelling allows us to focus on socioeconomic, political and financial news topics that are more likely to reflect sovereign default concerns. To quantify sovereign default risk using text similarity, we follow a similar approach as Engle et al. (2020) and Hassan et al. (2019) and compile a library of 448 documents focused

[3]Note, the number of countries used for our analysis depends on the country coverage of our news API provider. However, our methodology can be extended without modification to a larger sample of countries if one has access to news coverage for those countries through a different news API.

on sovereign default risk. We then calculate how similar the language used in a news article is to the sovereign default library using the cosine similarity algorithm. Intuitively, an article about a country would have a high score if it uses combinations of "default risk charged words" in similar proportions as the sovereign default library, which will be the case if the news article signals elevated sovereign default concern. Finally, we aggregate the news-level scores to a daily frequency for each country to obtain the NSRI. Intuitively, NSRI will be high during elevated news coverage about a country's default risk, which will typically coincide with heightened investor concern about the country's worsening default risk. This view is consistent with academic studies (Liu and Matthies, 2021; Trussler and Soroka, 2014; Robinson, 2007) suggesting that news often focuses on negative events because they are more attention-grabbing.

Figure 1 shows the evolution of the NSRI for two developed countries (Panels A and B) and two emerging ones (Panels C and D). First, we observe that the NSRI has rich time-series and cross-sectional variation, spiking during the great financial crisis and the subsequent European sovereign debt crisis, as well as during the Covid-19 pandemic when several sovereigns faced heightened default concerns. Second, NSRI reasonably tracks the evolution of sovereign CDS spreads with a correlation coefficient ranging between 0.48 and 0.86. More so, CDS spreads appear to be more persistent than NSRI. This is likely because CDS spreads capture the level of risk, whereas NSRI tends to reflect changes in risk, consistent with the idea that news captures changes in market participants' concerns and expectations.

We conduct a formal analysis of the relationship between CDS spreads and the NSRI using panel regressions and controlling for global and local macro-financial variables, including stock market returns, volatility, and general news sentiment. Overall, we find that the NSRI percentage change is strongly associated with that of CDS spreads. For instance, a standard deviation increase in the NSRI percentage change is associated with a 0.21 – 0.40 percent points (pp) increase in the same-week CDS spread percentage change relative to the previous week. NSRI is also significantly related to CDS spreads at the monthly frequency and predicts an increase in the CDS spread percentage change by 1.02 – 1.67 pp over the next month for a standard deviation increase in the NSRI percentage change.

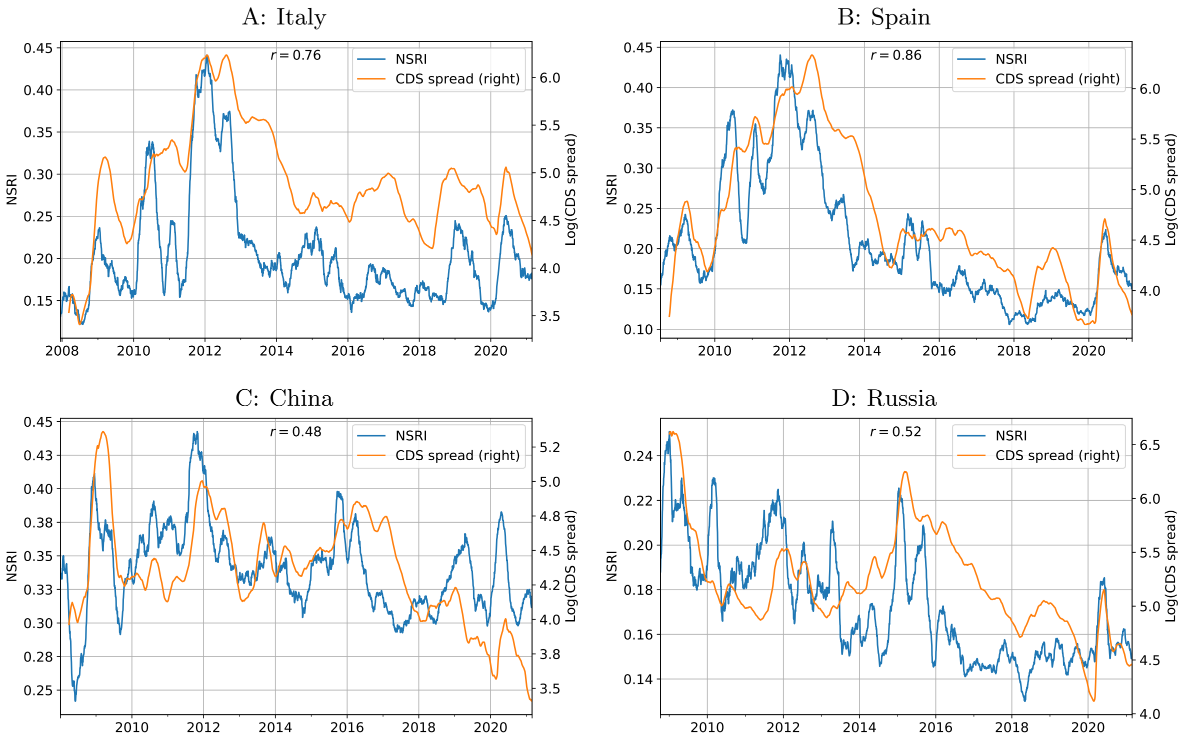

Figure 1: News-implied Sovereign Default Risk and CDS Spreads Dynamics. The figure shows the time series evolution of each country's NSRI (left axis) and the natural log of CDS spreads (right axis). Both variables are smoothed using a three-month moving average. r is the correlation coefficient between both variables over the period January 2007 - February 2021.

The use of CDS spreads to demonstrate NSRI's informativeness about default risk dynamics raises the question of whether the informativeness extends beyond the sample of countries for which CDS data is available. Furthermore, it is helpful to understand whether NSRI provides additional default risk information beyond that already captured by CDS spreads. To address these two questions, we examine whether the NSRI predicts future sovereign rating downgrades using the information on all countries rated by the credit rating agencies Moody's, Standard & Poor's, and Fitch. We find that NSRI has significant explanatory power in predicting sovereign rating downgrades. A standard deviation increase in NSRI over the past 30-day period ending two days before a rating change announcement predicts an increase in the probability of a downgrade by 8.8 pp, which corresponds to a 17% rise in the likelihood of a downgrade relative to the observed fraction of downgrades in the data.

The NSRI's ability to predict future rating downgrades holds even after controlling for selected local and global macro-financial variables, including CDS spreads and news sentiment.

Conversely, conditional on NSRI, CDS spreads do not provide significant incremental information about future rating downgrades. These results indicate that NSRI contains additional sovereign risk information not fully reflected in CDS spreads. A potential reason for this is that news media reflects agents' default concerns more timely before those concerns are fully incorporated in CDS spreads, ultimately leading to downgrades. Overall, NSRI can complement CDS spreads to track default concerns at a high frequency for countries with liquid CDS contracts. For other countries lacking daily market-based measures of sovereign default risk, NSRI can serve as a stand-alone high-frequency proxy for default risk.

Equipped with an informative proxy of sovereign default risk for a broad cross-section of countries, we revisit theoretical predictions on the influence of sovereign default risk in equity markets. Jeanneret (2017) suggests a risk-based channel through which an increase in sovereign default risk negatively impacts equity prices across countries by raising the threat of a future economic slow-down. However, a different channel supported by behavioural explanations (e.g., Dremen, 1983; De Bondt and Thaler, 1985; Gennaioli et al., 2015; Bordalo et al., 2019) is investors' overreaction to news of sovereign default concerns that may not be strongly associated with economic fundamentals. Such overreaction can stem from an aversion to an actual sovereign default's far-reaching adverse financial market impacts. Using panel regressions of weekly stock market returns and weekly-aggregated NSRI, we find that NSRI is associated with stock market returns in ways that support the overreaction hypothesis. Precisely, a standard deviation increase in the NSRI percentage change is contemporaneously associated with a 7.1 bps (3.4% annualized) decline in the equity returns and predicts an increase in equity returns by 6.4 bps (3.1% annualized) over the next week. The reversal of the initial decline in equity returns suggests that equity investors tend to overreact to news of sovereign default concerns.

Next, motivated by the literature on how countries' fiscal capacity can differently impact their ability to withstand adverse economic shocks and avoid default (e.g., Yared, 2019; Romer and Romer, 2018; Battaglini and Coate, 2016), we examine whether the equity market's reaction to sovereign default concerns is heterogeneous across countries. We find that countries with tighter fiscal constraints experience a more pronounced and persistent decline in equity returns following an increase in default concerns. Precisely, for countries facing a high fiscal constraint, a standard deviation increase in the NSRI percentage change is contemporaneously associated with a 10.7 bps (5.5% annualized) decline in weekly equity returns, more than four times the magnitude of the decline experienced by other countries. At the same time, the equity return reversal over the next week for the high fiscal constraint countries is only about 1.9 bps.

Finally, we examine the factor structure in countries' NSRI, following the literature on the common drivers of sovereign default risk reflected in CDS spreads (e.g., Pan and Singleton, 2008; Longstaff et al., 2011). We find that NSRI exhibits a non-negligible factor structure, as the first five principal components explain roughly 51% of the variation in countries' NSRI. Motivated by this evidence, we compare how equity markets respond to global sovereign default concerns versus country-specific default concerns. Surprisingly, we find that global default concerns are more strongly associated with local equity returns in terms of magnitude and statistical significance. A standard deviation increase in the global NSRI percentage change is contemporaneously associated with a decline in weekly equity returns by 21.6 bps, and predicts an increase in returns by 16.7 bps over the next week. The magnitude of the initial decline and subsequent reversal is more than twice that of the country-specific NSRI, indicating that equity investors overreact even more to global sovereign default concerns than country-specific concerns. Hence, the results underline the importance of global (or push) factors for local asset prices, in line with the existing literature (e.g., Colacito et al., 2018; Fraiberger et al., 2021).

Taken together, our paper shows that textual data from news media contains valuable information about countries' default risks that is useful for tracking default concerns in real-time. Our quantification of such default risk concerns from the news media, NSRI, is informative about observed countries' default risk dynamics and equity returns, and hence has useful applications in credit risk early warning systems, credit portfolio management, and for testing theories on the economic costs and financial market impacts of sovereign default risk.

Related Literature: This paper contributes to several strands of research. First, the paper is related to the literature on the measurement and evolution of sovereign default risk. Several papers (e.g., Chernov et al., 2020b; Longstaff et al., 2011; Remolona et al., 2008; Pan and Singleton, 2008) explore the default risk implied by sovereign CDS spreads. Others examine factors that drive sovereign CDS spreads as default risk proxies, including local factors (e.g., Augustin et al., 2021; Chernov et al., 2020a; Hilscher and Nosbusch, 2010) and common global factors (e.g., Ang and Longstaff, 2013; Longstaff et al., 2011; Augustin and Tédongap, 2016). However, one of the major shortcomings of using CDS premium to proxy sovereign default risk is that liquid CDS contracts are not available for most countries. Consequently, tracking default risk concerns for many emerging and developing markets at a daily frequency has proved impossible. Our contribution is to employ big data techniques to construct a novel *real-time* sovereign default risk proxy (NSRI) that depends only on news media data, is not constrained by market liquidity problems, and can be computed for *any* country. We show that, as with CDS spreads, NSRI has a strong factor structure, corroborating the importance of common global drivers of sovereign default risk (e.g., Longstaff et al., 2011).

This paper also relates to the literature on the economic impacts of sovereign credit risk. Several papers note that sovereign default risk has real economic effects (Gennaioli et al., 2014; Adelino and Ferreira, 2016; Bocola, 2016; Almeida et al., 2017; Lee et al., 2016), others focus on stock market implications of sovereign default risk (Andrade and Chhaochharia, 2018; Jeanneret, 2017). By showing that sovereign default risk concerns reflected in the news media affect global equity markets, our paper is more closely related to the latter strand of literature. We present new evidence along three dimensions: (i) Investors tend to overreact to sovereign credit risk news, resulting in stock market effects that typically reverse over the next week. (ii) Equity markets have a heterogeneous response to sovereign risk news, reacting more strongly and persistently to news about fiscally constrained countries, which supports Jeanneret's (2017) risk-based explanation for the effect of sovereign default risk on equity prices. (iii) Common (systemic) global sovereign risk concerns have a more pronounced influence on equity returns than country-specific sovereign risk concerns, highlighting the importance of global factors for local asset prices (e.g., Colacito et al., 2018; Fraiberger et al., 2021).

Finally, our paper is related to, but fundamentally different from, recent papers that measure important economic variables from textual data (e.g., Baker et al., 2016; Engle et al., 2020; Manela and Moreira, 2017; Hassan et al., 2019; Sautner et al., 2020). We differ from these and other related papers in the following ways: First, to our knowledge, we are the first to leverage news media and NLP to construct a high-frequency proxy of sovereign default risk.

Second, the papers that develop country-level variables from text data mainly focus on the US or a few advanced countries using a hand full of news sources. In contrast, our sample covers over 100 countries and uses various news sources, including local and global news publishers. Moreover, we employ a much larger text corpus, over 10 million articles, which enables us to construct NSRI for a broad sample of countries. Overall, our analysis provides strong support for applying text data to understand the dynamics of macro-financial variables.

The paper proceeds as follows. Section 2 describes the data and key variables. Section 3 covers the NLP framework for quantifying sovereign default concerns from news articles. Section 3.3 discusses the NSRI's validation exercises. Section 4 analyzes the NSRI's informativeness about sovereign CDS spreads and downgrades, and how equity markets react to sovereign default concerns. Section 5 discusses robustness tests, and Section 6 concludes the paper.

2 Data and Variables

Our news data comes from the PressReader News API. The sample period starts from January 2006, the inception of the API data, and ends in February 2021. PressReader is a news aggregator service provider that uses its proprietary technology to process thousands of newspapers daily, extracting text and images and making articles instantly translatable and searchable (PressReader, 2021). We use PressReader's data for this study because it provides historical news coverage for many countries, including local and global news publishers.

Since we are interested in socioeconomic and political news content that may be informative about default risk, we proceed as follows. For each country, we retrieve all English language news articles that mention the country's name and one of the key words "economy" and "government" and published either in the country in question, the US, UK or China.[4] In total, we retrieved approximately 10 million news articles published by roughly 2,100 publishers covering 184 countries. Table A1 of the Appendix provides the country coverage and number of news

[4]We use a keyword search for news retrieval because the PressReader news API does not provide the functionality to filter news by topic. However, after we retrieve news data using the broad search, we train a topic model to isolate relevant news from noise. Including the US, UK and China when filtering news for the country of publication allows us to obtain news published by popular international media about each country. This is particularly helpful in scenarios where PressReader does not have extensive coverage of local news publishers for a country.

articles we obtained for each country. There is substantial heterogeneity in news coverage per country, arising from our focus on English language news and different starting moments of PressReader's coverage of local news publishers in each country.[5] We measure news sentiment, i.e., the negativity of each news article, as the number of negative words divided by the sum of negative and positive words from the Loughran and McDonald (2011) dictionary.

We obtain data on sovereign CDS spreads and countries' stock market indices from Datastream. We use the more liquid CDS contracts with five years tenor, full restructuring clause, denominated in USD, and referencing external debt. For stock market returns, we use country-level MSCI total return indices and consider both USD and local-currency denominated indices for each available country. Table A2 of the Appendix lists the index ticker per country considered in this study. We collect historical sovereign credit rating changes and the announcement dates from Trading Economics.[6]

We follow a similar approach as Augustin et al. (2021) to characterize a country's fiscal constraint using variables that capture fiscal capacity and economic conditions. We obtain annual data on countries' government debt-to-GDP ratio, public external debt-to-GDP ratio, interest-to-GDP ratio, unemployment rate, GDP growth, and credit ratings from the World Bank.[7] Since most of these variables are available only with a significant delay, we use them with a two-year lag. Next, for each of the variables and each year, we calculate a country's score, such that the country with the largest fiscal constraint (e.g., the highest debt-to-GDP ratio) has a value of one, and the country with the lowest fiscal constraint has a value of zero. In doing so, variables (GDP growth and credit ratings) for which higher values mean better economic/fiscal conditions are signed so that the rankings are comparable across metrics. Finally, we construct a single measure of fiscal constraint by averaging across the ranks of all six metrics.

[5]A possible extension of our framework, which we do not pursue in this study, is to use machine translation of news in languages other than English to significantly expand the news coverage of some non-English speaking countries with little news coverage in English. Note, however, that the framework we develop in this paper will work with little or no modification for such machine-translated news, making our work more appealing.

[6]`https://tradingeconomics.com/country-list/rating`.

[7]We use the average rating of foreign currency long-term sovereign debt from rating agencies Moody's, Standard & Poor's, and Fitch, as provided by the World Bank (Kose et al., 2017). Each agency's rating is converted to a numerical scale, with 1 representing the least creditworthy country and 21 the most creditworthy. The other variables are from the World Development Indicators.

We use the following global macroeconomic and financial indicators: world stock market index from MSCI, US Economic Policy Uncertainty Index (EPU) from Baker et al. (2016),[8] US macroeconomic activity index (ADS) of Aruoba et al. (2009) from the Federal Reserve Bank of Philadelphia,[9] and the VIX volatility index from the Chicago Board Options Exchange (CBOE). To mitigate the influence of outliers, we winsorize all continuous variables in our study at the 1% and 99% levels.

3 Measuring Sovereign Default Risk from News Articles

The sheer size of our news corpus, with roughly 10 million articles, presents a large text corpus to construct a text-based multi-country measure of sovereign default risk. However, text data is inherently high-dimensional and may contain substantial noise. Therefore, we start by pre-processing the text corpus to reduce noise by eliminating uninformative text as described in Appendix A.1. Next, we employ two NLP techniques to extract sovereign default risk signals from text. First, we train a topic model to further reduce noise and dimensionality. Second, we employ text similarity to quantify the sovereign default concern reflected in news articles.

3.1 Topic Model

Our news data relies on a broad search of news that uses the words "economy" or "government". However, some of the resulting news articles may only be tangentially related to the broad economy, business, and sociopolitical topics likely relevant to sovereign risk. Therefore, we use a topic model to isolate articles that belong to topics that are more likely to be informative about sovereign risk, hence reducing noise and dimensionality.

We employ the Latent Dirichlet Allocation (LDA) algorithm of Blei et al. (2003), which is widely used in the economics and finance literature for topic modelling (e.g., Bellstam et al., 2021; Bybee et al., 2021; Hanley and Hoberg, 2019; Huang et al., 2018). LDA gives text a hierarchical structure, where documents (news articles) are composed of topics which in turn contain words. Precisely, each document has a probability distribution over latent topics, with parameter $\alpha > 0$,

[8] https://www.policyuncertainty.com/.
[9] https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/ads.

and each topic is defined by a probability distribution over words with parameter $\beta > 0$. α controls the sparsity of topics in a document, while β controls the sparsity of words in a topic. LDA treats a document as a mixture of topics and a topic as a mixture of words, such that documents overlap each other rather than being separated into discrete groups.

Training the model essentially boils down to finding the optimal number of latent topics $\mathcal{T}$ that best fit the data. A topic model fitted on a corpus of documents with a chosen $\mathcal{T}$ yields two outputs: (i) the distribution of word frequencies for each topic, common across documents; and (ii) the distribution of topics across documents. For each document, the topic distribution is a vector of loadings that describe how intensively the topic is reflected in a given document, such that if a document has a higher loading for a particular topic, it is more likely associated with the topic.

We train the topic model by randomly selecting 70% of the news corpus as the training set and the remainder as the test set. Next, we convert the training set into a document term matrix whose rows are the news articles (documents) and columns the unique single words (unigrams) and two-word combinations (bigrams) in the training set corpus. We then search for the number of topics $\mathcal{T}$ that minimizes the test set perplexity score (maximizes the log-likelihood). We search for a low-level $\mathcal{T} \in \{4, 5, ..., 10\}$, as we are interested in broad topical themes, and find that the test set likelihood score increases with $\mathcal{T}$, although for $\mathcal{T} > 9$ the resulting additional topics are arguably sub-topics of broader themes. Hence, we set $\mathcal{T} = 9$ and manually assign the resulting topics meaningful labels based on their word distributions.

Next, we classify each news article into a topic based on its maximum probability score across the nine topics. Table 1 shows the topics and their most important terms. The "Economy", "Business", "Security", and two "Governance" topics—which are more likely to matter for sovereign default risk—account for 72% of the text corpus. In the next section, we restrict attention to these topics to compute the NSRI.

3.2 Text Similarity

The second component of our framework is quantifying how much sovereign risk concern each news article reflects. We use a lexical text similarity algorithm, namely cosine similarity (as in

Hanley and Hoberg, 2010; Engle et al., 2020), because of its simplicity, low computational cost on large data sets, and the difficulty of mapping our problem to supervised learning.[10] Cosine similarity relies on the word combinations in documents to quantify the similarity between documents. The algorithm quantifies the distance between two documents' vectors such that a value close to 1 indicates a high degree of similarity between the documents and a value close to 0 a low similarity. Cosine similarity is calculated as the inner product of two documents' vectors divided by the product of the vectors' euclidean lengths:

$$CS = \cos(\boldsymbol{u}, \boldsymbol{v}) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{|\boldsymbol{u}| \times |\boldsymbol{v}|} \qquad \in \quad [0, 1]. \tag{1}$$

Let $\boldsymbol{v}$ in Eq. (1) be the vector of a representative sovereign default risk document whose language usage reflects elevated sovereign default concerns, and $\boldsymbol{u}$ that of an arbitrary news article. Then, for any article represented by $\boldsymbol{u}$, CS will be close to 1 if the article signals elevated sovereign risk by using default risk-related language in similar proportions as in $\boldsymbol{v}$.

To specify $\boldsymbol{v}$, we follow a similar approach as Engle et al. (2020) and Hassan et al. (2019) to build a training library of documents that signal elevated sovereign default risk. In building this library, we take special care to focus on texts that are primarily centered on sovereign credit risk issues and default concerns. This is important because we are interested in constructing a measure that reflects sovereign default risk dynamics. Hence, we collect 448 documents, comprised of three textbooks on sovereign default,[11] 31 sovereign credit rating downgrade reports by the rating agencies Moody's, Standard & Poors and Fitch, 367 excerpts from the European Investment Bank's sovereign rating downgrade rationales and reports for low rated sovereigns, 22 country risk analysis reports by the Economist Intelligence Unit, and 28 debt sustainability analysis reports by the World Bank and IMF as well as news articles on sovereign defaults and

[10] We experimented with supervised learning by training linear classifiers (penalized logistic regression and support vector machine) on a binary transformation of CDS spreads (outcome variable) on a feature space comprising unigrams and bigrams. However, we found that supervised learning could not effectively separate elevated sovereign risk concern periods reflected in CDS spreads from other periods. For example, the best classifier achieved an out-of-sample Area Under the Curve (AUC) score of 0.66. Reasons for this weak performance include micro-structure noise in CDS spreads and news re-iterations from different media outlets, which creates misalignment in the dating of news events and the CDS market's potential reactions.

[11] The books are (i) "Sovereign Debt: from Safety to Default" by Robert W. Kolb; (ii) "Sovereign Debt: A Guide for Economists" by Ali Abbas, Alex Pienkowski, And Kenneth Rogoff; (iii) "Why not Default? The Political Economy of Sovereign Debt" by Jerome Roos.

debt problems. We preprocess the documents using steps 2 – 5 in Appendix A.1 and then aggregate them into a representative sovereign default risk library d.

Next, to extract $\boldsymbol{v}$ from d, we first convert our news corpus into a "term frequency–inverse document frequency" (*tf-idf*) normalized matrix $\mathbf{X}$ using the Bag of Words text vectorization.[12,13] The rows of $\mathbf{X}$ are the $\boldsymbol{u}$'s in (1) and the columns are the unique single words (unigrams) and two-word combinantions (bigrams) in our news corpus. We then transform d to $\boldsymbol{v}$ using the text vectorizer and *tf-idf* weights fitted on $\mathbf{X}$, which ensures that the document-frequency weights of d match that of our news corpus.

Finally, we score each news article using Eq. (1) and re-scale the resulting *CS* scores to have a maximum value of 1 and a minimum of 0 in the sample. We then obtain the country-level News-implied Sovereign Risk Index (NSRI) by aggregating *CS* over a time interval. The preponderance of socioeconomic news served in real-time by several news feed providers suggests that the NSRI aggregation interval can range anywhere from a few minutes to several days, depending on the frequency of news availability. In this paper, we aggregate NSRI at a daily frequency by averaging across all news that belongs to the sovereign risk-relevant topics:

$$NSRI_{j,t} = \frac{1}{n^r_{j,t}} \sum_{i}^{n^r_{j,t}} CS_{i,j,t}, \tag{2}$$

where $CS_{i,j,t}$ is the cosine similarity score of article i published on day t about country j, and $n^r_{j,t}$ is the number of news articles in the sovereign risk-relevant topics published about country j on day t. We consider news articles in the "Economy", "Business", "Governance", and "Geopolitics" topics relevant for sovereign risk and use only those in (2).

Note that although our framework and its implementation rely solely on English language news and countries with news data from PressReader, the framework is applicable with little or no modification to alternative news data and non-English language news. The fact that our news

[12]Following standard NLP practice aimed at reducing noise and dimensionality, we omit terms (unigrams and bigramns) that occur very rarely in the text corpus — terms that occur in less than 10,000 documents.

[13]For a given document i, *tf-idf* normalization gives low weights to terms that are commonplace in all documents and high weights to terms that appear frequently in that specific document. It is computed as $\textit{tf-idf}(x, i) = \textit{tf} * \log(N + 1)/(N_x + 1) + 1$, where N is the number of documents in our news corpus, N_x is the number of documents in which term x appears, and *tf* is the number of times x appears in document i.

data covers several media outlets from various countries allows our framework to not depend on specific news outlets, enhancing generalization. Therefore, one can seamlessly apply our framework to other news sources, besides PressReader, with broader coverage. For non-English language news, one option is to translate the news to English using language translation APIs, then feed the resulting English text through our framework. Given that language translation algorithms are still developing, we anticipate that this option may result in potentially noisier default signals which will get less noisy with improvement in translation algorithms. We leave this interesting application for future research.

3.3 Framework Validation

We conduct a multi-pronged validation exercise to establish NSRI's robustness and informativeness about sovereign default risk. Starting with the topic model, Table 1 shows the top-20 relevant unigrams and bigrams for the identified topics. We measure term relevance using the approach of Sievert and Shirley (2014) based on the rescaled topic-specific term probabilities and setting the scaling parameter $\lambda = 0.5$. Table 1 shows that the most relevant words for each topic are distinct enough to permit assigning meaningful labels to the topics, which allows us to filter out news unlikely to matter for sovereign risk. For instance, Topic 9, which we label "Economy", includes terms such as "market", "bank", "economy", and "debt", which are likely relevant for sovereign risk. In contrast, Topic 8, which we label "Tourism", includes terms such as "tourist", "food", "flight", and "hotel", which are unlikely to be informative about sovereign risk. Therefore, by isolating news on the Environment, Health, Tourism and Miscellaneous topics, the topic model allows us to eliminate noise that can swamp the information about sovereign default risk potentially reflected in the news on the Security, Governance, Business, and Economy topics.

Next, we examine the most influential terms in our sovereign default risk library d used to obtain article-level scores. Figure 2 depicts the unigrams and bigrams with the highest weights (in vector $\boldsymbol{v}$ in Eq. (1)), indicating that words such as "sovereign", "default", "public debt" "crisis", "risk", "fiscal", "borrowing", "creditor", "currency", "repayment", "credit rating", and "downgrade", which are likely to figure prominently in discussions about sovereign credit risk, rank highly. We also observe other economic and financial terms that are likely to be used

S/n	Topic 1 Environment	Topic 2 Security	Topic 3 Miscellaneous	Topic 4 Governance I	Topic 5 Health	Topic 6 Business	Topic 7 Governance II	Topic 8 Tourism	Topic 9 Economy
1	oil	police	woman	right	health	business	president	water	market
2	energy	attack	world	law	pandemic	development	minister	food	bank
3	climate	kill	school	party	virus	project	leader	flight	growth
4	climate change	rebel	family	political	medical	technology	prime minister	tourist	economy
5	fuel	force	team	people	coronavirus	service	war	passenger	rate
6	coal	investigation	sports	court	vaccine	company	peace	tourism	investor
7	mining	intelligence	game	public	covid	sector	millitary	airline	tax
8	oil gas	report	cup	election	hospital	industry	sanction	ship	financial
9	emission	security	play	vote	disease	cooperation	opposition	airport	trade
10	electricity	arrest	love	corruption	patient	programme	administration	hotel	economic
11	oil price	government	film	government	outbreak	opportunity	state	road	debt
12	carbon	terrorist	book	act	infection	digital	meeting	island	price
13	power	millitary	friend	bill	lockdown	international	deal	transport	stock
14	renewable	millitant	football	freedom	people	local	political	travel	investment
15	natural gas	violence	player	state	doctor	partnership	nation	traffic	income
16	solar	army	young	rule	death	innovation	conflict	sea	currency
17	natural	civilian	life	voter	treatment	system	diplomatic	park	export
18	petroleum	crime	home	legal	health care	economic	international	vessel	sale
19	crude	protest	old	democracy	worker	infrastructure	negotiation	boat	gdp
20	gas	soldier	child	president	test	country	ally	route	business
% of News	4%	12%	12%	15%	6%	11%	15%	7%	17%

Table 1: Most Important Terms for Topic Labeling. The table shows the most important terms for the nine latent topics identified by our topic model. The columns indicate the topics and their assigned labels based on the word distributions. The last row shows the fraction of articles in our news corpus that belong to each topic. Each news article is assigned to the topic for which it has the highest probability score.

in context when discussing sovereign credit risk, including “interest”, “lender”, “economic”, “GDP”, and “bank”. Intuitively, a news article that uses combinations of these terms in a high proportion as our sovereign default risk library will have a high *CS* score.

To examine whether our news-level scores sufficiently differentiate news articles signalling elevated sovereign risk concerns from others, we randomly select three sets of news articles that are publicly available on the internet. We narrowed the search as follows: the first set comprises news articles related sovereign default concerns; the second relates to news articles on socio-economic problems; and the third comprises miscellaneous news unrelated to the former two. Importantly, these articles were not part of our training library. Next, we score each news article using Eq. (1) and assign it to a topic using our topic model. Furthermore, to evaluate whether our algorithm simply picks up positive/negative sentiment in the news, we compute a sentiment score for each article.[14]

Table 2 shows the result of this exercise. As the article titles in the table indicate, Panel A contains results for the set of articles specifically discussing sovereign default and credit risk issues. Panel B contains articles on socioeconomic problems which are not directly related to sovereign credit risk, and Panel C contains miscellaneous news unrelated to sovereign credit risk.

[14] We compute sentiment using the Loughran and McDonald (2011) dictionary as the number of negative words in a news article divided by the sum of positive and negative words in the article.

Figure 2: Most Important Terms in the Sovereign Default Risk Library. The figure shows the terms with the highest *tf-idf* scores in our Sovereign Default Risk Library. Font size is proportional to term weight.

Indeed, our algorithm assigns higher *CS* scores (Column (2)) to the default risk-related articles in Panel A compared to the rest (Panels B and C). Furthermore, as expected, the articles in Panel B are on average scored higher than the miscellaneous articles in Panel C. Comparing the sentiment scores (Column (3)) to the *CS* scores in (Column (2)), it is clear that *CS* is markedly different from sentiment, indicating that our algorithm does not simply capture negative news sentiment. Finally, Column (3) shows that our topic model correctly classifies news articles into reasonable topics, providing strong support for our reliance on the model for reducing noise.

We now turn to the NSRI's time-series dynamics to examine whether our algorithm picks up important periods of elevated global sovereign default concerns. Figure 1 already showed that this is the case for selected countries. Next, we aggregate the country-level NSRIs to the global level by taking a weighted average of the daily values using countries' stock market capitalization (in dollars) as of the end of the last calendar year as weights.[15] Figure 3 depicts the resulting **G**lobal **N**ew-implied **S**overeign **R**isk **I**ndex (GNSRI). GNSRI has rich time-series dynamics,

[15]We use market capitalization as weights for computing GNSRI to give more importance to countries with larger capital markets whose financial market and economic policy developments have stronger effects on the global economy. If a country's market capitalization is missing for the last calendar year, we use the most recent value before the last calendar year. If the market capitalization is missing throughout the sample period, we use the cross-sectional median of market capitalization for each year.

News Article Title	(1) CS	(2) Sentiment	(3) Topic
Panel A: News signaling sovereign default concerns			
Zambia defaults, economically and politically	0.54	0.79	Governance I
Zambia on brink of defaulting on foreign debt	0.81	0.72	Economy
Zambia's Eurobond default - What we have learned	0.69	0.82	Economy
Zambia's default fuels fears of African 'debt tsunami' as Covid impact bites	0.80	0.74	Economy
Argentina Defaults on Sovereign Debt Amid Coronavirus Crisis	0.69	0.98	Economy
Argentina defaults yet again, but hopes to get off lightly	0.63	0.94	Economy
Already in default, Argentina hits an impasse with creditors over debt restructuring	0.62	0.82	Governance II
'Painful' downgrades will raise South Africa's borrowing costs, minister says	0.55	0.90	Economy
What the latest rating downgrades mean for the average South African	0.64	0.87	Economy
Moody's, Fitch further downgrade South Africa	0.40	0.75	Economy
World Bank says world leaders moving away from debt cancellation for Africa	0.67	0.88	Business
In a surprise move, Fitch upgraded South Africa's banks - here's why	0.52	0.66	Economy
Panel B: News on socioeconomic problems			
Myanmar: Military Coup Kills Fragile Democracy	0.13	0.91	Security
Myanmar coup: No sign of end to brutal crackdown on all fronts	0.12	0.80	Security
Nigeria's Economy Faces Worst Recession in Four Decades, says New Report	0.23	0.73	Economy
Recession: Nigeria's Economic Crisis Requires a Political Solution	0.31	0.69	Economy
Turkey's Erdogan sacks central bank governor after rate hike	0.11	0.50	Economy
Venezuela hyperinflation hits 10 million percent. 'Shock therapy' may be only chance	0.42	0.83	Economy
Venezuela crisis in 300 words	0.08	0.57	Governance II
Venezuela crisis: How the political situation escalated	0.08	0.87	Governance I
Chadian President Idriss Deby dies on frontline, rebels vow to keep fighting	0.13	0.78	Governance I
Nigeria's inflation rises again, hit four-year high February	0.12	0.50	Economy
Panel C: Miscellaneous news			
Minister orders intelligence-led operation against violence in South-East	0.03	0.89	Security
Man who drove into News Cafe in Rosebank appears in court	0.03	0.98	Security
Five arrested for murder of KZN farmer	0.02	0.77	Security
Uber pledges to boost safety for SA drivers as accidents rise	0.07	0.67	Health
UK imposes sanctions on Russians, Guptas in first use of anti-corruption law	0.09	0.86	Security
Naomi Osaka withdraws from French Open amid row over press conferences	0.09	0.80	Miscellaneous
Chelsea beat Man City to win Champions League	0.07	0.58	Miscellaneous

Table 2: News-level scores and Topics. The table shows the cosine similarity (CS) scores, sentiment scores and topics for a randomly selected sample of news articles that were *not* part of our training library. Panel A contains articles highlighting sovereign default concerns, Panel B contains articles highlighting socioeconomic problems, and Panel C shows miscellaneous news unrelated to the economy or sovereign credit risk. "Sentiment" captures the negativity of each news and is computed as the number of negative words divided by the sum of negative and positive words using the Loughran and McDonald (2011) dictionary.

spiking during the 2009 financial crisis that sowed the seeds of the European debt crisis. GNSRI then attained its peak at the height of the European debt crisis (2011 – 2012). It spiked again in 2020 at the height of the Covid-19 pandemic during which the global economy experienced an unprecedented shutdown and governments launched massive fiscal packages, plunging many countries into fiscal problems that created sovereign debt concerns and a few defaults. Overall, the evolution of GNSRI provides strong evidence that our news-based measure of sovereign default risk captures major developments in global sovereign credit risk.

Finally, Table 3 shows the descriptive statistics for NSRI by income level, indicating substantial cross-sectional heterogeneity. For instance, the mean (median) daily NSRI for the

Figure 3: Global News-implied Default Risk. The figure shows the weekly time series evolution of the Global News-implied Risk Index (GNSRI). GNSRI is computed as the time t weighted average of the country-specific NSRIs, with the country weights based on the preceding year's dollar stock market capitalizations. The time series is then re-scaled to range between 0 and 1. The plotted values are three-month moving averages.

	Countries	Obs.	Mean	SD	Min	5%	10%	25%	50%	75%	90%	95%	Max
High income	57	150,156	0.151	0.075	0.008	0.064	0.078	0.102	0.136	0.180	0.241	0.295	0.894
Upper middle income	48	110,106	0.116	0.051	0.008	0.056	0.067	0.085	0.107	0.136	0.169	0.197	0.953
Lower middle income	51	89,120	0.113	0.049	0.007	0.051	0.063	0.084	0.107	0.134	0.164	0.191	0.841
Low income	27	46,231	0.095	0.044	0.007	0.044	0.054	0.069	0.087	0.109	0.141	0.169	0.825

Table 3: NSRI Summary Statistics by Income Level. The table shows the summary statistics of the daily NSRI by income level, following the World Bank's country classification. "Countries" indicates the number of countries in each group for which we have news data. The sample period is January 2006 – February 2021.

high-income countries is 0.15 (0.14), which declines to around 0.11 (0.11) for the middle-income countries and 0.10 (0.09) for the low-income countries. It is not surprising that NSRI, on average, is higher for the high-income countries. In our sample period (2006 – 2021), the major sovereign debt crises and the associated elevated default concerns mainly occurred among high-income countries, such as Spain, Italy, Portugal, and Greece. In summary, the validation exercises indicate that NSRI effectively reflects sovereign default concerns by exploiting narratives about sovereign credit risk in the news media.

4 Empirical Analysis

4.1 NSRI's Informativeness about CDS Spreads and Rating Downgrades

We saw in the previous section that NSRI has exciting properties, exhibiting cross-sectional and time-series variations that align with reasonable priors. We now turn to an analysis of the relationships between NSRI and the two standard indicators of sovereign credit risk—CDS spreads and credit ratings—to understand if and to what extent NSRI captures information contained in these variables.

Informativeness about CDS Spreads. CDS spread data is available only for a limited cross-section of countries with liquid CDS contracts. For these countries, we examine how the percentage change of the weekly/monthly averaged CDS spreads relate to the percentage change of the weekly/monthly averaged NSRI, using panel contemporaneous and predictive regressions. We use percentage changes instead of levels to prevent cross-country differences in the levels and volatilities of CDS spreads from influencing our results. Furthermore, we use a weekly/monthly frequency because the news data used to compute NSRI are from various sources with different time zones. By aggregating to a lower frequency, we side-step potential complications surrounding whether the news were published before or after the market trading hours. We use the following regression specification:

$$\Delta CDS_{j,t} = \beta_0 + \beta_1 \Delta NSRI_{j,t-\tau} + \boldsymbol{\gamma X} + \xi_j + \lambda_t + \epsilon_{j,t}, \tag{3}$$

where $\Delta CDS_{j,t}$ is the percentage change of country j's CDS spread in period t relative to $t-1$, $\Delta NSRI_{j,t-\tau}$ is the percentage change of country j's NSRI relative to period $t-1$ or its past three-month median. $\tau = 0$ for contemporaneous regression and $\tau = 1$ for one-period predictive regression. $\boldsymbol{X}$ is a vector of control variables: percentage changes of global macro-financial variables—the implied volatility index (VIX), US economic policy uncertainty Index (EPU), US economic activity index (ADS), and the world stock market return—that may drive international CDS spreads' dynamics (e.g., Pan and Singleton, 2008; Longstaff et al., 2011; Ang and Longstaff, 2013). $\boldsymbol{X}$ also includes country-specific controls, namely stock market return, volatility and news sentiment. λ_t is the year-month (year) fixed-effect for the weekly (monthly) frequency panel to absorb slower-moving common trends in CDS spreads. ξ_j is the country fixed-effect to absorb time-invariant characteristics that may influence CDS spread dynamics. We normalize all continuous regressors to unit variance. The sample includes only countries with liquid CDS contracts. We further restrict the analysis to country-week observations with at least 60 news articles in a calendar month to reduce noise and ensure sufficient information to compute the NSRI.[16,17]

Table 4 shows the regression results. Panel A shows results for the weekly frequency, while Panel B shows results for the monthly frequency. Starting with percentage changes relative to the previous period, we see that NSRI has a statistically significant and economically sizeable association with CDS spreads. Columns (1) and (2) of Table 4 show that a standard deviation

	NSRI % change rel. to $t-1$				NSRI % change rel. to 3-month median			
	Contemporaneous		Lagged		Contemporaneous		Lagged	
Dep. var.: CDS Spread % change	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Weekly frequency								
NSRI % change	0.399	0.258	-0.055	-0.095	0.391	0.213	-0.132	-0.158
	(5.05)	(4.39)	(-1.15)	(-2.12)	(5.35)	(3.58)	(-1.91)	(-2.43)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
R^2 (%)	25.45	38.73	25.19	28.71	24.86	38.04	24.64	28.32
Obs.	14,004	14,004	14,004	14,004	13,779	13,779	13,779	13,779
Panel B: Monthly frequency								
NSRI % change	3.117	1.713	1.495	1.021	4.804	2.384	1.673	1.100
	(7.11)	(9.59)	(3.53)	(3.47)	(9.08)	(8.62)	(3.87)	(3.27)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
R^2 (%)	10.81	46.33	8.66	18.99	14.67	47.64	8.97	19.27
Obs.	3,015	3,015	3,015	3,015	3,073	3,073	3,073	3,073

Table 4: CDS Spread and News-implied Sovereign Risk Index. The table shows results for the panel regression of percentage changes in countries' CDS Spreads (in %) on percentage changes in NSRI at the weekly (Panel A) and monthly (Panel B) frequencies. Percentage changes are based on average values over each frequency. For the first set of results (Columns (1) – (4)), the percentage change in NSRI is relative to the previous period. For the second set of results (Columns (5) – (8)), the percentage change in NSRI is relative to the past three-month median. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. Control variables are the one-period lag of the dependent variable, the weekly average of country j's news sentiment, country j's weekly stock market return and realized volatility, the weekly return of the MSCI world stock market index, weekly percentage changes of the US economic policy uncertainty index, the US economic activity index, and the implied volatility index. Continuous regressors are normalized to unit variance, and all regressions include country and time fixed effects (year-month fixed effects in Panel A and year fixed effects in Panel B). Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021.

increase in the NSRI percentage change relative to the previous week (month) is associated with an increase in CDS spreads by between 0.26% and 0.40% (1.71% and 3.12%) in the same period. Furthermore, Columns (3) and (4) show that NSRI also predicts CDS spreads dynamics. At the weekly frequency, the prediction is negative and small in magnitude, while the predictability is

[16]The final sample includes the following 39 countries Austria, Brazil, Chile, China, Colombia, Costa Rica, Croatia, Denmark, Egypt, France, Germany, Greece, Hungary, Indonesia, Israel, Italy, Japan, Malaysia, Mexico, Netherlands, Norway, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Saudi Arabia, Slovakia, Slovenia, South Africa, Spain, Sweden, Thailand, Turkey, UK, USA, and Venezuela

[17]Requiring at least 60 news articles per month is equivalent to an average of two articles per day, which is a reasonable requirement to avoid noise in the data. Our analysis is, however, robust to altering the threshold as we show in the robustness Section 5.

positive, sizeable, and strongly significant at the monthly frequency. For instance, a standard deviation increase in the NSRI percentage change predicts an increase in CDS spreads by 1.02% – 1.50% over the next month. This suggests that shocks to sovereign default risk concerns reflected in the news trigger insurance sellers to demand a higher premium for insurance against default, resulting in higher CDS spreads. At short windows (one week), the initial increase in CDS premiums tends to subside, indicating a potential overreaction to news, before increasing further as more news about default concerns emerge.

Columns (5) – (6) of Table 4 present further evidence that the association between NSRI and CDS spreads is robust. In these latter specifications, we use more persistent changes in NSRI by computing NSRI percentage change relative to the past three-month median. The results are even stronger in magnitude and statistical significance. For example, at the monthly frequency, a standard deviation increase in the NSRI percentage change relative to its past three-month median is associated with an increase in CDS spreads by between 2.38% and 4.80% contemporaneously and predicts an increase in CDS spreads by 1.10% – 1.67% over the next month. Overall, the analyses indicate that NSRI reflects information about sovereign default risk captured by CDS spreads — the predominantly used proxy for sovereign default risk at a daily frequency.[18] The results, therefore, point towards the possibility of using NSRI as a high-frequency proxy of default risk for countries without liquid CDS contracts.

Informativeness about Sovereign Rating Downgrades. To reinforce our claim that NSRI can be used as a proxy of sovereign default risk, we test if NSRI is informative about future sovereign credit rating downgrades. Like CDS spreads, sovereign credit ratings reflect the likelihood that a sovereign will default on its debt obligations in the future, such that sovereigns with poor credit ratings are perceived to have higher default risk. However, unlike CDS spreads, credit ratings are available for a much broader cross-section of countries and do not exhibit daily time-series variation. Instead, rating agencies update credit ratings typically during pre-specified intervals and significant economic fundamentals changes that weaken a sovereign's creditworthiness. Hence, if NSRI is truly informative about sovereign default risk, it should aggregate

[18]We conduct additional robustness analysis, using only observations for which the country has at least 90 news articles in the calendar month, to further reduce noise, and by dropping one country at a time from the panel regression to ensure that an outlier country does not drive the results. The results presented in Tables A3, A4 and A5, and discussed in Section 5, strongly reinforce the main results.

information about default concerns leading up to a sovereign rating downgrade announcement and therefore should predict rating downgrades. We test whether this is the case using historical sovereign rating information for all available countries.

For each country j we construct a dummy variable $Downgrade_{j,t}$ that equals 1 if any of rating changes on day t is a downgrade and 0 otherwise. We consider ratings by the three main credit rating agencies, Moody's, Standard & Poor's and Fitch, and define downgrade as any change in credit ratings towards a lower credit quality, including changes in rating outlook. We then estimate the following regression specification:

$$Downgrade_{j,t} = \beta_0 + \beta_1 NSRI_{j,t-\tau} + \boldsymbol{\gamma X} + \xi_j + \epsilon_{j,t}, \tag{4}$$

where j and t index country and time respectively, $Downgrade_{j,t}$ is the dummy variable for downgrade conditional on rating change on day t, and $NSRI_{j,t-\tau}$ is country j's average NSRI over the past 30-day window ending $t-\tau$, where $\tau = 2$ in the baseline analysis. To ensure that differences in the level of NSRI across countries do not drive the results, we normalize the time series of each country's daily NSRI to have a minimum value of 0 and a maximum value of 1 before computing the averages. $\boldsymbol{X}$ is a vector of control variables computed over the same window as $NSRI_{j,t-\tau}$: global macro-financial variables as before (VIX, EPU, ADS, and world stock market return) and country-specific controls, namely stock market return, volatility, news sentiment, and CDS spread. ξ_j captures time-invariant country fixed-effect. We normalize continuous regressors to unit variance and estimate the regression using ordinary least squares (OLS), which allows us to include fixed effects without raising an incidental parameters issue. We provide robustness results based on logistic regression in Section 5.

Table 5 shows the estimation results. Across different specifications, NSRI significantly predicts sovereign rating downgrades. A standard deviation increase in average NSRI over the 30-day window ending $t-2$ predicts an increase in the probability of a downgrade by $8-12$ pp, which corresponds to a $15\% - 22\%$ increase in the likelihood of a downgrade relative to the observed fraction of downgrades in the data. Importantly, this result holds after controlling for a set of global and country-specific macro-financial variables, including news sentiment and CDS spreads.

	(1)	(2)	(3)	(4)	(5)	(6)
NSRI	0.114	0.115	0.092	0.092	0.080	0.088
	(6.51)	(6.63)	(4.31)	(4.31)	(3.78)	(3.76)
Sentiment		0.051	0.048	0.047	0.037	0.092
		(3.12)	(1.80)	(1.77)	(1.38)	(3.60)
Ret			-0.017	-0.023	-0.025	-0.044
			(-1.45)	(-1.23)	(-1.44)	(-2.62)
Vol			0.155	0.158	0.115	0.119
			(9.46)	(8.95)	(5.35)	(3.98)
World Ret				0.011	0.031	0.066
				(0.48)	(1.48)	(4.08)
EPU					0.041	0.044
					(2.08)	(1.54)
ADS					0.010	0.012
					(0.57)	(0.65)
VIX					0.056	0.051
					(1.97)	(1.20)
CDS Spread						0.008
						(0.46)
R^2 (%)	16.5	17.3	27.5	27.5	28.9	33.5
Obs.	2,605	2,605	1,716	1,716	1,716	1,074

Table 5: Sovereign Rating Downgrades and News-implied Sovereign Risk Index. The table shows results for the panel regression of an indicator variable of rating downgrade ($Downgrade_{j,t}$) on the NSRI averaged over the past 30-day window ending $t-2$ before rating change on day t. $Downgrade_{j,t}$ equals equals 1 if a country's rating change on t is a downgrade and 0 otherwise. The other regressors are computed over the same 30-day window ending $t-2$ as NSRI. Sentiment is the average of country j's news sentiment. Ret is country j's cumulative stock market return. Vol is country j's stock market volatility. World Ret is the cumulative return of the MSCI World stock market index. EPU is the average US economic policy uncertainty index. ADS is the average US economic activity index. VIX is the average implied volatility index. CDS Spread is the average of country j's CDS Spread. We normalize continuous regressors to unit variance, and all regressions include country fixed-effects. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 - February 2021.

Columns (1) – (5) of Table 5 show once again that NSRI does not simply reflect news sentiment, given that we directly control for news sentiment, which turns out to be insignificant in most specifications. Furthermore, Column (6) of Table 5 shows that conditional on NSRI, CDS spreads do not provide additional significant information about future sovereign downgrades. Conversely, NSRI remains strongly significant with a sizeable effect size after controlling for CDS spreads. This result indicates that NSRI captures incremental default risk information not fully reflected in either CDS spreads or the other macro-financial variables.[19] A potential explanation for this is that the news text on which the NSRI is based reflects agents' default

[19]We conduct additional robustness analyses presented in Tables A6 and A7, and discussed in Section 5, by averaging NSRI over the past 60-days ending $t-2$ before downgrade announcement on t and by estimating Eq. (4) using logistic regression. The results reinforce the main findings.

concerns more timely before those concerns are fully incorporated in CDS spreads, ultimately leading to downgrades.

To summarize, the analyses in this section and several robustness analyses discussed in Section 5 provide strong evidence that NSRI is a reliable proxy for sovereign default risk. It captures CDS dynamics across countries and is informative about future rating downgrades. Importantly, NSRI is not redundant in the face of CDS spreads and other macro-financial variables. Hence, NSRI provides complementary default risk information for countries with liquid market-based measures, such as CDS contracts, and can directly serve as a high-frequency default risk proxy for those countries lacking liquid market-based default risk measures.

4.2 Sovereign Default Concerns and Equity Returns

We have shown that NSRI is a reliably proxy for sovereign default risk. In this section, we use NSRI to examine the role of sovereign default concerns in equity markets around the world. The theoretical literature has proposed different channels through which sovereign default (e.g., Gennaioli et al., 2014; Cole and Kehoe, 1998) and heightened default risk (Jeanneret, 2017) influence output, investments, and equity valuations. Jeanneret (2017) suggests a risk-based channel where a rise in sovereign default risk translates into a threat of future economic slowdown, which in turn impacts equity prices across countries. However, a different channel supported by behavioural explanations (e.g., Dremen, 1983; De Bondt and Thaler, 1985; Gennaioli et al., 2015) is investors' overreaction to news of sovereign default concerns that may not be strongly associated with economic fundamentals. Overreaction occurs, for example, when investors become excessively pessimistic about a country's economic fundamentals following a string of default concern news, hence depressing equity prices. However, as more news emerges, which turn out to be better than the previous forecasts now identified as too pessimistic, prices adjust leading to a reversal of the initial decline.

Since NSRI is a news-based measure of sovereign risk covering a sizeable cross-section of countries, it is well suited to determine whether sovereign default risk concerns affect equity returns through the risk-based or overreaction channel. If sovereign default risk affects equity markets by primarily reflecting a higher likelihood of future economic slowdown, we expect its impact on equity valuations to be more persistent, lasting several months. Conversely, if investors mainly overreact to news of sovereign default concerns, we expect a reversal of NSRI's influence on equity returns over a short time frame. We use the following regression specification to examine these hypotheses:

$$Ret_{j,t} = \beta_0 + \beta_1 \Delta NSRI_{j,t-\tau} + \boldsymbol{\gamma} \boldsymbol{X} + \xi_j + \lambda_t + \epsilon_{j,t}, \tag{5}$$

where $Ret_{j,t}$ is country j's stock market index return on week t and $\Delta NSRI_{j,t-\tau}$ is the percentage change in the country's NSRI in week $t-\tau$ relative to its median over the past three-month period, $\tau = 0$ for contemporaneous regression and $\tau = 1$ for one-week predictive regression. $\boldsymbol{X}$ is a vector of control variables that include news sentiment, world stock market return, and one-period lag of the dependent variable. Given the lack of high-frequency macroeconomic variables for most countries, we follow Edmans et al. (2021) and Gao et al. (2020) to use US variables to capture economic conditions. Although US macro-financial variables does not fully capture local economic conditions, they likely reflect sizeable local macro-financial dynamics, given the finding of Brusa et al. (2020) that US macroeconomic policy has a stronger impact on foreign country stock markets than local macroeconomic policy. Hence, we control for percentage changes in the same US macro-financial variables in Eq. (3), namely VIX, EPU, and ADS. ξ_j captures other time-invariant country characteristics that may influence the equity market, and λ_t captures year-month fixed effects to control for time-varying global drivers of the equity market. We normalize continuous regressors to unit variance and cluster standard errors by country. We restrict the analysis to country-week observations with at least 60 news articles in a calendar month to reduce noise and ensure sufficient information to compute NSRI.[20]

We use weekly returns for the analysis because, as earlier discussed, the news data used to compute NSRI are from various sources with different time zones. Hence, by aggregating weekly, we side-step potential complications surrounding whether the news was published before or after market trading hours. Furthermore, in this section and the remainder of the paper, we use NSRI growth relative to its three-month median for three reasons: first, to control for country-level differences in the average level of NSRI; second, to capture more persistent news shocks compared

[20] In the robustness analysis, Section 5, we show that this threshold does not materially influence our results.

to week-on-week changes; third, because results from Table 4 shows that, compared to week-on-week percentage change, the NSRI percentage change relative to the past three-month median is more informative about future default risk dynamics captured by CDS spreads.

Table 6 shows the results of estimating Eq. 5. Panel A shows results for the dollar-denominated stock market index returns. Focusing on Column (2), we find a negative contemporaneous association between stock market returns and sovereign default risk concerns reflected in NSRI. A standard deviation increase in the NSRI percentage change is associated with a 7.1 bps (3.7% annualized) decline in same-week market returns. Columns (3) and (4) report results for the one-week lagged NSRI percentage change, indicating a reversal of the initial decline in equity returns following an increase in sovereign default risk concerns. Precisely, Column (4) shows that a standard deviation increase in the NSRI percentage change predicts an increase in market returns by 6.4 bps (3.4% annualized) over the next week, indicating a price-reversal pattern consistent with investors' overreaction to news of sovereign default concerns.

For robustness, Panel B of Table 6 presents results for estimations using local-currency denominated stock market index returns. The results are similar and comparable to those in Panel A. Once more, we find that stock market returns have a negative contemporaneous association with an increase in sovereign default concerns, which reverses over the next week. Further robustness analyses—dropping one country at a time from the panel regression to ensure that an outlier country does not drive the results, using only observations with at least 90 news articles in the calendar month to potentially reduce noise in NSRI, and stopping the sample in 2019 to exclude the Covid-19 period—presented in Tables A8 and A9 and discussed in Section 5 support the finding that equity investors tend to overreact to news of sovereign default concerns.

4.2.1 Fiscal Constraint, Sovereign Default Concerns and Equity Returns

While Table 6 indicates that, on average, equity markets overreact to an increase in sovereign default concerns in the news media, there might be heterogeneity in the response to countries' default concerns. For instance, consistent with theories on the economic costs of high debt levels and default risk, the equity markets of countries with tight fiscal constraints, and hence

	Panel A: Dollar index return				Panel B: Local-currency index return			
	Contemporaneous		Lagged		Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NSRI	-9.316	-7.120	7.560	6.443	-6.111	-3.978	6.294	5.466
	(-4.25)	(-3.50)	(2.75)	(2.39)	(-3.14)	(-2.31)	(2.70)	(2.37)
Ret($t-1$)		-14.100		-26.247		-13.159		-21.231
		(-2.99)		(-4.16)		(-2.80)		(-3.77)
Sentiment		-11.684		10.474		-8.677		9.305
		(-3.00)		(2.93)		(-2.35)		(2.77)
World Ret		179.999		-18.791		139.780		-16.724
		(14.00)		(-3.13)		(13.07)		(-3.35)
EPU		-5.446		0.363		-2.468		2.408
		(-4.19)		(0.23)		(-1.86)		(2.08)
ADS		-8.164		6.681		-1.797		5.108
		(-4.76)		(4.27)		(-1.32)		(3.59)
VIX		-14.838		13.745		-24.143		11.853
		(-3.27)		(4.26)		(-5.97)		(4.66)
R^2 (%)	11.74	36.78	11.72	13.40	10.03	32.50	10.03	11.61
Obs.	24,101	24,101	24,101	24,101	24,004	24,004	24,004	24,004

Table 6: Equity Returns and Sovereign Risk Concerns. The table shows results for the panel regression of countries' weekly stock market returns (Ret), in basis points, on the percentage change of the weekly-averaged NSRI relative to its past three-month median. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. Panel A shows results for US dollar stock market indices, and Panel B shows results for local-currency stock market indices. Ret($t-1$) is the one period lag of country j's weekly stock market return. Sentiment is the weekly average of country j's news sentiment. World Ret is the weekly return of the MSCI World stock market index, EPU is the percentage change in the weekly-averaged US economic policy uncertainty index. ADS is the percentage change in the weekly-averaged US economic activity index. VIX is the percentage change in the weekly-averaged implied volatility index. We normalize continuous regressors to unit variance and include and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021.

limited ability to accommodate adverse fiscal shocks, may be impacted more by elevated default concerns. We explore this conjecture to further tease out the risk versus overreaction hypotheses.

To examine the heterogeneity in the equity markets' response to news of default concerns, we use a similar specification as Eq. (5):

$$\begin{aligned} Ret_{j,t} = \beta_0 + \beta_1 \Delta NSRI_{j,t-\tau} + \beta_2 \Delta NSRI_{j,t-\tau} \times HighFiscalConstraint_{j,t} \\ + HighFiscalConstraint_{j,t} + \boldsymbol{\gamma X} + \xi_j + \lambda_t + \epsilon_{j,t}, \end{aligned} \tag{6}$$

where, $HighFiscalConstraint_{j,t}$ is a dummy variable that equals 1 if the measure of a country's fiscal constraint is higher than the sample median and 0 otherwise. The measurement of countries' fiscal constraints follows a similar approach as Augustin et al. (2021) and is described in Section 2. The rest of the variables are as described under Eq. (5).

In Eq. (6), we are interested in the estimates of β_1 and β_2, which are reported in Table 7. We find that the coefficient of the interaction term is sizable and significant, indicating heterogeneity in the equity markets' reaction to sovereign default concerns. For the contemporaneous regression, Columns (1) and (2) of Table 7 show that much of the equity markets' response to sovereign default concern news is driven by countries with high fiscal constraints for which elevated default concerns are more likely to result in actual default, with potentially more economic costs.[21] For instance, the contemporaneous decline in equity market returns for a standard deviation increase in the NSRI percentage change is 9.1 bps higher for the fiscally constrained countries. At the same time, the reversal in equity returns over the next week is significantly lower for the fiscally constrained countries by roughly 10 bps. The results indicate that for countries experiencing high fiscal constraints, the influence of sovereign default risk news on equity returns is more persistent and more in line with a risk-based story. Conversely, for the rest of the countries, investors mainly overreact to sovereign default concerns.

	Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)
NSRI	-3.078	-1.651	12.023	11.863
	(-1.11)	(-0.68)	(2.61)	(2.61)
NSRI × High Fiscal Constr.	-10.005	-9.098	-8.586	-9.958
	(-2.48)	(-2.44)	(-1.74)	(-2.04)
High Fiscal Constr.	8.221	6.912	9.242	10.431
	(1.81)	(1.55)	(2.05)	(2.13)
Controls	NO	YES	NO	YES
R^2 (%)	11.74	36.73	11.72	13.41
Obs.	23,862	23,862	23,862	23,862

Table 7: Equity Returns, Sovereign Risk Concerns and Fiscal Constraint. The table shows results for the panel regression of countries' weekly stock market returns (Ret), in basis points, on the percentage change of the weekly-averaged NSRI relative to its past three-month median and its interaction with a dummy variable for high fiscal constraint (High Fiscal Constr.). The regressors are either contemporaneous (Columns 1 and 2) or one-period lagged (Columns and 3, 4). Ret(t-1) is the one-period lag of country j's weekly stock market return. The control variables included in Columns (2) and (4) are defined under Table 6. We normalize continuous regressors to unit variance and include country and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 - February 2021.

[21] Romer and Romer (2018), Battaglini and Coate (2016) and Yared (2019), among others, note that a high fiscal constraint can limit governments' responses to future catastrophes, such as financial crises, natural disasters, or wars, and can result is default through debt repudiation or inflation.

4.2.2 Global versus Local Sovereign Default Concerns and Equity Returns

Having seen how stock markets react to country-specific default risk concerns measured by NSRI, we now study the common component in NSRI and how equity markets react to global versus country-specific sovereign risk concerns. The motivation for this analysis is the extensive literature documenting a strong factor structure in countries' default risk dynamics, reflected in CDS spreads and its common drivers (e.g., Pan and Singleton, 2008; Longstaff et al., 2011; Augustin et al., 2014).

We start by examining whether, like CDS spreads, NSRI has a strong factor structure. Using weekly averaged NSRI and a panel of 74 countries with sufficient news coverage to compute the time-series of NSRI, we find that the first five principal components account for roughly 51% of the variation in the countries' NSRI.[22] This indicates that NSRI has a strong common component, corroborating the existing literature that observes a strong co-movement in countries' sovereign default risk. A possible explanation for this co-movement proposed by Benzoni et al. (2015) is as follows: a negative country-specific shock induces agents to revise their expectations about the default risk of other countries, resulting in greater credit spread correlations across countries compared to if spreads depended only on macroeconomic fundamentals. A similar intuition suggests that a surge in one country's default concern can prompt a higher discussion of other countries' default risks in the news media, resulting in the co-movement we observe in NSRI. Alternatively, comovement in countries' default concerns reflected in CDS spreads and NSRI could be jointly driven by global factors.

To analyze equity market's response to the average perception of default risk globally compared to country-specific default risk perception, we compute **G**lobal **N**ews-implied **S**overeign **D**efault **R**isk Index (GNSRI) as the cross-sectional value-weighted average of the weekly-averaged NSRI following the description in Footnote 15. We then augment Eq. (5) with $\Delta GNSRI_t$ computed in the same way as $\Delta NSRI_{j,t}$, i.e., the weekly percentage change in GNSRI relative to its three-month median. We exclude the US macroeconomic variables from the regression be-

[22]We use country-week observations with at least three news articles available to compute NSRI, which reduces noise in the data. Furthermore, we start the sample from 2007 as several countries had little coverage from PressReader in 2006. Finally, we drop countries with missing values for more than one-third of the sample period, then replace missing values with the cross-sectional median to obtain a complete time series for each country.

cause, as shown in Pan and Singleton (2008), the common component in sovereign default risk is strongly related to global macro-financial variables such as VIX. Hence, including such variables absorbs much of the information in GNSRI.[23]

Table 8 shows the estimation results. Focusing on Panel A, which shows results for the dollar-denominated stock market index return, Columns (1) and (2) show that the magnitude of the contemporaneous association between global sovereign default concerns and equity returns is more than twice the association between country-specific NSRI and equity returns. In particular, a standard deviation increase in the percentage change of the country-specific NSRI is contemporaneously associated with a decline in the stock market returns by roughly 10 bps, with t-statistic of 4.2. In comparison, for the global NSRI (GNSRI), the magnitude is 22 bps, with a much larger t-statistic of 7.9. We observe a similar result for the predictive regression (Columns (3) and (4)), where the regressors are one-week lagged. Here the initial declines in market returns following the increases in the NSRI and GNSRI percentage changes reverse in the following week, with the strength of the reversal again more pronounced by more than twice for GNSRI compared to NSRI. Precisely, a standard deviation increase in the percentage change of country-specific NSRI predicts an increase in market returns by 5.1 bps (with t-statistic of 1.80) over the next week compared to 16.7 bps (with t-statistic of 5.7) for the global NSRI.

For robustness, Panel B of Table 8 presents results using returns on local-currency denominated stock market indices. The estimates are quantitatively and qualitatively similar to those in Panel A, indicating that the choice of the stock market indices does not drive the results. Overall, the results once more support the view that equity markets around the world overreact to sovereign default risk concerns reflected in the news media. However, compared to country-specific concerns, the overreaction to global default concerns is even more pronounced. The results, therefore, underline the importance of global (or push) factors for local asset prices, in line with the existing literature (e.g., Colacito et al., 2018; Fraiberger et al., 2021).

[23] For example, we find that only the levels of VIX, EPU, ADS and the world stock market index explain roughly 40% of the time-series variation in GNSRI.

	Panel A: Dollar index return				Panel B: Local-currency index return			
	Contemporaneous		Lagged		Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NSRI	-7.093	-9.908	5.287	5.138	-3.770	-5.765	4.129	4.350
	(-3.07)	(-4.23)	(1.86)	(1.80)	(-1.90)	(-2.89)	(1.72)	(1.80)
GNSRI	-18.900	-21.554	19.640	16.683	-19.914	-21.741	18.551	15.976
	(-6.53)	(-7.93)	(6.45)	(5.71)	(-7.41)	(-8.24)	(6.21)	(5.66)
Ret($t-1$)		-41.845		-39.172		-33.777		-31.676
		(-6.22)		(-5.96)		(-5.36)		(-5.15)
Sentiment		-18.085		10.578		-13.709		9.520
		(-3.94)		(2.85)		(-3.25)		(2.70)
R^2 (%)	11.88	13.27	11.87	13.06	10.23	11.45	10.20	11.26
Obs.	24,101	24,101	24,101	24,101	24,004	24,004	24,004	24,004

Table 8: Impact of Local Vs. Global Sovereign Risk Concerns on Equity Returns. The table shows results for the panel regression of countries' weekly stock market returns (Ret), in basis points, on the percentage change of the weekly-averaged NSRI relative to its past three-month median and its global equivalent (GNSRI). GNSRI is computed as the time t weighted average of the country-specific NSRIs, with country weights based on the preceding year's dollar stock market capitalizations. The regressors are either contemporaneous (Columns 1, 2, 5, and 6) or one-period lagged (Columns 3, 4, 7 and 8). Ret(t-1) is the one period lag of country j's weekly stock market return. Sentiment is weekly average of country j's news sentiment. All right-hand-side variables are normalized to unit variance, and all regressions include country and year-month fixed effects. t-statistics based on standard errors clustered by country are in parentheses. Sample period is January 2006 - February 2021.

5 Additional Analysis and Robustness Tests

Equity Market Volatility and Sovereign Default Concerns. The theoretical model of Jeanneret (2017) suggests that an increase in sovereign default risk translates into an increase in equity market volatility. We examine if this prediction holds in our sample using our news-based measure of sovereign default risk (NSRI). Using the regression specification of Eq. (5) and weekly stock market realized volatility as the left-hand-side variable, we find that NSRI is positively and significantly associated with the same-week stock market volatility and predicts an increase in volatility over the next week. Table A10 presents the results, indicating that a standard deviation increase in the NSRI percentage change is associated with a contemporaneous increase in the realized stock market volatility by 0.08 – 0.10 pp. Furthermore, it predicts an increase in next week's volatility by 0.02 – 0.05 pp. Hence, sovereign default concerns reflected in NSRI is informative about both stock market returns and risk measured by realized volatility.

CDS Spreads and NSRI. We conduct additional tests to ensure that the observed informativeness of NSRI about CDS spread dynamics is robust. In the main analysis, we used country-week observations for which at least 60 news articles were available in a given calendar month to compute NSRI, which reduces noise. We verify that this choice does not influence our results by requiring a minimum of 90 news articles in a calendar month, i.e., an average of three articles per day. Table A3 shows the results of estimating Eq. (3) with the sample that meet the requirement. We find that requiring more news articles does not influence our results. In particular, the coefficient estimates for the NSRI percentage change for the weekly contemporaneous regressions, shown in Panel A of Table A3, range between 0.21 and 0.40, are significant and similar to those in the main analysis (Table 4). The same picture obtains for the one-week predictive regressions. Similarly, for the monthly contemporaneous (Columns (5) and (6)) and predictive (Columns (7) and (8)) regressions shown in Panel B of Table A3, the percentage change of NSRI is significantly and positively associated with CDS spreads percentage change (returns). A standard deviation increase in the NSRI percentage change is associated with an increase in CDS spread returns by 2.4% in the same month and predicts an increase of 1.1% in the next month.

We further check whether some specific countries drive our results. We re-estimate Eq. (3) dropping one country at a time from the panel regression. Table A4 collects the coefficient estimates for the NSRI percentage change for the weekly frequency panel regressions and Table A5 collects those for the monthly frequency. In both cases, we find that dropping any of the countries for which we have CDS spreads data and sufficient news coverage from the analysis does not materially affect the association between NSRI and CDS spreads. For instance, Column (2) of Table A4 shows that for the contemporaneous weekly regressions, the coefficient of the NSRI percentage change across all estimations is around 0.20 and significant in all cases. For the one-week predictive regressions, the coefficients are also significant in all cases, ranging between -0.14 and -0.19. These coefficients are quantitatively and qualitatively similar to those in the main analysis. We find a similar picture for the monthly panel regressions. In all estimations, the NSRI percentage change is positively associated with CDS spreads returns contemporaneously and predicts an increase in CDS spreads returns over the next month.

Rating Downgrades and NSRI. In the main analysis, we examined the informativeness of NSRI about future sovereign rating downgrades by predicting future downgrades on day t by NSRI averaged over the preceding 30 days ending $t-2$. We check whether NSRI remains informative when older information is used by computing the average over the preceding 60 days ending $t - 2$ and then re-estimate Eq. (4) using OLS. Table A6 shows the estimation results, indicating that averaging NSRI over a longer period does not wither away NSRI's informativeness about future sovereign rating downgrades. In particular, we find that a standard deviation increase in the average NSRI over the 60 days ending $t-2$ predicts an increase in the probability of a rating downgrade by between 9 – 12 pp. Again, we observe that CDS spread has no predictive power about future rating downgrade after conditioning on NSRI. This result underlines the incremental sovereign default information NSRI provides on top of CDS spread. Our main analysis and the initial robustness test relies on OLS estimations. We show that the results hold when we estimate Eq. (4) using conditional logistic regression. Table A7 collects the results for NSRI averaged over the past 30-day and 60-day periods ending $t - 2$. Across board, we find the NSRI significantly predicts sovereign rating downgrade even after controlling for CDS spread.

Equity Returns and NSRI. We conduct additional tests on the association between NSRI and stock market returns to ensure that our results in the main analysis are robust. First, we check that the observed equity market overreaction to news of sovereign default concerns is not driven by our requirement of at least 60 news articles per country in a calendar month to reduce noise in NSRI. We re-estimate Eq. (5) using country-week observations with at least 90 news articles in a calendar month. Furthermore, we check that the wild equity market fluctuations during the Covid-19 pandemic do not drive our results by stopping our sample in December 2019.

Table A8 shows the result of the two tests. Panel A shows that requiring more news per country in a calendar month to compute NSRI does not influence our main results. Panel B yields the same conclusion: our results are robust to excluding the Covid-19 period. For example, focusing on results for the dollar stock index returns (Columns (1) – (4) of Table A8), we find that, as in the main analysis, a standard deviation increase in the NSRI percentage change is contemporaneously associated with around 9 bps decline in the weekly stock market return, which significantly reverses over the next week. This result confirms our initial evidence that equity investors overreact to news of sovereign default concerns. Turning to the local-currency stock market index returns (Columns (5) – (8) of Table A8) we obtain a similar pattern: the

NSRI percentage change is significantly and negatively associated with same-week stock market returns and predicts an increase in stock market returns over the next week.

Next, we examine if our results are driven by specific outlier countries in the sample and find that this is not the case. Table A9 collects the coefficient estimates for the NSRI percentage change from re-estimating Eq. (5) using the dollar-denominated stock market index returns and dropping one country at a time from the panel regression. We find that for the contemporaneous regressions (Columns (1) and (2) of Table A9), the coefficient of the NSRI percentage change across all estimations is statistically significant and sizeable, ranging between -6 bps to -10 bps. Similarly, for the predictive regressions (Columns (3) and (4) of Table A9), we find a reversal of the initial decline in returns following an increase in NSRI. Precisely, across all estimations, the coefficient of the lagged NSRI percentage change is positive, statistically significant, and ranges between 5 bps and 9 bps. These results provide convincing evidence that NSRI is informative about equity market movements. In particular, equity investors tend to overreact to news of sovereign default concerns, resulting in a decline in returns in the same week news broke and a reversal over the following week.

6 Conclusion

This paper develops a novel framework for quantifying sovereign default risk in real-time using machine learning and natural language processing techniques. Our framework relies on two building blocks: topic modelling for dimensionality and noise reduction and text similarity for quantifying default concerns in the news media. Applying our framework to a corpus of 10 million news articles, we obtain the News-implied Sovereign Risk Index (NSRI) for a broad cross-section of countries. We conduct several validation exercises that showcase the intuitiveness and robustness of our methodology.

We analyze the informativeness of NSRI about sovereign CDS spread dynamics and credit rating downgrades. Our analyses provide convincing evidence that NSRI is strongly associated with CDS spreads contemporaneously and predicts future CDS spread dynamics and rating downgrades. Importantly, we find that NSRI contains additional sovereign default risk information not fully reflected in CDS spreads. For example, controlling for CDS spreads, NSRI significantly predicts rating downgrades. Conversely, CDS spreads have no incremental information on top of NSRI about future rating downgrades. These results imply that NSRI can complement CDS spreads in tracking sovereign default risk at a high frequency for countries with liquid CDS contracts. More importantly, for countries without liquid CDS contracts and other daily market-based proxies of default risk, NSRI can serve as a reliable stand-alone high-frequency proxy for sovereign default risk.

We apply NSRI to examine how equity markets respond to news of sovereign default concerns and find that NSRI is informative about equity returns. For example, a standard deviation increase in the NSRI percentage change is contemporaneously associated with a significant decline in the weekly stock market returns by 7.1 bps (3.7% annualized) and predicts an increase in returns by 6.4 bps (3.4% annualized) over the next week. The reversal in returns suggests that equity investors tend to overreact to news of sovereign default concerns. Furthermore, we find that compared to country-specific default concerns, global default concerns have a much stronger influence on stock market returns, underlining the relevance of global "push" factors for local asset prices.

There are plausible extensions of our framework which may yield even stronger sovereign default risk signals from news text: (i) supervised machine learning methods that use news text and an alternative high-frequency indicator of default risk for model training; (ii) use of language translation algorithms to avoid reliance on only news in the English language, thereby providing even more signals for many countries; (iii) use of alternative news API providers to achieve more local news coverage for certain countries. We leave these extensions for future research.

In sum, textual data from the news media contains valuable information about countries' default risks useful for monitoring default concerns in real-time. The NSRI can serve as an input to credit risk early warning systems, credit portfolio management, and can be applied to test theories on the economic costs and financial market impacts of sovereign default risk.

References

Adelino, M. and Ferreira, M. A. (2016). Bank ratings and lending supply: Evidence from sovereign downgrades. *The Review of Financial Studies*, 29(7):1709–1746.

Almeida, H., Cunha, I., Ferreira, M. A., and Restrepo, F. (2017). The real effects of credit ratings: The sovereign ceiling channel. *The Journal of Finance*, 72(1):249–290.

Andrade, S. C. and Chhaochharia, V. (2018). The costs of sovereign default: Evidence from the stock market. *The Review of Financial Studies*, 31(5):1707–1751.

Ang, A. and Longstaff, F. A. (2013). Systemic sovereign credit risk: Lessons from the US and Europe. *Journal of Monetary Economics*, 60(5):493–510.

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427.

Augustin, P. (2014). Sovereign credit default swap premia. *Journal of Investment Management.*

Augustin, P., Sokolovski, V., Subrahmanyam, M. G., and Tomio, D. (2021). In sickness and in debt: The covid-19 impact on sovereign credit risk. *Journal of Financial Economics.*

Augustin, P., Subrahmanyam, M. G., Tang, D. Y., and Wang, S. Q. (2014). Credit default swaps–a survey. *Foundations and Trends® in Finance*, 9(1-2):1–196.

Augustin, P. and Tédongap, R. (2016). Real economic shocks and sovereign credit risk. *Journal of Financial and Quantitative Analysis*, 51(2):541–587.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Battaglini, M. and Coate, S. (2016). A political economy theory of fiscal policy and unemployment. *Journal of the European Economic Association*, 14(2):303–337.

Bellstam, G., Bhagat, S., and Cookson, J. A. (2021). A text-based analysis of corporate innovation. *Management Science*, 67(7):4004–4031.

Benzoni, L., Collin-Dufresne, P., Goldstein, R. S., and Helwege, J. (2015). Modeling credit contagion via the updating of fragile beliefs. *The Review of Financial Studies*, 28(7):1960–2008.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Bocola, L. (2016). The pass-through of sovereign risk. *Journal of Political Economy*, 124(4):879–926.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic expectations and stock returns. *The Journal of Finance*, 74(6):2839–2874.

Brusa, F., Savor, P., and Wilson, M. (2020). One central bank to rule them all. *Review of Finance*, 24(2):263–304.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2021). Business news and business cycles. *NBER Working Paper w29344.*

Chernov, M., Creal, D. D., and Hördahl, P. (2020a). Sovereign credit and exchange rate risks: evidence from Asia-Pacific local currency bonds. Technical report, National Bureau of Economic Research.

Chernov, M., Schmid, L., and Schneider, A. (2020b). A macrofinance view of us sovereign CDS premiums. *The Journal of Finance*, 75(5):2809–2844.

Colacito, R., Croce, M. M., Gavazzoni, F., and Ready, R. (2018). Currency risk factors in a recursive multicountry economy. *The Journal of Finance*, 73(6):2719–2756.

Cole, H. L. and Kehoe, P. J. (1998). Models of sovereign debt: Partial versus general reputations. *International Economic Review*, pages 55–70.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3):793–805.

Dremen, D. (1983). The new contrarian investment strategy. the psychology of stock market success.

Edmans, A., Fernandez-Perez, A., Garel, A., and Indriawan, I. (2021). Music sentiment and stock returns around the world. *CEPR Discussion Paper No. DP15756.*

Eichengreen, B., El-Ganainy, A., Esteves, R., and Mitchener, K. J. (2021). *In Defense of Public Debt*. Oxford University Press.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.

Fraiberger, S. P., Lee, D., Puy, D., and Ranciere, R. (2021). Media sentiment and international asset prices. *Journal of International Economics*, 133:103526.

Gao, Z., Ren, H., and Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, 55(2):549–580.

Gennaioli, N., Martin, A., and Rossi, S. (2014). Sovereign default, domestic banks, and financial institutions. *The Journal of Finance*, 69(2):819–866.

Gennaioli, N., Shleifer, A., and Vishny, R. (2015). Neglected risks: The psychology of financial crises. *American Economic Review*, 105(5):310–14.

Hanley, K. W. and Hoberg, G. (2010). The information content of ipo prospectuses. *The Review of Financial Studies*, 23(7):2821–2864.

Hanley, K. W. and Hoberg, G. (2019). Dynamic interpretation of emerging risks in the financial sector. *The Review of Financial Studies*, 32(12):4543–4603.

Hassan, T. A., Hollander, S., Van Lent, L., and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4):2135–2202.

Hilscher, J. and Nosbusch, Y. (2010). Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt. *Review of Finance*, 14(2):235–262.

Huang, A. H., Lehavy, R., Zang, A. Y., and Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64(6):2833–2855.

IIF (2021a). Chipping away at the mountain? *Global Debt Monitor.*

IIF (2021b). Covid drives debt surge—stabilization ahead? *Global Debt Monitor.*

IMF (2021). Fiscal monitor - October 2021, chapter 1. *Fiscal Monitor Reports.*

Jeanneret, A. (2017). Sovereign default risk and the us equity market. *Journal of Financial and Quantitative Analysis*, 52(1):305–339.

Kose, M. A., Kurlat, S., Ohnsorge, F., and Sugawara, N. (2017). A cross-country database of fiscal space. *World Bank Policy Research Working Paper 8157, World Bank, Washington, DC.*

Lee, J., Naranjo, A., and Sirmans, S. (2016). Exodus from sovereign risk: Global asset and information networks in the pricing of corporate credit risk. *The Journal of Finance*, 71(4):1813–1856.

Liu, Y. and Matthies, B. (2021). Long run risk: Is it there? *Journal of Finance, Forthcoming*.

Longstaff, F. A., Pan, J., Pedersen, L. H., and Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.

Manela, A. and Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1):137–162.

Mitchener, K. J. and Trebesch, C. (2021). Sovereign debt in the 21st century: Looking backward, looking forward. Technical report, National Bureau of Economic Research.

Pan, J. and Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign cds spreads. *The Journal of Finance*, 63(5):2345–2384.

PressReader (2021). About us. https://about.pressreader.com/about-us/. Accessed: 2021-10-01.

Remolona, E. M., Scatigna, M., and Wu, E. (2008). The dynamic pricing of sovereign risk in emerging markets: Fundamentals and risk aversion. *The Journal of Fixed Income*, 17(4):57–71.

Robinson, M. J. (2007). Two decades of american news preferences. *Pew Research Center*.

Romer, C. D. and Romer, D. H. (2018). Phillips lecture–why some times are different: Macroeconomic policy and the aftermath of financial crises. *Economica*, 85(337):1–40.

Sautner, Z., van Lent, L., Vilkov, G., and Zhang, R. (2020). Firm-level climate change exposure. *TRR 266 Accounting for Transparency Working Paper Series No. 33*.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4):967–1004.

Sievert, C. and Shirley, K. (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.

Trussler, M. and Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, 19(3):360–379.

Yared, P. (2019). Rising government debt: Causes and solutions for a decades-old trend. *Journal of Economic Perspectives*, 33(2):115–40.

A Appendix

A.1 Text Preprocessing

The goal of text preprocessing is to reduce noise and dimensionality in text data to improve signal extraction. We use the following steps:

1. Screen retrieved news articles for country relevance: We use regular expression to check that each news article retrieved for a country sufficiently relates to that country. To do so, we search for the number of times variants of a country's name is mentioned in a news article and its first ten sentences. Then, we keep only articles that mention the country's name at least once in the first ten sentences or at least twice in the entire article.

2. Remove entity names: To further reduce noise and improve signal extraction, we use Named Entity Recognition (NER) algorithms to identify and remove country names, names of persons, geographical locations, political entities, and places from the news corpus.

3. Remove stopwords: We filter out common English language stopwords using standard stopword lists found in NLP libraries.

4. Remove non-English words: We filter out non-English words using the English language word list from the NLTK python library. We retain common economics and financial acronyms such as "gdp", "imf", "cds", "fx", "cpi".

5. Normalize words: Finally, we lemmatize words in our news corpus to reduce different forms of a word to their root word while avoiding ambiguities that could arise from word stemming.

A.2 Additional Tables and Figures

Name	# Articles	Country	# Articles	Country	# Articles	Country	# Articles
Afghanistan	162,041	Dominican Rep.	7,786	Libya	66,892	Samoa	2,444
Albania	4,238	DR Congo	39,595	Liechtenstein	1,733	Sao Tome & Principe	159
Algeria	9,686	Ecuador	10,950	Lithuania	7,169	Saudi Arabia	90,194
Angola	5,087	Egypt	78,413	Luxembourg	11,910	Senegal	5,405
Argentina	30,441	El Salvador	13,153	Macao	10,515	Serbia	11,591
Armenia	3,007	Equatorial Guinea	1,214	Madagascar	2,324	Seychelles	1,756
Aruba	999	Eritrea	4,187	Malawi	3,919	Sierra Leone	8,107
Australia	150,633	Estonia	6,720	Malaysia	195,572	Singapore	117,999
Austria	6,426	Eswatini	69	Maldives	2,794	Slovakia	6,969
Azerbaijan	9,618	Ethiopia	15,588	Mali	9,082	Slovenia	5,637
Bahamas	7,542	Fiji	20,589	Mauritania	1,841	Somalia	25,961
Bahrain	15,634	Finland	5,261	Mauritius	2,260	South Africa	382,812
Bangladesh	17,901	France	85,164	Mexico	226,711	South Korea	98,657
Barbados	3,269	Gabon	1,373	Micronesia	486	South Sudan	12,129
Belarus	8,560	Gambia	1,717	Moldova	2,306	Spain	117,483
Belgium	32,098	Georgia	114,482	Monaco	6,397	Sri Lanka	71,847
Belize	2,184	Germany	95,886	Mongolia	12,483	Sudan	27,914
Benin	1,372	Ghana	7,397	Montenegro	2,867	Suriname	680
Bermuda	4,629	Greece	95,438	Morocco	10,920	Sweden	10,724
Bhutan	4,540	Grenada	1,693	Mozambique	4,234	Switzerland	14,056
Bolivia	8,514	Guatemala	15,897	Myanmar	30,557	Syria	173,367
Botswana	5,538	Guinea	14,443	Namibia	5,357	Taiwan, Province of China	32,715
Brazil	73,248	Guinea-Bissau	500	Nepal	9,783	Tajikistan	2,434
Brunei Darussalam	224	Guyana	36,033	Netherlands	11,402	Tanzania	8,920
Bulgaria	9,968	Haiti	17,134	New Caledonia	1,025	Thailand	101,392
Burkina Faso	2,738	Honduras	14,812	New Zealand	114,858	Timor-Leste	145
Burundi	2,564	Hong Kong	109,717	Nicaragua	9,274	Togo	1,775
Cabo Verde	105	Hungary	19,038	Niger	6,602	Tonga	1,190
Cambodia	26,669	Iceland	17,630	Nigeria	250,843	Trinidad and Tobago	2,335
Cameroon	6,091	India	1,148,893	North Korea	70,456	Tunisia	18,235
Canada	527,203	Indonesia	81,286	Norway	10,373	Turkey	113,541
Central African Rep.	4,850	Iran (Islamic Rep.)	183,588	Oman	41,298	Turkmenistan	1,599
Chad	18,404	Iraq	232,631	Palestine	15,513	Tuvalu	463
Chile	20,689	Ireland	246,719	Panama	15,760	UAE	87,892
China	577,578	Israel	201,846	Papua New Guinea	2,761	Uganda	11,199
Colombia	25,791	Italy	124,875	Paraguay	3,103	Ukraine	77,985
Comoros	370	Jamaica	42,792	Peru	14,994	United Kingdom	347,792
Costa Rica	6,181	Japan	85,857	Philippines	416,841	Uruguay	6,156
Cote d'Ivoire	189	Kazakhstan	9,568	Poland	39,310	USA	347,792
Croatia	8,153	Kenya	36,767	Portugal	31,436	Uzbekistan	4,522
Cuba	54,944	Kyrgyzstan	3,829	Puerto Rico	22,524	Venezuela	43,693
Cyprus	18,135	Laos	6,465	Qatar	41,747	Vietnam	63,682
Czechia	37	Latvia	6,371	Rep. of Congo	6,980	Western Sahara	1,215
Denmark	8,336	Lebanon	53,035	Romania	13,181	Yemen	43,505
Djibouti	2,538	Lesotho	4,169	Russia	258,694	Zambia	12,769
Dominica	1,101	Liberia	7,937	Rwanda	9,192	Zimbabwe	67,750

Table A1: News API Country Coverage. The table shows the countries, dependent territories, and special areas of geographical interest for which we have news articles from the PressReader News API, as well as the number of news articles we obtained. To save space we use abbreviated names in some cases instead of the official names.

Name	MSCI Index (USD)	MSCI Index (local)	Country	MSCI Index (USD)	MSCI Index (local)
Argentina	MSARGT$	MSARGTL	Malta	MSIMLT$	MSIMLTL
Australia	MSAUST$	MSAUSTL	Mauritius	MSMAUR$	MSMAURL
Austria	MSASTR$	MSASTRL	Mexico	MSMEXF$	MSMEXFL
Bahrain	MSBAHR$	MSBAHRL	Morocco	MSMORC$	MSMORCL
Bangladesh	MSBNGS$	MSBNGSL	Namibia	IFFMNA$	IFFMNAL
Belgium	MSBELG$	MSBELGL	Netherlands	MSNETH$	MSNETHL
Botswana	MSBTSW$	MSBTSWL	New Zealand	MSNZEA$	MSNZEAL
Brazil	MSBRAZ$	MSBRAZL	Nigeria	MSNGRA$	MSNGRAL
Bulgaria	MSBLGN$	MSBLGNL	Norway	MSNWAY$	MSNWAYL
Canada	MSCNDA$	MSCNDAL	Oman	MSOMAN$	MSOMANL
Chile	MSCHIL$	MSCHILL	Pakistan	MSPAKI$	MSPAKIL
China	MSCHIN$	MSCHINL	Palestine	MSIPLE$	MSIPLEL
Colombia	MSCOLM$	MSCOLML	Panama	IFFPNM$	IFFMPAL
Cote d'Ivoire	IFFMCI$	IFFMCIL	Peru	MSPERU$	MSPERU$
Croatia	MSCROA$	MSCROAL	Philippines	MSPHLF$	MSPHLFL
Czechia	MSCZCH$	MSCZCHL	Poland	MSPLND$	MSPLNDL
Denmark	MSDNMK$	MSDNMKL	Portugal	MSPORD$	MSPORDL
Ecuador	IFFMEC$	IFFMECL	Qatar	MSQATA$	MSQATAL
Egypt	MSEGYT$	MSEGYTL	Romania	MSROMN$	MSROMNL
Estonia	MSESTN$	MSESTNL	Russia	MSRUSS$	MSRUSSL
Finland	MSFIND$	MSFINDL	Saudi Arabia	MSSAUD$	MSSAUDL
France	MSFRNC$	MSFRNCL	Serbia	MSSERB$	MSSERBL
Georgia	IFGEGR$	IFGEGR$	Singapore	MSSING$	MSSINGL
Germany	MSGERM$	MSGERML	Slovenia	MSSLVN$	MSSLVNL
Ghana	MSGHAN$	MSGHANL	South Africa	MSSARF$	MSSARFL
Greece	MSGREE$	MSGREEL	South Korea	MSKORE$	MSKOREL
Hong Kong	MSHGKG$	MSHGKGL	Spain	MSSPAN$	MSSPANL
Hungary	MSHUNG$	MSHUNGL	Sri Lanka	MSSRIL$	MSSRILL
Iceland	ICEXALL	ICEXALL	Sweden	MSSWDN$	MSSWDNL
India	MSINDI$	MSINDIL	Switzerland	MSSWIT$	MSSWITL
Indonesia	MSINDF$	MSINDFL	Taiwan, Province of China	MSTAIW$	MSTAIWL
Ireland	MSEIRE$	MSEIREL	Thailand	MSTHAF$	MSTHAFL
Israel	MSISRL$	MSISRLL	Trinidad and Tobago	MSTRTG$	MSTRTGL
Italy	MSITAL$	MSITALL	Tunisia	MSTNSA$	MSTNSAL
Jamaica	MSJMCA$	MSJMCAL	Turkey	MSTURK$	MSTURKL
Japan	MSJPAN$	MSJPANL	Uganda	ALSIUGI	ALSIUGI
Jordan	MSJORD$	MSJORDL	Ukraine	MSUKRN$	MSUKRNL
Kazakhstan	MSKZKT$	MSKZKTL	UAE	MSUAEI$	MSUAEIL
Kenya	MSKNYA$	MSKNYAL	United Kingdom	MSUTDK$	MSUTDKL
Kuwait	MSKUWA$	MSKUWAL	USA	MSUSAM$	MSUSAML
Laos	LAOLSXI	LAOLSXI	Vietnam	MSVIET$	MSVIETL
Latvia	RIGSEIN	RIGSEIN	Zambia	IFFMZB$	IFFMZBL
Lithuania	MSLITH$	MSLITHL	Zimbabwe	MSZMBW$	MSZMBWL
Luxembourg	SBBLUX$	SBBLUXL	World	MSWRLD$	MSWRLD$
Malaysia	MSMALF$	MSMALFL			

Table A2: Stock Market Indices per Country. The table shows the MSCI dollar-denominated and local-currency denominated stock market indices used per country, dependent territory, or special areas of geographical interest. To save space we use abbreviated names in some cases instead of the official names.

	Panel A: Weekly frequency				Panel B: Monthly frequency			
	Contemporaneous		Lagged		Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NSRI	0.397	0.210	-0.124	-0.165	4.647	2.286	1.667	1.134
	(4.54)	(3.20)	(-1.71)	(-2.55)	(9.27)	(7.52)	(3.71)	(3.13)
CDS Spread ($t-1$)		1.226		0.699		0.999		-0.854
		(11.33)		(3.79)		(2.54)		(-1.52)
Sentiment		0.402		-0.050		1.228		0.744
		(4.20)		(-0.74)		(1.80)		(1.55)
Ret		-1.689		-0.832		-2.442		-6.119
		(-10.74)		(-5.43)		(-7.38)		(-10.99)
Vol		0.727		-0.102		5.507		-1.569
		(4.57)		(-0.99)		(6.06)		(-2.46)
World Ret		-0.082		-0.903		-4.708		2.186
		(-0.56)		(-4.36)		(-10.05)		(4.56)
EPU		-0.010		-0.092		2.297		0.262
		(-0.16)		(-1.21)		(7.85)		(0.54)
ADS		0.111		0.086		0.140		2.267
		(2.79)		(2.39)		(0.90)		(6.77)
VIX		1.517		-0.843		0.858		2.508
		(12.17)		(-12.53)		(1.83)		(4.82)
R^2 (%)	23.39	36.60	23.15	27.04	13.90	47.32	8.43	19.09
Obs.	12,673	12,673	12,673	12,673	2,838	2,838	2,838	2,838

Table A3: Panel Regression of CDS Spreads on NSRI Filtering for Min 90 News Articles. The table shows results for the panel regression of percentage changes in countries' CDS Spreads (in %) on percentage changes in NSRI relative to its three-month median. We drop observations with less than 90 news articles for a country in a calendar month. The analysis uses only observations for which Time is either at the weekly (Panel A) and monthly (Panel B) frequency. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. CDS Spread ($t-1$) is the one-period lag of country j's CDS Spread (the dependent variable). Sentiment is the weekly average of country j's news sentiment. Ret and Vol are country j's weekly stock market return and realized volatility, respectively. World Ret is the weekly return on the MSCI World stock market index. EPU is the weekly percentage change of the US economic policy uncertainty index. ADS is the weekly percentage change of US economic activity index, and VIX is the weekly percentage change of the implied volatility index. We normalize continuous regressors to unit variance and include country and time fixed effects (year-month fixed effects in Panel A and year fixed effects in Panel B) in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021.

	Contemporaneous				Lagged			
	(1)		(2)		(3)		(4)	
Excluded country	Without controls		With controls		Without controls		With controls	
Brazil	0.419	(6.10)	0.243	(4.46)	-0.103	(-1.52)	-0.135	(-2.08)
Chile	0.392	(5.41)	0.219	(3.68)	-0.122	(-1.77)	-0.148	(-2.28)
China	0.404	(5.50)	0.230	(3.86)	-0.143	(-2.08)	-0.172	(-2.67)
Colombia	0.394	(5.29)	0.217	(3.57)	-0.121	(-1.72)	-0.141	(-2.19)
Croatia	0.391	(5.39)	0.216	(3.62)	-0.126	(-1.82)	-0.152	(-2.33)
Denmark	0.391	(5.39)	0.216	(3.62)	-0.126	(-1.82)	-0.152	(-2.33)
Egypt	0.412	(5.69)	0.219	(3.52)	-0.125	(-1.71)	-0.149	(-2.20)
France	0.377	(5.02)	0.212	(3.41)	-0.120	(-1.63)	-0.143	(-2.08)
Germany	0.365	(5.09)	0.198	(3.29)	-0.105	(-1.50)	-0.132	(-1.99)
Greece	0.390	(4.99)	0.200	(3.18)	-0.148	(-2.03)	-0.171	(-2.44)
Hungary	0.391	(5.36)	0.215	(3.59)	-0.128	(-1.84)	-0.153	(-2.34)
Indonesia	0.390	(5.26)	0.216	(3.54)	-0.125	(-1.76)	-0.150	(-2.26)
Israel	0.394	(5.33)	0.214	(3.48)	-0.136	(-1.92)	-0.161	(-2.41)
Italy	0.360	(4.52)	0.200	(3.00)	-0.161	(-2.31)	-0.185	(-2.78)
Japan	0.403	(5.35)	0.224	(3.56)	-0.114	(-1.66)	-0.139	(-2.15)
Malaysia	0.398	(5.37)	0.224	(3.69)	-0.125	(-1.76)	-0.149	(-2.25)
Mexico	0.406	(5.44)	0.228	(3.76)	-0.103	(-1.57)	-0.134	(-2.14)
Netherlands	0.394	(5.43)	0.218	(3.66)	-0.123	(-1.78)	-0.150	(-2.29)
Panama	0.390	(5.38)	0.215	(3.61)	-0.127	(-1.83)	-0.153	(-2.34)
Peru	0.393	(5.41)	0.216	(3.61)	-0.126	(-1.81)	-0.152	(-2.32)
Philippines	0.397	(5.37)	0.221	(3.64)	-0.135	(-1.91)	-0.160	(-2.41)
Poland	0.393	(5.23)	0.218	(3.53)	-0.125	(-1.71)	-0.158	(-2.31)
Portugal	0.384	(5.23)	0.209	(3.46)	-0.134	(-1.90)	-0.157	(-2.35)
Romania	0.393	(5.40)	0.216	(3.62)	-0.127	(-1.82)	-0.153	(-2.34)
Russia	0.359	(5.28)	0.186	(3.41)	-0.124	(-1.75)	-0.154	(-2.31)
Saudi Arabia	0.406	(5.60)	0.224	(3.71)	-0.114	(-1.66)	-0.139	(-2.18)
Slovakia	0.391	(5.39)	0.216	(3.62)	-0.126	(-1.82)	-0.152	(-2.33)
South Africa	0.393	(5.26)	0.216	(3.52)	-0.130	(-1.84)	-0.155	(-2.33)
Spain	0.378	(4.81)	0.213	(3.29)	-0.158	(-2.24)	-0.181	(-2.70)
Sweden	0.391	(5.39)	0.216	(3.62)	-0.126	(-1.82)	-0.152	(-2.33)
Thailand	0.422	(5.96)	0.239	(4.13)	-0.114	(-1.66)	-0.145	(-2.23)
Turkey	0.360	(4.85)	0.190	(3.13)	-0.137	(-1.82)	-0.161	(-2.29)
United Kingdom	0.386	(5.19)	0.205	(3.38)	-0.130	(-1.80)	-0.157	(-2.30)
USA	0.387	(5.24)	0.227	(3.71)	-0.123	(-1.76)	-0.142	(-2.20)
Venezuela	0.391	(5.39)	0.216	(3.62)	-0.127	(-1.83)	-0.153	(-2.34)

Table A4: Panel Regression of CDS Spreads on NSRI Dropping One Country at a Time I. The table shows results for the panel regression of weekly percentage changes in countries' CDS Spreads (in %) on the weekly percentage changes in NSRI relative to its three-month median ($\Delta NSRI_{j,t}$) based on Eq. (3). We drop one country at a time from the regression and report only the resulting coefficient of $\Delta NSRI_{j,t}$ from each regression. The column "Excluded country" indicates the country dropped from the analysis. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. Columns (2) and (4) include the control variables described under Table 4. We normalize continuous regressors to unit variance and include country and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021 and includes observations with at least 60 news articles in a calendar month to compute NSRI.

	Contemporaneous				Lagged			
	(1)		(2)		(3)		(4)	
Excluded country	Without controls		With controls		Without controls		With controls	
Brazil	4.702	(8.32)	2.363	(8.00)	1.782	(4.18)	1.194	(3.65)
Chile	4.773	(8.97)	2.392	(8.65)	1.657	(3.86)	1.090	(3.28)
China	4.758	(8.81)	2.488	(8.76)	1.583	(3.62)	1.045	(3.10)
Colombia	4.805	(8.94)	2.408	(8.65)	1.726	(4.07)	1.120	(3.35)
Croatia	4.772	(8.97)	2.392	(8.66)	1.656	(3.86)	1.089	(3.28)
Egypt	4.942	(9.10)	2.448	(8.51)	1.710	(3.88)	1.098	(3.22)
France	4.544	(9.08)	2.285	(8.39)	1.478	(3.42)	0.977	(2.82)
Germany	4.551	(9.01)	2.246	(8.82)	1.528	(3.43)	0.998	(2.84)
Greece	4.910	(9.08)	2.258	(8.18)	1.675	(3.64)	1.022	(2.87)
Hungary	4.780	(8.97)	2.392	(8.64)	1.657	(3.86)	1.088	(3.28)
Indonesia	4.709	(8.91)	2.414	(8.57)	1.664	(3.79)	1.115	(3.29)
Israel	4.887	(8.93)	2.408	(8.49)	1.699	(3.88)	1.103	(3.26)
Italy	4.669	(8.20)	2.384	(8.66)	1.363	(3.04)	0.892	(2.49)
Japan	4.736	(8.70)	2.437	(8.39)	1.665	(3.77)	1.119	(3.25)
Malaysia	4.795	(8.83)	2.436	(8.66)	1.706	(3.88)	1.141	(3.39)
Mexico	4.782	(8.77)	2.424	(8.62)	1.741	(3.99)	1.137	(3.30)
Netherlands	4.772	(8.97)	2.391	(8.65)	1.656	(3.86)	1.090	(3.28)
Panama	4.773	(8.97)	2.390	(8.63)	1.656	(3.86)	1.089	(3.28)
Peru	4.774	(8.97)	2.393	(8.65)	1.652	(3.86)	1.086	(3.28)
Philippines	4.840	(8.87)	2.468	(8.83)	1.655	(3.77)	1.036	(3.11)
Poland	4.869	(8.96)	2.444	(8.64)	1.641	(3.70)	1.051	(3.09)
Portugal	4.772	(8.90)	2.346	(8.49)	1.654	(3.79)	1.087	(3.22)
Romania	4.772	(8.97)	2.393	(8.66)	1.657	(3.86)	1.090	(3.29)
Russia	4.687	(8.66)	2.372	(8.40)	1.739	(3.96)	1.165	(3.46)
Saudi Arabia	4.672	(8.85)	2.341	(8.44)	1.627	(3.70)	1.066	(3.15)
South Africa	4.832	(8.79)	2.432	(8.49)	1.688	(3.82)	1.124	(3.30)
Spain	4.602	(7.91)	2.255	(7.42)	1.484	(3.20)	0.969	(2.69)
Thailand	4.901	(8.88)	2.519	(9.31)	1.713	(3.90)	1.130	(3.30)
Turkey	4.948	(9.42)	2.341	(7.94)	1.875	(5.15)	1.269	(4.65)
United Kingdom	4.837	(8.83)	2.448	(8.59)	1.644	(3.72)	1.123	(3.28)
USA	4.819	(8.70)	2.439	(8.39)	1.699	(3.85)	1.181	(3.60)
Venezuela	4.777	(8.98)	2.392	(8.62)	1.658	(3.85)	1.091	(3.28)

Table A5: Panel Regression of CDS Spreads on NSRI Dropping One Country at a Time II. The table shows results for the panel regression of monthly percentage changes in countries' CDS Spreads (in %) on the monthly percentage changes in NSRI relative to its three-month median ($\Delta NSRI_{j,t}$) based on Eq. (3). We drop one country at a time from the regression and report only the resulting coefficient of $\Delta NSRI_{j,t}$ from each regression. The column "Excluded country" indicates the country dropped from the analysis. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. Columns (2) and (4) include the control variables described under Table 4. We normalize continuous regressors to unit variance and include country and year fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021 and includes observations with at least 60 news articles in a calendar month to compute NSRI.

	(1)	(2)	(3)	(4)	(5)	(6)
NSRI	0.119	0.120	0.095	0.095	0.089	0.089
	(6.11)	(6.17)	(4.39)	(4.38)	(4.37)	(3.83)
Sentiment		0.054	0.049	0.048	0.038	0.095
		(2.94)	(1.58)	(1.57)	(1.27)	(2.96)
Ret			-0.090	-0.092	-0.092	-0.122
			(-6.28)	(-4.03)	(-4.22)	(-4.87)
Vol			0.115	0.116	0.088	0.098
			(6.26)	(6.00)	(4.28)	(3.30)
World Ret				0.002	0.019	0.067
				(0.07)	(0.76)	(3.14)
EPU					0.038	0.044
					(1.94)	(1.66)
ADS					-0.009	-0.006
					(-0.54)	(-0.31)
VIX					0.025	0.022
					(0.87)	(0.52)
CDS Spread						-0.001
						(-0.08)
R^2 (%)	2606	2606	1721	1721	1721	1076
Obs.	0.165	0.172	0.294	0.294	0.303	0.351

Table A6: Panel Regression of Sovereign Rating Downgrade on NSRI. The table shows results for the panel OLS regression of an indicator variable of rating downgrade ($Downgrade_{j,t}$) on the NSRI averaged over the past 60-day window ending $t-2$ before rating change on day t. $Downgrade_{j,t}$ equals equals 1 if a country's rating change on t is a downgrade and 0 otherwise. The other regressors are computed over the same 30-day window ending $t-2$ as NSRI. Sentiment is the average of country j's news sentiment. Ret is country j's cumulative stock market return. Vol is country j's stock market volatility. World Ret is the cumulative return of the MSCI World stock market index. EPU is the average US economic policy uncertainty index. ADS is the average US economic activity index. VIX is the average implied volatility index. CDS Spread is the average of country j's CDS Spread. We normalize continuous regressors to unit variance, and all regressions include country fixed-effects. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 - February 2021.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: NSRI Averaged over the past 30-days ending $t-2$						
NSRI	0.105	0.086	0.036	0.036	0.024	0.012
	(11.03)	(10.76)	(4.52)	(4.45)	(3.32)	(2.64)
Sentiment		0.036	0.015	0.015	0.009	0.014
		(3.92)	(1.74)	(1.73)	(1.21)	(2.38)
Ret			-0.013	-0.014	-0.012	-0.009
			(-1.94)	(-1.58)	(-1.79)	(-1.40)
Vol			0.083	0.083	0.054	0.035
			(4.24)	(4.23)	(2.89)	(1.84)
World Ret				0.002	0.007	0.010
				(0.15)	(0.95)	(1.91)
EPU					0.017	0.013
					(2.03)	(1.81)
ADS					-0.009	-0.018
					(-0.65)	(-1.56)
VIX					0.014	0.004
					(1.48)	(0.50)
CDS Spread						-0.001
						(-0.21)
Obs.	2582	2582	1696	1696	1696	1059
Panel B: NSRI Averaged over the past 60-days ending $t-2$						
NSRI	0.106	0.080	0.038	0.039	0.031	0.012
	(15.90)	(8.95)	(4.30)	(4.30)	(3.41)	(2.12)
Sentiment		0.032	0.016	0.016	0.012	0.016
		(4.09)	(1.68)	(1.67)	(1.27)	(2.28)
Ret			-0.039	-0.038	-0.035	-0.023
			(-2.57)	(-2.32)	(-2.25)	(-1.56)
Vol			0.063	0.063	0.049	0.034
			(3.10)	(3.09)	(2.23)	(1.54)
World Ret				-0.003	0.003	0.009
				(-0.30)	(0.29)	(1.45)
EPU					0.024	0.016
					(1.95)	(1.52)
ADS					-0.021	-0.014
					(-1.36)	(-1.40)
VIX					-0.003	-0.004
					(-0.21)	(-0.32)
CDS Spread						-0.001
						(-0.93)
Obs.	2582	2582	1701	1701	1701	1061

Table A7: Panel Logistic Regression of Sovereign Rating Downgrade on NSRI. The table shows results for the panel conditional logistic regression of an indicator variable of rating downgrade ($Downgrade_{j,t}$) on the NSRI averaged over the past 30-day (Panel A) and 60-day (Panel B) windows ending $t-2$ before rating change on day t. $Downgrade_{j,t}$ equals equals 1 if a country's rating change on t is a downgrade and 0 otherwise. The other regressors are computed over the same 30-day window ending $t-2$ as NSRI. Sentiment is the average of country j's news sentiment. Ret is country j's cumulative stock market return. Vol is country j's stock market volatility. World Ret is the cumulative return of the MSCI World stock market index. EPU is the average US economic policy uncertainty index. ADS is the average US economic activity index. VIX is the average implied volatility index. CDS Spread is the average of country j's CDS Spread. We normalize continuous regressors to unit variance, and all regressions include country fixed-effects. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 - February 2021.

	Dollar return				Local-currency return			
	Contemporaneous		Lagged		Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Excluding months with less than 90 news articles per country								
NSRI	-9.935	-8.608	9.607	8.382	-6.862	-5.339	7.876	6.938
	(-4.07)	(-3.50)	(4.54)	(4.00)	(-3.27)	(-2.56)	(4.25)	(3.68)
Ret($t-1$)		-15.411		-26.066		-13.954		-20.353
		(-2.97)		(-3.84)		(-2.71)		(-3.37)
Sentiment		-14.508		10.245		-10.418		9.507
		(-3.61)		(2.69)		(-2.74)		(2.66)
World Ret		179.083		-22.097		137.359		-19.386
		(14.36)		(-3.44)		(13.06)		(-3.60)
EPU		-4.861		-0.698		-1.649		1.865
		(-3.67)		(-0.40)		(-1.27)		(1.46)
ADS		-8.667		7.256		-1.950		5.576
		(-4.52)		(4.13)		(-1.33)		(3.74)
VIX		-13.945		12.276		-24.024		11.111
		(-3.10)		(3.52)		(-5.91)		(4.36)
R^2 (%)	11.97	37.00	11.96	13.77	10.19	32.38	10.21	11.88
Obs.	21,767	21,767	21,767	21,767	21,680	21,680	21,680	21,680
Panel B: Excluding Covid-19 period								
NSRI	-10.503	-8.789	7.205	6.354	-7.219	-5.320	6.057	5.472
	(-4.18)	(-3.44)	(3.38)	(2.93)	(-3.26)	(-2.38)	(3.48)	(2.98)
Ret($t-1$)		-13.044		-26.746		-12.460		-21.505
		(-2.73)		(-4.08)		(-2.55)		(-3.61)
Sentiment		-13.351		12.496		-10.127		10.758
		(-3.22)		(3.33)		(-2.56)		(3.15)
World Ret		166.353		-1.920		126.385		-3.905
		(13.44)		(-0.31)		(11.69)		(-0.73)
EPU		-3.785		-7.397		-0.375		-3.752
		(-2.99)		(-3.55)		(-0.33)		(-2.54)
ADS		-11.496		6.650		-4.455		5.562
		(-5.53)		(3.49)		(-2.86)		(3.40)
VIX		-11.042		20.601		-21.293		16.518
		(-2.35)		(5.71)		(-4.97)		(6.33)
R^2 (%)	11.57	34.63	11.52	12.92	9.35	29.68	9.33	10.64
Obs.	19,296	19,296	19,296	19,296	19,257	19,257	19,257	19,257

Table A8: Panel Regression of Stock Market Returns on NSRI: Filter for Article Count and Covid-19. The table shows results for the panel regression of countries' weekly stock market returns (Ret), in basis points, on the percentage change of the weekly-averaged NSRI relative to its past three-month median. Panel A shows results for analysis excluding observations with less than 90 news articles in a calendar month to compute NSRI. Panel B shows results for analysis with data up to December 2019, hence excluding the Covid-19 period. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. As shown in the column headers, stock market returns are based on either the dollar or local-currency denominated stock market indices. Ret($t-1$) is the one period lag of country j's weekly stock market return. Sentiment is the weekly average of country j's news sentiment. World Ret is the weekly return of the MSCI World stock market index, EPU is the percentage change in the weekly-averaged US economic policy uncertainty index. ADS is the percentage change in the weekly-averaged US economic activity index. VIX is the percentage change in the weekly-averaged implied volatility index. We normalize continuous regressors to unit variance and include and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021.

	Contemporaneous				Lagged			
	(1)		(2)		(3)		(4)	
Excluded country	Without control		With control		Without control		With control	
Argentina	-8.643	(-4.06)	-6.570	(-3.29)	8.061	(2.96)	7.066	(2.66)
Australia	-9.128	(-4.11)	-7.156	(-3.45)	7.416	(2.65)	6.338	(2.31)
Bahrain	-9.226	(-4.21)	-6.851	(-3.38)	7.623	(2.76)	6.487	(2.40)
Bangladesh	-9.445	(-4.30)	-7.226	(-3.54)	7.392	(2.68)	6.255	(2.32)
Belgium	-9.321	(-4.25)	-7.113	(-3.49)	7.514	(2.73)	6.402	(2.37)
Botswana	-9.343	(-4.26)	-7.110	(-3.49)	7.567	(2.74)	6.436	(2.38)
Brazil	-9.574	(-4.29)	-7.762	(-3.94)	7.535	(2.64)	6.306	(2.25)
Bulgaria	-9.310	(-4.25)	-7.118	(-3.50)	7.547	(2.74)	6.432	(2.38)
Canada	-9.274	(-4.14)	-7.258	(-3.50)	7.436	(2.65)	6.307	(2.30)
Chile	-9.312	(-4.25)	-7.119	(-3.50)	7.547	(2.74)	6.431	(2.38)
China	-9.475	(-4.24)	-7.240	(-3.49)	7.906	(2.82)	6.758	(2.46)
Colombia	-9.014	(-4.13)	-6.894	(-3.40)	7.816	(2.84)	6.714	(2.49)
Egypt	-9.810	(-4.50)	-7.424	(-3.60)	7.647	(2.69)	6.384	(2.30)
France	-9.255	(-4.10)	-7.375	(-3.56)	7.304	(2.58)	6.225	(2.24)
Georgia	-8.762	(-4.05)	-6.933	(-3.34)	7.789	(2.79)	6.643	(2.43)
Germany	-9.451	(-4.20)	-7.438	(-3.60)	7.234	(2.57)	6.158	(2.23)
Greece	-10.233	(-5.11)	-7.646	(-3.82)	7.704	(2.70)	6.438	(2.29)
Hong Kong	-9.392	(-4.18)	-7.109	(-3.41)	7.007	(2.55)	5.932	(2.20)
Hungary	-9.283	(-4.25)	-7.092	(-3.49)	7.542	(2.74)	6.431	(2.39)
Iceland	-9.379	(-4.27)	-7.190	(-3.52)	7.581	(2.75)	6.468	(2.40)
India	-9.409	(-4.19)	-6.959	(-3.35)	7.991	(2.85)	6.874	(2.51)
Indonesia	-9.214	(-4.14)	-6.795	(-3.33)	7.874	(2.82)	6.816	(2.50)
Ireland	-9.764	(-4.44)	-7.544	(-3.69)	7.588	(2.69)	6.472	(2.34)
Israel	-9.298	(-4.15)	-6.924	(-3.37)	7.830	(2.79)	6.647	(2.42)
Italy	-8.617	(-3.81)	-6.652	(-3.14)	8.068	(2.85)	7.086	(2.57)
Jamaica	-9.522	(-4.32)	-7.065	(-3.43)	7.360	(2.66)	6.199	(2.29)
Japan	-9.267	(-4.10)	-6.944	(-3.33)	7.592	(2.69)	6.433	(2.33)
Kenya	-9.234	(-4.18)	-7.084	(-3.45)	7.463	(2.69)	6.308	(2.33)
Malaysia	-9.660	(-4.36)	-7.408	(-3.58)	7.572	(2.70)	6.432	(2.35)
Mexico	-9.641	(-4.29)	-7.337	(-3.51)	7.032	(2.54)	5.966	(2.19)
New Zealand	-9.022	(-4.07)	-7.129	(-3.41)	7.794	(2.75)	6.774	(2.45)
Nigeria	-9.367	(-4.21)	-7.108	(-3.43)	7.798	(2.80)	6.475	(2.37)
Oman	-9.409	(-4.23)	-6.870	(-3.35)	7.499	(2.69)	6.356	(2.32)
Palestine	-9.316	(-4.25)	-7.122	(-3.50)	7.557	(2.74)	6.439	(2.38)
Philippines	-9.532	(-4.28)	-7.259	(-3.49)	7.735	(2.77)	6.586	(2.41)
Poland	-9.308	(-4.18)	-7.001	(-3.39)	7.371	(2.65)	6.317	(2.32)
Portugal	-9.161	(-4.17)	-6.998	(-3.42)	7.770	(2.81)	6.666	(2.46)
Qatar	-9.591	(-4.34)	-7.606	(-3.77)	7.532	(2.69)	6.477	(2.35)
Romania	-9.302	(-4.25)	-7.110	(-3.49)	7.546	(2.74)	6.432	(2.38)
Russia	-9.485	(-4.23)	-7.056	(-3.39)	6.995	(2.53)	5.959	(2.19)
Saudi Arabia	-9.664	(-4.41)	-7.402	(-3.60)	7.190	(2.60)	5.992	(2.23)
Singapore	-9.386	(-4.18)	-7.304	(-3.51)	7.432	(2.65)	6.303	(2.30)
South Africa	-8.845	(-4.05)	-6.841	(-3.34)	6.701	(2.52)	5.651	(2.15)
South Korea	-8.703	(-3.97)	-7.029	(-3.37)	6.922	(2.50)	5.911	(2.17)
Spain	-9.080	(-3.98)	-6.734	(-3.20)	8.211	(2.95)	7.193	(2.65)
Sri Lanka	-9.193	(-4.12)	-6.872	(-3.32)	7.488	(2.68)	6.287	(2.30)
Taiwan, Province of China	-9.373	(-4.22)	-7.244	(-3.52)	8.044	(2.92)	6.843	(2.52)
Thailand	-9.723	(-4.38)	-7.437	(-3.59)	7.432	(2.63)	6.359	(2.30)
Tunisia	-8.997	(-4.14)	-6.816	(-3.38)	7.475	(2.71)	6.403	(2.36)
Turkey	-8.679	(-3.91)	-6.332	(-3.18)	6.891	(2.47)	5.833	(2.13)
UAE	-9.495	(-4.29)	-7.043	(-3.41)	6.956	(2.56)	5.775	(2.19)
Ukraine	-9.423	(-4.24)	-7.254	(-3.51)	7.555	(2.70)	6.385	(2.33)
United Kingdom	-9.587	(-4.33)	-7.486	(-3.66)	7.802	(2.79)	6.700	(2.44)
USA	-9.004	(-4.07)	-7.208	(-3.46)	7.436	(2.65)	6.435	(2.33)
Vietnam	-9.658	(-4.35)	-7.500	(-3.64)	9.099	(3.84)	7.900	(3.34)
Zambia	-8.824	(-4.05)	-6.406	(-3.29)	7.687	(2.77)	6.646	(2.44)
Zimbabwe	-9.328	(-4.22)	-7.121	(-3.45)	7.569	(2.74)	6.378	(2.35)

Table A9: Panel Regression of Stock Market Returns on NSRI Dropping one Country at a Time. The table shows results for the panel regression of countries' weekly stock market returns (Ret), in basis points, on the percentage change of the weekly-averaged NSRI relative to its past three-month median ($\Delta NSRI_{j,t}$) based on Eq. (5). We drop one country at a time from the regression and report only the resulting coefficient of $\Delta NSRI_{j,t}$ from each regression. The column "Excluded country" indicates the country dropped from the analysis. The regressors are either contemporaneous or one-period lagged, as shown in the column headers. Columns (2) and (4) include the control variables described under Table 6. We normalize continuous regressors to unit variance and include country and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021 and includes observations with at least 60 news articles in a calendar month to compute NSRI.

	Contemporaneous		Lagged	
	(1)	(2)	(3)	(4)
NSRI	0.1045	0.0882	0.0536	0.0177
	(6.83)	(7.06)	(4.92)	(2.05)
Vol($t-1$)		0.6694		0.6490
		(15.91)		(14.73)
Sentiment		0.1171		0.0573
		(4.23)		(2.62)
World Ret		0.1476		-0.1540
		(9.03)		(-10.85)
EPU		0.0426		0.0166
		(4.53)		(1.67)
ADS		0.0054		0.0113
		(0.70)		(1.49)
VIX		0.2338		0.0338
		(12.73)		(2.30)
R^2 (%)	46.74	53.19	46.57	52.78
Obs.	24,101	24,101	24,101	24,101

Table A10: Panel Regression of Weekly Stock Market Volatility on NSRI. The table shows results for the panel regression of countries' weekly stock market volatility (Vol), in percent, on the percentage change of the weekly-averaged NSRI relative to its past three-month median. Volatility is based on the US dollar denominated stock market indices. The regressors are either contemporaneous or one-period lagged, as indicated in the column headers. Vol($t-1$) is the one period lag of country j's weekly stock market volatility. Sentiment is the weekly average of country j's news sentiment. World Ret is the weekly return of the MSCI World stock market index, EPU is the percentage change in the weekly-averaged US economic policy uncertainty index. ADS is the percentage change in the weekly-averaged US economic activity index. VIX is the percentage change in the weekly-averaged implied volatility index. We normalize continuous regressors to unit variance and include and year-month fixed effects in all regressions. Shown in parentheses are t-statistics based on standard errors clustered by country. The sample period is January 2006 – February 2021 and includes observations with at least 60 news articles in a calendar month to compute NSRI.