

Relationship between country risk volatility and indices based on unstructured information*

Relación entre volatilidad del riesgo país e índices basados en información no estructurada

MARTÍN LLADA**

Abstract

This work assesses whether certain indicators constructed from unstructured information published in newspapers contain useful information regarding dynamics of Argentina's country risk volatility, estimated from a GARCH(1,1) model. The analysis covers the period 1998-2019. One standard deviation increment in the indicator that captures manifestations of pessimism is followed by an increment of approximately 0.2% in expected country risk volatility in the consecutive quarter. Out-of-sample exercises confirm that these non-traditional indicators allow for gains in forecast accuracy. These findings are robust to changes in the set of predictors, the specification of the model and the incorporation of new media content.

Key words: *Macroeconomic forecasting, natural language processing, uncertainty, country risk volatility.*

JEL Classification: *E47, E70, G17.*

Resumen

Este trabajo evalúa si ciertos indicadores contruidos a partir de información no estructural publicada en los periódicos contienen información útil respecto de la dinámica de la volatilidad del riesgo país de Argentina, la que es estimada

* I would like to recognize the significant support and important comments provided by Daniel Aromí. I would also like to thank Daniel Heymann, Gabriel Montes Rojas, and Diego Fresoli for their comments and suggestions.

** Universidad de Buenos Aires. Facultad de Ciencias Económicas. Buenos Aires, Argentina. CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política de Buenos Aires. Buenos Aires, Argentina. E-mail: lladamartin@gmail.com

a partir de un modelo GARCH(1,1). El análisis cubre el periodo 1998-2019. Se evidencia que un incremento de una desviación estándar en el indicador que captura manifestación de pesimismo anticipa, en promedio, un aumento de 0,2% en la volatilidad del riesgo país durante el trimestre subsiguiente. Un conjunto de ejercicios de pronóstico fuera de la muestra evidencia que los indicadores no tradicionales permiten mejorar la precisión del pronóstico. Estos resultados son robustos a cambios en el conjunto de regresores, la especificación del modelo y la incorporación de nuevos contenidos difundidos en la prensa.

Palabras clave: *Pronósticos macroeconómicos, procesamiento del lenguaje natural, incertidumbre, volatilidad del riesgo país.*

Claificación JEL: *E47, E70, G17.*

1. INTRODUCTION

The perception of investors associated with the ability of a given country to honor its external commitments is captured by the sovereign risk premium. The pessimistic perception associated with these phenomena is reflected in increases in country risk. In this sense, the level of uncertainty about this variable is especially relevant for policymakers, taking into account its potential impacts over the business cycle. Fernández-Villaverde *et al.* (2011) and de Ferra and Mallucci (2020) show how changes in the volatility of the real interest rate at which emerging economies borrow have an effect on real economic variables (e.g., output, consumption, investment, hours worked, etc).

The level of uncertainty is a non-observable concept, which could be approximated using unstructured information (Baker *et al.*, 2016, Ghirelli *et al.*, 2019, Aromi, 2020). Can valuable indicators of uncertainty about the country risk variable be built using natural language processing tools? Can these indicators based on unstructured data provide information on the dynamics of the country risk volatility? Is the information provided by media content different from the information provided by traditional indicators? In this work we implement an empirical analysis to generate a proxy for the perception of uncertainty about the country risk variable. More specifically, economic press content is processed to generate a set of indicators that approximate different manifestations linked to uncertainty, pessimism, and fear in a specific topic¹: country risk.

There is a wide set of studies that have focused on studying the determinants of the country spreads (Edwards 1985, Kieguel and Lopetegui, 1997, Nogués and Grandes, 2001, Rowland, 2004, Uribe and Yue, 2006, Azar *et al.*, 2007, Grandes, 2007, González-Rosada and Levi Yeyati, 2008, Bellas *et al.*, 2010,

¹ These subjective states are identified by inspecting literature and by recurrence to subjective judgment. The non-traditional indicators aim to capture an expectation of negative scenarios.

Csonto and Ivaschenko, 2013, Mpapalika and Malikane, 2019). In general, these studies conclude that the sovereign risk is influenced by changes in variables related to macroeconomic aggregates, intertemporal liquidity, contagion effects and external shocks. However, there have been very few papers that focus on understanding the dynamics of country risk volatility (Fernández-Villaverde *et al.*, 2011, Palic *et al.*, 2017, de Ferra and Mallucci, 2020). In this sense, this work aims to shed light on this variable of interest, given its key importance in public debt and macroeconomic uncertainty.

This paper focuses on the behavior of the country risk volatility of Argentina, whose evolution is major considering its impact on the real interest rate at which this country borrows. Additionally, this work is timely taking into account that Argentina returns to international capital markets in 2001, following a long period of exclusion. In this sense, we approximated the sovereign risk by the EMBI spread (Emerging Market Bond Index spread) computed and reported by J.P. Morgan from the United States. The analysis covers the period 1998-2019. The volatility is approximated by the conditional standard deviation of the difference in logarithms of the sovereign risk estimated from a GARCH(1,1) model. In turn, the indicators based on unstructured information are built from a large collection of text from *La Nación*, a prominent Argentine newspaper.

Every day, structured and unstructured information is available and potentially relevant to be used by agents of the economy to formulate expectations and make decisions, which influence the dynamics of real economic and financial variables. The literature has focused on studying the information captured by traditional economic indicators; however, there is a growing body of studies that shows the usefulness of exploiting unstructured information in the field of economics (for example, Tetlock, 2007, García, 2013, Loughran and McDonald, 2011, Baker *et al.*, 2016, Ghirelli *et al.*, 2019, Aromi, 2020, among others). In this sense, to exploit these sources of unstructured information, non-traditional indicators based on information published in the economics section of the newspaper *La Nación* are developed. To this aim, first, an unsupervised machine-learning model called Dirichlet Latent Allocation (LDA) was trained to identify all those articles associated with the country risk topic. Finally, following Aromi (2020), quantitative indicators were computed in order to approximate different manifestations of subjective states associated with this topic, using lists of words that belong to the same semantic space that arise from training an unsupervised machine learning model of words vector representation (Pennington *et al.*, 2014).

Preliminary evaluations show that certain non-traditional indicators based on the unstructured information on the country risk topic published by the media are able to explain the future levels of the country risk volatility of Argentina. More specifically, an unexpected increase in pessimism generates a significant increase in the country risk volatility of Argentina. The initial increase amounts to about 0.17 standard deviations in the second quarter, and gradually fades away. These results are robust to changes in the way the country risk volatility is modeled and also the model specification.

Additionally, this work studies to what extent the set of non-traditional economic indicators proposed have relevant information regarding the dynamics of the country risk volatility of Argentina. In this sense, first, an in-sample forecast exercise is carried out from which it is evaluated whether the set of proposed predictors provides useful information on the future values of country risk volatility. Based on these exercises, the results confirm that non-traditional indicators have valuable information regarding the future evolution of country risk volatility. In particular, estimated forecasting models indicate that a one standard deviation increment in the indicator that captures manifestations of pessimism is associated to an increment of 0.21 standard deviations in the next quarter expected country risk volatility of Argentina. Second, out-of-sample forecasts exercises confirm that indicators that exploit unstructured information allows for significant gains in forecast accuracy of country risk volatility computed for Argentina. In this sense, the positive results provide evidence on the usefulness of the proposed indicators to explain the future evolution of the interest variable.

In order to strengthen the patterns documented in this work, we perform a series of robustness exercises. First, the in-sample forecast model is estimated in order to evaluate if the observed regularities still stand when the volatility of the country risk is approximated from an alternative methodology. Second, a set of traditional economic indicators that approximate the internal and external position of the country are incorporated into the forecasting models. Third, dummies variables are incorporated in the forecast model in order to evaluate whether the effects of non-traditional indicators intensify during periods of high volatility. Fourth, new content published by other media are incorporated, namely: *Página 12* and *Ámbito Financiero*. In all cases, the finding previously mentioned is consistent with the robustness exercises. In sense, the results remain unaltered by changes in the model specification, the incorporation of new regressors and new unstructured content. Also, the statistical exercise would seem to suggest that the information provided by the proposed non-traditional indicators based on unstructured information is different from that provided by traditional macroeconomic indicators. Finally, the in-sample forecast model is used to assess whether the regularities reported for Argentina are present in other emerging countries: Brazil, Chile and Peru. The results suggest that non-traditional indicators have useful information regarding the future levels of country risk volatility for the case of Brazil, while they do not seem to capture valuable information beyond what the autoregressive model communicates for Chile and Peru.

The findings show that empirical studies that focus on subjective states² can contribute to improving our understanding of economic dynamics. This result emerges once it is recognized that economic processes emerge as a result of the coevolution of structural and subjective elements. Taking this perspective

² We understand subjective states as perceptions, beliefs or opinions about a specific topic.

into account, it is necessary to find sources of information that allow estimating elements associated with subjective states that can provide new knowledge.

This work contributes to three main strands of the literature. First, this document is aligned with a literature that aims to model the volatility of the risk premium and study its association with the business cycle. Fernández-Villaverde *et al.* (2011) show that the movements in the country-spread volatility, following a likelihood-based approach, faced by four emerging economies (Argentina, Brazil, Ecuador and Venezuela) are negatively correlated with a set of real variables such as the output, consumption and investment. Furthermore, they find that periods of high sovereign spreads are associated with periods of high volatility. de Ferra and Mallucci (2020) estimate the volatility of the risk premium for Argentina following the work of Fernández-Villaverde *et al.* (2011). The authors show times of high interest rate spread volatility are associated with high spreads, low consumption, low output, and a positive trade balance. The current study contributes to this strand of research from a simple model that manages to capture in a favorable way the dynamics of the country risk volatility of Argentina.

Second, this work contributes to a growing set of contributions that use unstructured information to describe dynamics in macroeconomic and financial settings. For example, there are studies that show that measures of optimism or pessimism based on newspaper content predict the stock markets returns (for example, Tetlock, 2007, García, 2013, Loughran and McDonald, 2011). Baker *et al.* (2016) propose an index that measures economic policy uncertainty (the so-called Economic Policy Uncertainty - EPU) computing the fraction of news articles that make a reference to uncertainty and to economic policy. The authors show that this index is closely related to macroeconomic events and is shown to anticipate macroeconomic trajectories in VAR estimations. Ghirelli *et al.* (2019), provide a new EPU index for Spain, based on the methodology of Baker *et al.* (2016). They show that an uncertainty shock yields a significant negative response of real economic variables (e.g., GDP, private consumption and private investment). On the other hand, Baker *et al.* (2019) propose an indicator of volatility in the stock market based on news published in the press for the US, which shows a strong association with the indicator of volatility in the stock market: VIX (Chicago Board Options Exchange Volatility Index) and the realized volatility on the S&P 500. In turn, Aromi (2020) shows that quantitative indicators associated with different subjective states on unstructured press information provide valuable information on future dynamics of US economic activity. Although the current study uses different methodology and sources of unstructured information with respect to the works mentioned previously, the spirit of the statistical exercises is similar. In this sense, the present work contributes to this growing literature based on a set of exercises that show the usefulness of incorporating indicators built from information published by the press into economic models. In particular, this work shows that certain non-traditional indicators contain valuable information that explains the future evolution of the country risk volatility of Argentina. Additionally, a novel contribution of this

work is related to the construction of non-traditional indicators, which is made on a set of specific topics, identified through the LDA model. We believe that this contribution helps to improve the accuracy of the indicators.

Third, our work contributes to a literature that has proven the usefulness of non-traditional data sources (so-called big data) in order to provide valuable information in low-income developing countries. In this sense, Futoshi & Yin (2018) find that online search frequencies about a country, using Google Trends' data for a wide sample of developing countries, significantly correlate with macroeconomic variables, conditional on other covariates. In turn, nighttime lights extracted from processed satellite imagery have been used in economic contexts (Elvidge *et al.*, 1997, Henderson *et al.*, 2009 and Zhao *et al.*, 2020). Analogously, the current work develops indicators based on unstructured press information and evaluates its usefulness in macroeconomics contexts. These new sources of data could be used to reduce the missing data and the long-time lags in data release and improve real-time assessment of economic conditions and the ability to set sound policies.

After this introduction, the next section describes the methodology. Section 3 details the variables and the data sources used. Next, section 4 discusses the results of contemporary associations, while section 5 develops forecasting exercises. Then, in section 6 the results derived from the different robustness exercises are presented. Finally, section 7 summarizes the conclusions.

2. METHODOLOGY

A large quantity of economic data (structured and unstructured) is generated and disseminated everyday through multiple channels that individuals can use to make their economic decisions. Empirical studies usually evaluate the relationship between economic fundamentals and country risk. In this work, we incorporate a novel attribute that attempts to capture different subjective manifestations in the newspaper news in order to evaluate its ability to explain the future evolution of the country risk volatility. Therefore, in this section, we present the methodology associated with the construction of these indicators. In this sense, first, we apply a topic model in order to identify those articles that are associated with the country risk topic. Second, we build a set of quantitative indicators using lists of words that arise from applying an algorithm called GloVe. Lastly, the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is presented to estimate the volatility of the country risk that Argentina faces.

2.1. Topic model

As previously mentioned, in this work we implement the Latent Dirichlet Allocation (LDA) model, first introduced by Blei *et al.* (2003). This model allows to automatically discover the pattern of latent topics or themes within a corpus

(that is, a collection of unstructured documents), and assign the documents to those topics and exploit these documents for a specific purpose. Therefore, this model is a statistical tool within the field of machine learning, which allows the processing of large volumes of unstructured data in order to organize, understand and summarize the information they contain. In contrast to Baker *et al.* (2016) and Ghirelli *et al.* (2019), through the LDA model, we must not define an ad-hoc list of keywords, predefined dictionaries or subjective judgments to identify the articles associated with our topic of interest.

Given a collection of documents, LDA discovers not only the topics that each document addresses, but also the prevalence of these topics in each document. The only arbitrary decision that the researcher must make is to choose the number of topics to be extracted. The objective of LDA is to infer the latent topic structure, which is performed by using the information provided by the words of the documents since they are the only observable variable. In this sense, LDA assumes that documents are random probability distributions over topics, and that topics are random probability distributions over words. In other words, each document contains a set of topics with different probabilities, and the topics contain words with different probabilities. Finally, LDA assumes the independence of words by ignoring the order in which they appear in a document (bag-of-words assumption) since this model exploits the prevalence of words.³

The LDA model assumes that the observed documents were generated through a probabilistic generative process. Therefore, the key inferential task of LDA is estimating the latent parameter structure of this generative process. LDA performs this task by using the words of each document to estimate the parameters of the generative process that are most likely to have generated the collection of documents observed. More precisely, LDA model assumes that a document is composed of topics subject to Dirichlet distribution, θ_d whose parameter is α , which describes the prior knowledge about how topics are distributed in documents.⁴ Each topic is composed of words subject to Dirichlet distribution, φ_k , whose parameters is β , which describes the prior knowledge about how the words are distributed in each topic.⁵ Then, each word (ω_n), that belong to the vocabulary N , in each document is randomly assigned to a topic. Finally, a topic is reassigned to the word given the topic assignments (Z_n), the prevalence of each topic in the document (φ) and the prevalence of the word in each topic (θ). More details can be found in Appendix A.

As we mentioned previously, the key inferential task of LDA consist in estimating the latent structure (the distribution of the parameters θ , Z and φ)⁶

³ More details can be found in Blei *et al.* (2003) and Blei (2012).

⁴ A high alpha means that each document is likely to contain a mixture of most of the topics, while a low alpha value means that each document is likely to contain fewer topics.

⁵ A high beta value means that each topic is likely to be made up of most of the words in the corpus, while a low beta indicates that each topic will have fewer words.

⁶ Where Z represents the per-word topic assignments, θ represents the proportion of topics per-document (that is, the topic distribution of each document), which indicates the extent

that are most likely to have generated the observed document. In this sense, the conditional distribution of the topic structure must be calculated given the collection of documents and the Dirichlet priors. Therefore, the objective of LDA consists in computing the posterior conditional distribution of the latent variables⁷ given the observed variables (the documents and the parameters of the Dirichlet distribution):

$$p(\theta, Z, \varphi | \omega, \alpha, \beta) = \frac{p(\theta, Z, \varphi | \alpha, \beta)}{p(\omega | \alpha, \beta)}$$

The numerator is the joint distribution of the random variables, which can be easily calculated. The denominator is the marginal probability of the observations. However, the posterior distribution cannot be computed, because the evaluation of the denominator is intractable to compute, since the distributions of the hidden variables requires marginalizing over the hidden variables to obtain the model's probability for a given corpus w and priors β and α . In order to approximate this posterior distribution, we use a Collapsed Gibbs sampling as our inference technique, a commonly used alternative introduced by Griffith and Steyvers (2004), to estimate φ , θ and Z , the latent parameters.

2.1.1. LDA implementation

The flow diagram to estimate the LDA model is summarized in Figure 1. First, a set of standard practices in the field of natural language processing is performed (Text Pre-processing Module). In this sense, a tokenization of the articles is performed (a task that is associated with breaking the text into individual words). Then we remove all punctuation, numbers, and stops words (terms such as articles, prepositions, pronouns and other words that do not have a relevant semantic value on the definition of the topics). Once the cleaning tasks are carried out, we perform a vocabulary reduction. In this sense, we only concentrate on those words that occur at least 30 times in order to reduce the word vector space.

The LDA Model Module in Figure 1 summarizes this approach described above. As previously indicated, this work uses the Collapsed Gibbs Sampling proposed by Griffiths and Steyvers (2004) to carry out the iterative process of topic inference. This approach requires to define the specification of values for the parameters of the prior distributions: α (parameter of the per-document topic distributions) and β (parameter the per-topic term distributions), which are defined in $50 / k$ (where k is the number of topics) and 0.1, respectively. The

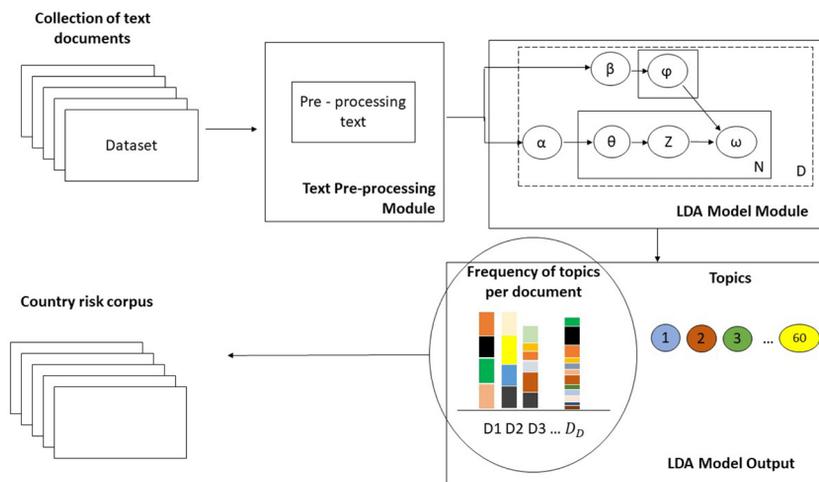
to which each document belongs to each topic, and φ represents the distribution of words in topic k , which is used to define the semantic content of each topic.

⁷ In the context of LDA, θ , Z , φ are hidden or latent variables.

number of topics is set at 60.⁸ On the other hand, the parameter that defines the proportion of topic that the document displays (θ)⁹ is used to select those articles whose proportion parameter associated with the country risk topic is maximum.

The LDA model is applied to the economics section of the newspaper *La Nación* (the Collection of text documents in Figure 1) in order to identify those articles that address the country risk topic. The LDA model was trained with all the articles covering the period 1996-2004 in order to avoid overfitting the model. The topics associated with country risk selected are topic 7 (whose 10 words associated with it with the highest relative probability are: fmi, fondo, argentina, acuerdo, internacional, monetario, organismo, programa, washington y banco) and topic 47 (deuda, bonos, acreedores, argentina, gobierno, canje, reestructuracion, default, títulos, oferta).¹⁰

FIGURE 1
STRUCTURAL OVERVIEW OF THE LDA IMPLEMENTATION



2.2. Indexes based on unstructured information

Once the corpus of documents associated with the topic of interest has been built, we proceed to compute the quantitative indexes based on unstructured

⁸ An analysis with different topics values was carried out, which allows us to arrive at a conclusive distribution of topics by setting $k = 60$.

⁹ θ represents the proportion of topics per-document (that is, the topic distribution of each document).

¹⁰ In English, topic 1: imf, fund, argentina, agreement, international, monetary, agency, program, washington and bank, while topic 7: debt, bonds, creditors, argentina, government, exchange, restructuring, default, securities, offer.

information published in the press. In this section we present the methodology followed in order to compute the non-traditional indices.

2.2.1. Word vector representation

In this work, word vectors models are used to learn word meaning structures. In particular, the GloVe (Global Vectors) model proposed by Pennington *et al.* (2014) is used, which produces linear structures of meaning. This model represents words as vectors and captures the global statistics of the corpus. The objective is to generate quantitative representations that summarize word semantic content by using the global statistics of the corpus and the trained vectors. In other words, this model allows us to identify relationships between words by computing the distance between their numeric representations, through simple algebraic operations. This unsupervised machine learning technique allows us to generate quantitative indicators that summarize information from the press. Furthermore, as Aromi (2020) suggests, the GloVe model has good results in tasks such as: i) resolution of ambiguity in the meaning of a word, ii) entity identification through vector composition, and iii) identification of words indicative of tone or topic. Finally, in contrast to Baker *et al.* (2016) and Ghirelli *et al.* (2019), through GloVe model, we must not define an ad-hoc list of keywords, predefined dictionaries or subjective judgments to identify relevant words in the press related to negative scenarios.

Through the GloVe model, several indices designed to capture manifestations of different subjective states in an economic context can be computed. For this task, first, word vectors are trained using a corpus through which a structure of meaning is built. In other words, for each word, a vector of numbers is obtained, which represents its meaning. Second, a keyword or set of words is chosen that expresses the subjective state that is to be represented. Then, the words strongly associated with the key term defined in the second step are identified, through computing the distance between their respective representative vectors. Finally, the quantitative indicator is given by the frequency of these words (the key term and the list of words strongly associated with it) appear in the corpus. The ability of word vectors to capture the meaning of a word depends on how informative the training corpus is and the effectiveness of the learning model.

The GloVe model trains word vectors in order to capture information about the co-occurrence of words in the training corpus, that is, information on the number of times a word appears in the context of other words. This method is global in the sense that all vectors are computed through an individual optimization exercise. Let W be the size of the vocabulary¹¹ and let X_{ij} denote the number of times word i occurs in the context of (i.e., is close to) word j . The loss function of the GloVe model is given by:

¹¹ The vocabulary size will be discussed later.

$$\sum_{i,j=1}^W f(x_{ij}) [v_i^T * \tilde{v}_j + b_i + \tilde{b}_j - \log(x_{ij})]^2$$

where v_i and \tilde{v}_j are word vectors, $f(x_{ij})$ is a concave increasing weighting function whose function is to weight the most frequent co-occurrences in order to limit the influence of frequent word co-occurrences¹². b_i and \tilde{b}_j are word biases, which are used to account for differences in the frequency of words. This is a log-bilinear regression model. In other words, the training objective of GloVe is to learn word vectors such that their dot product equals the logarithm of the words' probability of co-occurrence. The model is fitted by using stochastic gradient descent (Duchi *et al.*, 2011). More details can be found in Pennington *et al.* (2014).

Following the parameters values that are in line with those used in the natural language processing literature, the vector dimensionality is 100 and the window size used to compute the term co-occurrence is 5. The vocabulary used in the implementation is given by words with a frequency of 100 or higher in the training corpus. An analysis of the robustness of this implementation indicates that the results are not sensitive to variations in the values of these parameters. Vector representations of words are computed by using the package `text2vec` in platform R, using as a training corpus the articles of the economics section of the newspaper *La Nación* from 1996-2004 (as mentioned in section 2.1.1, cleaning tasks were carried out).

2.2.2. *Quantitative indexes*

As previously mentioned, this paper aims to analyze the information content of a set of non-traditional indicators regarding the dynamics of country risk volatility. Therefore, in this subsection, we are going to mention how to generate the quantitative indicators from unstructured information published in the press using the knowledge captured by word vectors, following several techniques of natural language processing. The flow diagram to build the quantitative indexes is summarized in Figure 2.

The procedure that generates the indicators which capture different manifestations of subjective states in the articles that address the country risk topic published by the press involves several steps, once the GloVe model was trained. First, identifying a keyword that expresses the subjective state that is to be represented in the content (e.g., "uncertainty").¹³ Next, the set of K most closely related terms are found based on the cosine distance, a similarity measure, between the respective vectors (which arise from the GloVe model), that is, the distance between the vector associated with the keyword and all words in the

¹² More specifically, following Pennington *et al.* (2014), the weighting function

$$f(x) = \left(\frac{x}{100}\right)^\alpha, \text{ si } x < 100, \text{ otherwise } f(x) = 1.$$

¹³ The keyword is defined by the user.

vocabulary W . Finally, the indicator is given by the frequency of the selected terms in the set of articles that address the country risk topic. See Figure 2.

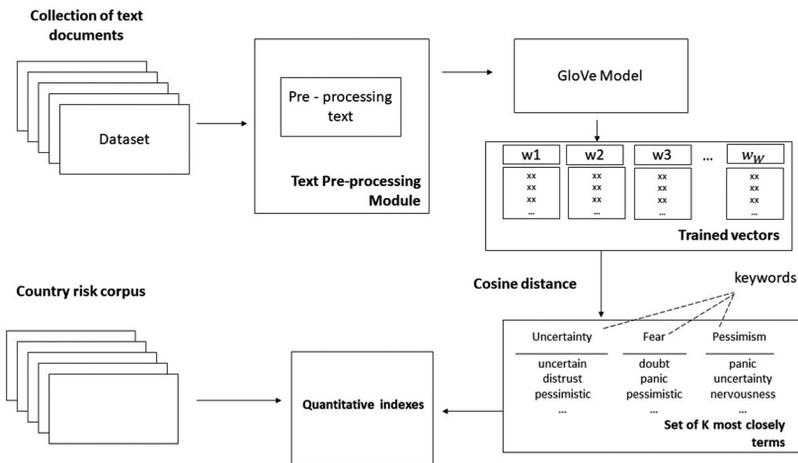
More formally, given a word $w \in K$, where K is the list of words and T is the total number of words in the selected text C . The computed index for the selected text subset is given by:

$$I_C^k = \frac{\sum_{w \in K} c_w^C}{\sum_{w \in T} c_w^C}$$

where c_w indicates the number of times the word w appears in the selected set of text C . The set of selected text corresponds to the articles that address the country risk topic identified through the LDA topic model (Country risk corpus in Figure 2).

Taking into account the considerable level of attention placed on the concept of uncertainty (Baker *et al.*, 2016, Jurado *et al.*, 2015; Rossi & Sekhposyan, 2015), an indicator that captures manifestations of “uncertainty” is computed and evaluated. Furthermore, indicators that approximate related but different manifestations of different subjective states are constructed. More specifically, quantitative indicators that capture manifestations of “pessimism”, “fear” are proposed.

FIGURE 2
STRUCTURAL OVERVIEW OF THE QUANTITATIVE INDEXES



2.3. GARCH model

In order to obtain the country risk volatility of Argentina, a univariate generalized autoregressive conditional heteroskedasticity (GARCH) model, proposed by Bollerslev (1986), is estimated. This model allows to model the unobserved variance of series, which are characterized by large fluctuations with respect to their mean and whose variability changes over time. The basic idea of the GARCH(1,1) model is that the conditional variance has an autoregressive structure and has to be positively correlated with past values.

More formally, let RP_t denote the natural logarithm of the sovereign country risk in day t , then we define the first differences of the country risk is defined as $\Delta RP_t = RP_t - RP_{t-1}$. Assuming ΔRP_t follows a process with the following characteristics:

$$\begin{aligned}\Delta RP_t &= \mu + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2\end{aligned}$$

The unobservable variance of ΔRP_t is estimated through a GARCH(1,1) model and it is assumed ω , α and $\beta > 0$ in order to assure the non-negativity and stationarity of the unconditional variance process, that is, $\sigma_t^2 > 0$ and $\alpha + \beta < 1$. The specification of GARCH model is selected using the Bayesian Information Criterion (BIC). Additionally, the motivation to estimate a GARCH(1,1) model is associated with its advantage to describe the dynamics of the daily volatility of financial variables quite accurately (Bollerslev, 1986, Andersen *et al.*, 2013 and Palic *et al.*, 2017). Also, considering that the conditional variance is an unobservable variable, it must be estimated as well as the parameters that compose it. In this sense, the parameters values (that is, μ , ω , α and β) are estimated using the maximum likelihood method, which estimate the parameters of the GARCH model that are most likely to have generated the ΔRP_t series. Finally, once the conditional variance is obtained, the conditional standard deviation is computed, which will be used as a metric of the country risk volatility of Argentina.

3. DATA

The sovereign spread is approximated through the Emerging Markets Bond Index (EMBI+) computed for Argentina calculated and disclosed by J. P. Morgan. The data for this indicator is taken from the newspaper *Ámbito Financiero*.¹⁴ Moreover, the conditional variance of the country risk of Argentina is estimated

¹⁴ <https://www.ambito.com/>

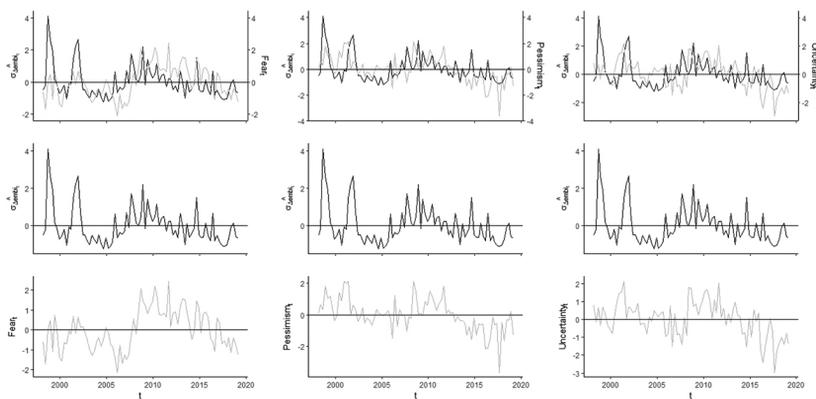
using the first difference of the natural logarithm of EMBI computed for Argentina. The model to be estimated considers that the country risk volatility of Argentina (*vol_emiarg*) is associated with non-traditional economic variables, which exploit the unstructured information disclosed in the press. In this sense, a database at quarterly frequency was developed, that covers the period 1998: Q1-2019: Q1. In turn, daily estimates of conditional volatility are transformed to quarterly frequency using the arithmetic mean, in order to obtain a series at quarterly frequency (Palic *et al.*, 2017).¹⁵

The second type of data is newspaper articles that are used to construct several non-traditional indicators that capture manifestations of different subjective states which are conjectured to be relevant. The procedure carried out to compute them has been detailed in the previous sections. The computed indicators are uncertainty, pessimism and fear. These three indexes correspond to those performed from the GloVe model, using the 50, 500 and 1000 words most strongly associated with each of these keywords. In order to compute a smoother measure for these indicators, the index constructed for the 50, 500 and 1000 words most associated with the keyword was averaged. In this way, the indicators *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted* were obtained. Therefore, given the structure of these indicators, a higher value is associated with higher expectations of negative events regarding country risk indicators or more pessimistic views regarding this variable.

Figure 3 shows that the non-traditional economic indicators that capture important information regarding manifestations of uncertainty, pessimism and fear in the articles regarding the country risk topic, capture the evolution of the country risk volatility of Argentina (*vol_emiarg*). In particular, the notable increment in the three non-traditional indicators coincides with the international economic instability period at the end of the 1990s, which accelerated the surcharge paid by bonds issued in Argentina. Also, the indices capture the period of great country risk volatility after Argentina entered default (2001Q4), and increments are detected around the 2008-2009 international crisis and during the legal dispute between Argentina and the holdouts (2014Q3). Finally, figure 3 shows that the fear indicator seems to clearly capture the contemporary behavior of the country risk volatility of Argentina, while the pessimism and uncertainty indicators seem to favorably anticipate the future evolution of the country risk volatility of Argentina. In the Appendix B, the estimated densities of each series can be found.

¹⁵ The days June 13rd, 2005 and June 30th, 2005, were excluded from the analysis due to their strong influence on the volatility indicator. This influence is associated with the substantial fall (-1.98 and -0.60, respectively) that shows the daily variation of the Argentine EMBI, calculated as the difference in logarithms. In the first case, the observed fall responds to the debt swap. In the second case, the observed fall responds to the exclusion of bonds in default in the computation of the Argentine EMBI.

FIGURE 3
COUNTRY RISK VOLATILITY OF ARGENTINA INDEX AND NON-TRADITIONAL INDICATORS BASED ON PRESS ARTICLES



Note: To facilitate comparisons, the time series were standardized.

Table 1 shows descriptive statistics for the variables used in this work. The table shows, first, that the average quarterly country risk volatility for the period was 2.9%. The period was characterized by high volatility as indicated by the standard deviation of, approximately, 1%. The country risk volatility reached its maximum sample value of 6.5% in September 1998, which is associated with the contagion effect produced by the Russian crisis that occurred in August of that year. In addition, the normality tests suggest the departure from normality, as measured by the test statistic, is statistically significant.

As shown in the following rows of table 1, the distribution of the non-traditional economic indicators based on unstructured information seem to differ from each other. In particular, the average quarterly for the period was 4%, which indicates that 4 out of every 100 words in the articles associated with the country risk topic manifest subjective states associated with pessimism, uncertainty or fear. Also, a certain asymmetry can be seen in the distribution of the indicators, but none is statistically significant. On the other hand, the pessimism and uncertainty indicator reached their maximum sample value in the second quarter of 2001 (2001:Q2), a quarter characterized by the negotiations associated with the *megacanje*, whose objective was to implement the external public debt restructuring. On the other hand, the fear indicator reached its maximum sample value in the third quarter of 2011 (2011:Q3). Also, non-traditional economic indicators show a volatile behavior throughout the analyzed period, which is deduced from their respective dispersion statistics (standard deviation and interquartile range). Finally, only the normality test of the pessimism index suggests the departure from normality is statistically significant.

Lastly, a measure that summarizes the information contained in the non-traditional indicators that capture manifestations linked to uncertainty, pessimism and fear is developed. In this sense, the variable *combined_indices* is constructed as the average of indexes: *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*. Table 1 shows that the *combined_indices* has a symmetric distribution, whose average value reaches 4%.

As we mentioned previously, the level of uncertainty is a non-observable concept. The empirical literature has developed several ways to proxies its behavior, such as the realized volatility of stock market returns (Bloom, 2009), the cross-sectional dispersion of subjective (survey-based) forecasts (Bachmann *et al.*, 2013), or the appearance of certain “uncertainty-related” key words in news publications (Baker *et al.*, 2016, Ghirelli *et al.*, 2019, Aromi, 2020). Therefore, one relevant question is: how does the performance of indices proposed in this paper compare to the performance observed with alternative methods? We consider five indicators to address this question: the volatility of stock market returns (*std_dev_merval*)¹⁶, the disagreement in analyst forecasts about inflation rate and exchange rate (*disagreement_inflation* and *disagreement_er*, respectively)¹⁷, an index that capture uncertainty in news articles regarding the country risk topic (*uncertainty_country_risk*)¹⁸ and an index that capture uncertainty in news (*uncertainty*)¹⁹.

Table 2 shows Pearson correlation coefficient between each indicator. There was a positive correlation between indexes proposed in this paper and indexes computed through alternative methods, except *fear_weighted* that show a negative correlation with the variables that capture analyst forecast dispersion about inflation rate and exchange rate. Therefore, these results suggest that our indexes compare favorably with indices that capture uncertainty level using alternative methods.

¹⁶ More specifically, the indicator is the standard deviation of the difference in logarithms of the daily S&P Merval, the Argentina stock market index. The sample period of these variables is 1998.Q1:2019.Q1. Source: <https://es.finance.yahoo.com/>.

¹⁷ More specifically, these indicators are the mean of the standard deviation for each variable, which are forecast by specialized analysts. We use the Market Expectations Survey (REM) by Central Bank. The sample period of these variables is 2016.Q2:2019.Q1. Source: <http://www.bcra.gob.ar/>.

¹⁸ More specifically, the indicator is the ratio between the number of articles that contain terms referring to two categories: uncertainty (we use the following list of words: “incierto*”, “incertidumbre*”, “inestable*”, “inestabilidad*”) and country risk (“riesgo país”, “riesgo pais”, “embi”), and total of articles published in the newspaper La Nación.

¹⁹ More specifically, the indicator is the ratio between the number of articles that contain terms referring to uncertainty and total of articles published in the newspaper La Nación.

TABLE 1
DESCRIPTIVE STATISTICS

Sample period is 1998:Q1-2019:Q1. Data frequency is quarterly. *vol_embiarg*: country risk volatility of Argentina. *fear_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*. Columns Skewness, Kurtosis and Jarque-Bera perform skewness test, kurtosis test, and Jarque-Bera test for the composite hypothesis of normality, respectively.

Variables	Mean	Median	St. Dev.	Q1	Q3	Minimum	Maximum	Skewness		Kurtosis		Jarque-Bera	
								Stat.	P-value	Stat.	P-value	Stat.	P-value
<i>vol_embiarg</i>	0.029	0.027	0.009	0.023	0.032	0.018	0.065	1.246	0.000	4.147	0.022	26.324	0.002
<i>fear_weighted</i>	0.035	0.034	0.006	0.031	0.039	0.023	0.049	0.220	0.38	2.374	0.164	2.048	0.259
<i>uncertainty_weighted</i>	0.043	0.044	0.004	0.04	0.046	0.031	0.052	-0.282	0.263	3.106	0.843	1.155	0.483
<i>pessimism_weighted</i>	0.039	0.038	0.005	0.036	0.041	0.022	0.048	-0.295	0.24	4.360	0.012	7.692	0.031
<i>combined_indices</i>	0.039	0.039	0.004	0.036	0.041	0.028	0.049	0.085	0.745	3.029	0.948	0.104	0.945

TABLE 2
CORRELATION COEFFICIENT BETWEEN UNCERTAINTY PROXIES

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *fear_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*. *std_dev_merval*: stock market returns volatility of Argentina. *disagreement_inflation*: dispersion of inflation rate forecast, *disagreement_er*: dispersion of exchange rate forecast: *uncertainty_country_risk*: country risk uncertainty index. *Uncertainty*: uncertainty index.

	fear weighted	uncertainty weighted	pessimism weighted	combined indices
<i>fear_weighted</i>	1			
<i>uncertainty_weighted</i>	0.56	1		
<i>pessimism_weighted</i>	0.33	0.82	1	
<i>combined_indices</i>	0.78	0.92	0.82	1
<i>std_dev_merval</i>	0.22	0.16	0.11	0.20
<i>disagreement_inflation</i>	-0.21	0.00	0.40	0.14
<i>disagreement_er</i>	-0.27	0.11	0.37	0.15
<i>uncertainty_country_risk</i>	0.04	0.25	0.35	0.24
<i>uncertainty</i>	0.28	0.14	0.22	0.26

4. ESTIMATION OF CONTEMPORANEOUS ASSOCIATION

In this section, an econometric exercise that evaluates the contemporary association between the country risk volatility of Argentina and a set of non-traditional economic indicators is carried out, conditional on other covariates. Understanding this association is useful, considering that country risk does play an important role in the dynamics of the economy of Argentina, the dynamics of the interest rate levels that Argentina faces in the international credit market, and the dynamics of certain real variables (Fernández-Villaverde *et al.*, 2011, de Ferra and Mallucci, 2020).

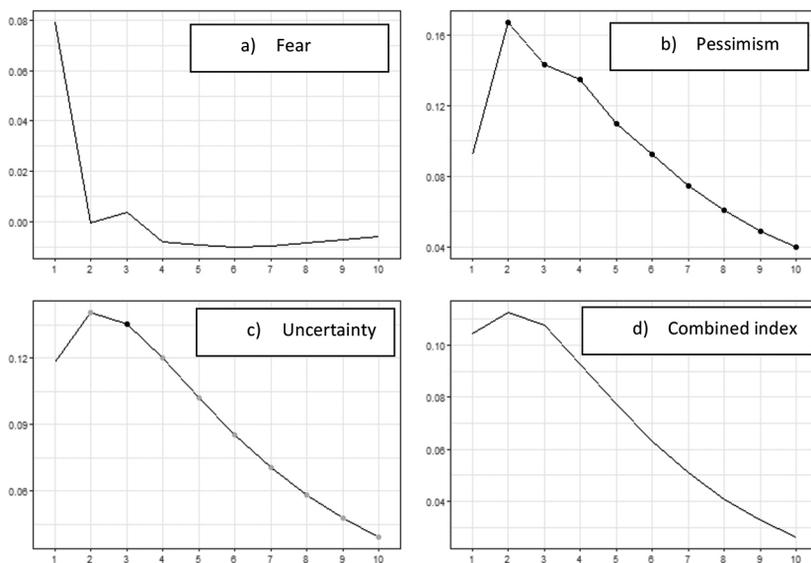
Following Ghirelli *et al.* (2019), we estimate Vector Autoregressive Models (VAR) for these variables in order to identify the effect of non-traditional indicator shock, which capture manifestations of subjective states associated with an expectation of negative events regarding the country risk topic. The VAR models is given by: $X_t = \Gamma(L)X_t + \varepsilon_t$, where X_t is a set of endogenous variables, Γ is a matrix of VAR coefficients capturing the dynamics of the system, and $\varepsilon_t \sim N(0, \Omega)$ is the vector of reduced-form residuals having zero-mean and variance-covariance matrix Ω . The VAR models are estimated by OLS. In each model, we include lags according to the optimal lag length.

To identify shocks we impose the following ordering for the Cholesky decomposition: the volatility of country risk (*vol_embiarg*) was put first, assuming that it is highly probable that this variable can have a contemporaneous effect

on the non-traditional indicators in the system, while the opposite is much less probable. The country risk volatility of Argentina is then followed by a non-traditional indicator.

Figure 3 compares the responses of the country risk volatility of Argentina for an unexpected shock in the indicator that capture manifestations of fear, uncertainty, and pessimism in articles regarding the country risk topic and the variable that combined these indices.²⁰

FIGURE 4
RESPONSE TO NON-TRADITIONAL INDICATORS



Note: Each graph shows responses to a positive shock of one standard deviation in a non-traditional indicator. Black (gray) circles indicate statistical significance at 5 (10)%; solid line, no statistical significance.

Looking at impulse-response functions, a first striking result is the positive response of the country risk volatility of Argentina to changes in the non-traditional indicators proposed in this paper. In particular, an unexpected increase in pessimism generates a significant increase in the country risk volatility of Argentina, which is persistent. The initial increase amounts to about 0.167 standard

²⁰ Results are robust to: i) modeling different VAR lags; ii) ordering country risk volatility indicator last in the vectors, to control for possible contemporaneous effects of the non-traditional indicators on country risk volatility indicator. See Figure C.1 and C.2 in the Appendix C.

deviations in the second quarter, and gradually fades away. On the contrary, the response of volatility of country risk to shocks to the fear indicator is positive but weak and non-statistically significant, while the response to shocks to the combined indicator is positive and persistent but non-statistically significant.

The results shown in this section allow us to conclude that the levels of certain non-traditional indicators contain valuable information about the future evolution of the country risk volatility of Argentina. In this sense, as previously mentioned, the pessimism and uncertainty indicators seem to capture notably the dynamics behavior of the country risk volatility regarding the fear indicator.

5. FORECAST COUNTRY RISK VOLATILITY

The future levels of the country risk volatility are a source of information whose relevance is essential for the design of economic policy. The relevance of understanding the dynamics of the country risk volatility lies in the strong relationship between this variable and the interest rate that countries face in the international capital market and the dynamics of certain real variables (Fernández-Villaverde *et al.*, 2011, de Ferra and Mallucci, 2020). Additionally, a growing set of studies shows favorable results when incorporating indicators based on unstructured information as predictors in models that aim to predict economic phenomena (Tetlock, 2007, García, 2013, Loughran and McDonald, 2011, Baker *et al.*, 2016, Ghirelli *et al.*, 2019, Aromi, 2020). Therefore, this section performs several forecasting exercises in order to evaluate the informational content of a set of non-traditional indicators regarding the future dynamics of the country risk volatility of Argentina. In this sense, the extent to which non-traditional indicators are able to contribute to improve our understanding of the dynamics of these phenomena, inherent in economies exposed to frequent uncertainty shocks such as the Argentine economy, is evaluated.

5.1. In-sample forecast

In this section, the information content of the non-traditional indicators based on unstructured information is evaluated through a series of in-sample forecasting exercises. The informative content of the non-traditional indicators and its ability to anticipate country risk volatility is evaluated. The forecasting models are given by an autoregressive specification that is complemented with an indicator of lagged press content. This model allows us to evaluate the information content of the proposed indicators, conditional on the other covariates' information. The number of lags is selected minimizing the Bayesian Information Criterion (BIC).

More formally, let $\sigma_{\Delta r_t}$ be the country risk volatility of Argentina (*vol_embiarg*) in quarter t , approximated by the conditional standard deviation estimated from the GARCH(1,1) model. The baseline autoregressive model satisfies:

$$\sigma_{\Delta r_t, t+h} = \alpha + \sum_{s=0}^P \beta_s \sigma_{\Delta r_t, t-s} + \mu_{t+h}$$

where P is the number of lags and $\mu_{t+h} \sim N(0, \sigma_i^2)$ is the forecast error, which follows a white noise process. In turn, the predictive ability of press content is evaluated through extended models that incorporate, as a predictor, one of four specifications of the non-traditional indicators. Formally, the forecasting models used to estimate the information content of non-traditional indices based on press content are given by the following equation:

$$\sigma_{\Delta r_t, t+h} = \alpha + \sum_{s=0}^P \beta_s \sigma_{\Delta r_t, t-s} + \beta_I I_t + \mu_{t+h}$$

where I_t can be one of the four non-traditional indicators that exploit unstructured information regarding country risk topic published in the press. The parameter of interest is β_1 . If $\beta_1 = 0$, the baseline model is estimated, while otherwise we estimate an extended forecast model which incorporates only one proposed regressor, namely: *uncertainty_weighted*, *pessimism_weighted*, *fear_weighted* or *combined_indices*. Also, the relative metric of model fit, as indicated by increments in *Adjusted R²*, is analyzed to assess the in-sample forecasting performance of the indicator. This metric allows us to compare how the benchmark model performs with the performance of the extended model. The predictors are standardized to facilitate the comparison of the economic significance of different estimated parameters. Finally, in order to capture the higher uncertainty derived from incorporate generated variables as regressors, residuals bootstrap is implemented (Berkowitz & Kilian, 2000)²¹.

Table 3 shows the estimations of the different specifications of the forecasting models. Column 1 shows the baseline model, while the following columns show the extended model. The baseline models indicate that lagged quarterly country risk volatility is statistically and economically significant predictor of country risk volatility. First, column 1 shows that the country risk volatility process contains an inertia component. Additionally, column 1 shows that the intercept is positive and statistically significant, which indicates that the country risk volatility reaches values of around 3% even when the value of the covariates is equal to zero. These results are in line with those reported in the previous sections. Finally, *Adjusted R²*'s suggest that these variables contain substantive information regarding subsequent levels of the country risk volatility of Argentina.

The estimated extended models indicate that press content adds information regarding future country risk volatility levels. Column 1 shows that the indicator that captures manifestations of fear does not seem to contain relevant information on the future evolution of the country risk volatility. In this sense, the fear

²¹ Taking into account that residuals bootstrap fails for time-dependent data, we implemented this method after to check there is no autocorrelation in the residuals.

indicator is able to provide information regarding the contemporary evolution of country risk volatility (see table 2), while displaying limitations to anticipate the future levels of the country risk volatility of Argentina. Second, columns 3 and 4 show the results associated with the extended model that incorporates the indicator of pessimism and uncertainty, respectively. Those models suggest that increases in the levels of pessimism and uncertainty in newspaper articles regarding the country risk topic anticipate increases in future country risk volatility levels. In particular, a one standard deviation increment in the pessimism index anticipates, on average, an increment of 0.4% in country risk volatility over the next quarter. In other words, a one standard deviation increment in the pessimism index anticipates, on average, an increment of 0.21 standard deviations in country risk volatility over the next quarter. This effect is not economically negligible if the estimated coefficient associated with the indicator of lagged country risk volatility is considered. Additionally, *Adjusted R²*'s point to noticeable gains in anticipatory ability. For example, in the baseline model, the *Adjusted R²* increases from 0.25 to more than 0.28 and 0.27 as the pessimism and uncertainty indicator are incorporated as predictors, respectively (see column 3 and 4). Lastly, the

TABLE 3
FORECASTING MODELS

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *vol_embiarg*: country risk volatility of Argentina. *fear_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with "fear". *uncertainty_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with "uncertainty". *pessimism_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with "pessimism". *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*.

	(1)	(2)	(3)	(4)	(5)
<i>vol_embiarg_t</i>	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>fear_weighted_t</i>		-0.001 (0.001)			
<i>pessimism_weighted_t</i>			0.002** (0.001)		
<i>uncertainty_weighted_t</i>				0.002** (0.001)	
<i>combined_indices_t</i>					0.001 (0.001)
Constant	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)
Observations	84	84	84	84	84
<i>R</i> ²	0.264	0.267	0.304	0.295	0.279
Adjusted <i>R</i> ²	0.255	0.249	0.287	0.277	0.261
<i>F</i> Statistic	29.456 ***	14.776***	17.685***	16.910***	15.671***

Note: nonparametric bootstrapped standard errors in parentheses. **p*<0.1; ***p*<0.05; ****p*<0.01.

indicator that combines the information from the three indicators that exploit the content published in the press does not seem to capture valuable information regarding the dynamics of country risk volatility (see column 5).

The results of the present subsection show that indicators based on unstructured information regarding country risk topic provide valuable information regarding future levels of the country risk volatility of Argentina. In particular, the results suggest that models which incorporate indicators that capture manifestations of pessimism and uncertainty add valuable information regarding future levels of the country risk volatility of Argentina and improve upon baseline forecasts by increasing 3 percentage points the *Adjusted R*². However, the indicator that captures manifestations of fear and the indicator that combines the information from non-traditional indices seem to be unable to capture valuable information regarding future country risk volatility levels. These results are consistent with those results reported in section 4.

5.2. Out-of-sample forecast

To provide further insights on the information content of press media we implement out of sample forecasts exercises in which models are trained recursively with past information. The performance of forecast generated by the baseline autoregressive model is compared to forecasts produced by models that incorporate an additional predictor. Four predictors are considered: *uncertainty_weighted*, *pessimism_weighted*, *fear_weighted* and *combined_indices*.

In this subsection, the predictive ability of non-traditional indicators on the country risk volatility of Argentina will be evaluated. As previously mentioned, this work aims to evaluate the information content of the regressors proposed regarding future levels of the country risk volatility. In this sense, positive results are observed in previously reported in-sample forecast exercises (see section 5.1). In order to evaluate the marginal contribution of each predictor based on unstructured information, we implement out-of-sample forecast exercise. In turn, the aim of this subsection is not to find the best forecasting model, but to evaluate the predictive ability of each regressor. In this sense, the predictive power of each regressor will be evaluated based on its ability to increase the forecast accuracy, which will be evaluated from the gain in terms of reduction of the two proposed objective measures, the RMSE and MAE (Faust and Wright, 2013).

Each forecast model is evaluated computing the root-mean-square prediction error (RMSE) and the mean absolute error (MAE). For extended models these measure of accuracy is also expressed as a fraction of the RMSE (and the MAE) of the baseline model. The benchmark is in the denominator so that numbers less than one indicate that the alternative model outperforms the benchmark. The baseline model is given by an autoregressive model, while the extended model is given by the autoregressive specification that is complemented with an indicator of lagged press content (see section 5.1). The performance of the models is assessed by using a specific starting date for pseudo out-of-sample forecast exercise. The starting dates are selected so that the smallest training

subsample represents 80% of the full sample. Following Faust *et al.* (2013), resampling techniques are implemented to compute the statistical significance of the differences in accuracy.

Following Faust *et al.* (2013) the bootstrap resampling techniques is used to approximate the distribution of the Diebold-Mariano statistic under the null hypothesis. The Bootstrap exercise follows the following steps. We estimate two models: (a) a restricted model that involves estimating an AR(4) process for $vol_embiar\ g_t$ and (b) an unrestricted model that consists of a regression of $vol_embiar\ g_t$ on four lags of itself and three predictors: the first three main components that capture the greatest variability of the set of regressors used in this work. In each bootstrap replication (500 replications), we then resample the residuals of the unrestricted model using wild bootstrap and construct a bootstrap sample of $vol_embiar\ g_{p_t}$ using these resampled residuals, together with the coefficients from the restricted model.

Table 5 shows the results for out-of-sample forecast exercises. For some extended models, the estimated forecast accuracy is higher than that observed in the case of the baseline model. In this sense, three of the four proposed indicators (*weighted_pessimism*, *weighted_uncertainty* and *combined_indices*) contain useful information that allows improving performance in forecasting the country risk volatility. These differences are statistically significant in all cases. However, the fear indicator does not seem to capture valuable information beyond

TABLE 4
OUT-OF-SAMPLE FORECAST

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *vol_embiarg*: country risk volatility of Argentina. *fear_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*. The forecast combination models are done through simple forecast averages. In this sense, *combination_A* combines the extended forecast models that incorporate *fear_weighted*, *pessimism_weight*, *uncertainty_weighted* and *combined_indices*, while *combination_B* averages the baseline forecast model and the extended forecast models.

	RMSE	RATIO	p-value	MAE	RATIO	p-value
base	0.00585			0.00510		
fear_weighted	0.00583	0.996	[0.260]	0.00508	0.996	[0.316]
pessimism_weighted	0.00542	0.926	[0.034]	0.00446	0.874	[0.018]
uncertainty_weighted	0.00511	0.873	[0.010]	0.00426	0.835	[0.014]
combined_indices	0.00552	0.943	[0.058]	0.00468	0.918	[0.05]
combination_A	0.00525	0.898	[0.006]	0.00435	0.853	[0.008]
combination_B	0.00533	0.911	[0.006]	0.00450	0.883	[0.014]

Note: forecast combinations are implemented through simple averages. p-values in brackets.

what the autoregressive model communicates. The result associated with the fear indicator is consistent with those results observed in the in-sample forecast exercise (see section 5.1). On the other hand, it is worth noting that the results obtained from the models that incorporate the indicator *combined_indices* in the out-of-sample estimation differ from those presented in section 5.1. The last two rows of the table show that forecast combinations allow for further gains in accuracy. In summary, these out-of-sample forecast exercises provide further support to the idea that press content regarding country risk topic provides valuable information regarding future levels of the country risk volatility of Argentina.

6. ROBUSTNESS ANALYSIS

This section aims to carry out a set of robustness exercises that provides evidence to strengthen the patterns documented in this work. Therefore, this section is organized as follows. First, in-sample forecast exercises where the country risk volatility indicator is approximated by an alternative methodology are carried out. Second, a set of traditional economic indicators are incorporated into the in-sample forecast model. Third, it evaluates whether the effects of non-traditional indicators intensify during periods of high volatility. Fourth, non-traditional indicators based on new content published by other press media, namely: *Página 12* and *Ámbito Financiero*, are incorporated into in-sample forecast model. Finally, new countries are incorporated, namely: Brazil, Chile and Peru.

6.1. An alternative volatility measures

This section proposes a new way of modeling volatility, the interest variable in this work, as a robustness exercise. In this sense, the mean of the standard deviation of the difference in logarithms of the daily Argentine country risk (Δr_t), approximated by the EMBI index computed for Argentina, is considered as a measure of volatility in quarter t . In this way, we hope that the relationship between the country risk volatility and non-traditional indicators based on unstructured information about the country risk topic will remain. The number of lags is selected minimizing the Bayesian Information Criterion (BIC).

Table 5 shows the results for in-sample forecast exercises, proposed in section 5.1, using a new volatility metric. The baseline and extended models indicate that both lagged quarterly country risk volatility and intercept are statistically and economically significant predictors of the country risk volatility of Argentina. On the other hand, some indicators based on unstructured information published by the press seem to capture valuable information regarding future evolution of the country risk volatility. Finally, *Adjusted R²*'s suggest that these variables contain substantive information regarding subsequent levels of country risk volatility, which was computed by a new methodology.

Therefore, in this subsection it can be seen that, in general, the set of proposed regressors has a statistically significant association with the variable of

interest in this work. The results are consistent with those reported in section 5.1. In this sense, the regularities reported in this subsection allow not only to strengthen the associations between the variables, but also to make these associations independent from the methodology used to estimate the dependent variable, the volatility of the country risk.

Summarizing, the evidence reported above suggests that indices based on press content have valuable information regarding future levels of the country risk volatility of Argentina. These findings are robust to changes in the methodology used to estimate the dependent variable, the country risk volatility. These results are consistent with those results reported in table 3.

TABLE 5
FORECASTING MODELS USING AN ALTERNATIVE MEASURE OF VOLATILITY

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *vol_embiarg_t*: country risk volatility of Argentina using a new methodology. *fear_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices_t*: average of indexes *uncertainty_weighted_t*, *pessimism_weighted_t* and *fear_weighted_t*.

	(1)	(2)	(3)	(4)	(5)
<i>vol_embiarg_t</i>	0.005*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
<i>fear_weighted_t</i>		-0.001 (0.001)			
<i>pessimism_weighted_t</i>			0.002** (0.001)		
<i>uncertainty_weighted_t</i>				0.002** (0.001)	
<i>combined_indices_t</i>					0.001 (0.001)
Constant	0.026*** (0.001)	0.026*** (0.001)	0.026*** (0.001)	0.026*** (0.001)	0.026*** (0.001)
Observations	84	84	84	84	84
R^2	0.238	0.242	0.277	0.267	0.252
Adjusted R^2	0.229	0.224	0.259	0.249	0.233
F Statistic	25.679***	12.950***	15.492***	14.739***	13.637***

Note: nonparametric bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

6.2. In-sample forecast considering traditional indicators

This subsection seeks to evaluate the extent to which the information provided by non-traditional indicators is different from that provided by traditional economic indicators. In this sense, the explanatory capacity of traditional economic

indicators versus the predictive capacity of indicators based on unstructured information published in the press can be evaluated. Additionally, the evaluation of these economic indicators allows a potentially more precise description of the phenomenon analyzed.

Following the literature on the determinants of the country risk premium several indicators related to macroeconomic aggregates, intertemporal liquidity, contagion effects, and external shocks are incorporated (Edwards, 1985, Kieguel and Lopetegui, 1997, Nogués and Grandes, 2001, Rowland, 2004, Uribe and Yue, 2006, Azar *et al.*, 2007, Grandes, 2007, González-Rosada and Levi Yeyati, 2008, Bellas *et al.*, 2010, Csonto and Ivaskenko, 2013, Mpapalika and Malikane, 2019). In particular, the ratio between international reserves²² and GDP²³, the ratio between external debt²⁴ and GDP, national interest rate²⁵, the EMBI+ index computed for emerging countries²⁶ and the EMBI+ index computed for Argentina²⁷. These indicators allow us to capture country's internal and external economic situation, as well as its liquidity and solvency conditions. The sources of data are the Central Bank²⁸ (BCRA), the official open data site²⁹, and *Ámbito Financiero*³⁰. Most of these indicators are available on a monthly basis. Hence, a transformation into quarterly data is necessary. Taking into account the statistical behavior of the series, we transform our data to obtain stationary time series³¹: both EMBI+ index computed for emerging countries and EMBI+ index computed for Argentina. Lastly, we use the natural logarithm of the national interest rate.

Table 6 shows the results for in-sample forecast exercises³², proposed in section 5.1, incorporating a new metric as a regressor: *traditional_indices*. In order to summarize economic information provided by traditional economic indicators, a Principal Component Analysis (PCA) method is implemented. Therefore, the indicator *traditional_indices* correspond to the first principal component which accounts for most variance in the set of traditional economic regressors used in this work. Columns 1 to 5 show that the estimated coefficient

²² Quarterly average of monthly balances at the end of the month (in millions of dollars).

²³ Gross Domestic Product (GDP) interpolated current (in dollars).

²⁴ Total external debt of Public sector and Central Bank from an estimate of gross external debt by resident sector at market value. Quarterly average of monthly of the balances at the end of the month (in millions of dollars)

²⁵ Quarterly average of monthly interest rate, which corresponds to 30- through 59-day term deposits. We decided to use interest rates for term deposits because in the case of Argentina, there is no interest rate that was used consistently as an instrument of monetary policy during the period under analysis.

²⁶ Quarterly average of monthly EMBI+ values for emerging countries.

²⁷ Quarterly average of monthly EMBI+ values for Argentina

²⁸ <http://www.bcra.gov.ar/>

²⁹ <https://www.argentina.gob.ar/modernizacion>

³⁰ <https://www.ambito.com/>

³¹ We conduct the Augmented Dickey-Fuller and Phillips-Perron unit-root test.

³² The number of lags is selected minimizing the Bayesian Information Criterion (BIC).

associated with the traditional indicator carries the expected positive sign. However, this variable does not seem to contain information to explain the future evolution of the country risk volatility of Argentina. Additionally, the estimated coefficient remains mostly unaltered as an indicator that summarizes economic information is incorporated into the model. These results suggest that the indicators that capture different manifestations of subjective states in the content of the press provide valuable information that is different from that provided by traditional economic indicators.

Summarizing, the evidence reported above suggests that indices based on press content have valuable information regarding future levels of the country risk volatility of Argentina. These findings are robust to changes in the set of predictors. These results are consistent with those results reported in table 3.

TABLE 6
NON-TRADITIONAL INDICES VS. TRADITIONAL VARIABLES

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *vol_embiarg*: country risk volatility of Argentina. *fear_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*. *traditional_indices_t*: first principal component which accounts for most variance in the set of traditional economic regressor proposed.

	(1)	(2)	(3)	(4)	(5)
<i>vol_embiarg_t</i>	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>fear_weighted_t</i>		-0.0004 (0.001)			
<i>pessimism_weighted_t</i>			0.002** (0.001)		
<i>uncertainty_weighted_t</i>				0.002** (0.001)	
<i>combined_indices_t</i>					0.001 (0.001)
<i>traditional_indices_t</i>	0.0007 (0.0009)	0.0006 (0.0009)	0.0001 (0.0009)	0.0004 (0.0009)	0.0006 (0.0009)
Constant	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)
Observations	84	84	84	84	84
R^2	0.269	0.271	0.304	0.296	0.283
Adjusted R^2	0.251	0.243	0.278	0.270	0.256
F Statistic	14.922***	9.90***	11.651***	11.221***	10.517***

Note: nonparametric bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6.3. In-sample forecast considering interactions

The model proposed in section 4 assumes that the relationships between two explanatory variables do not change their relationships with the dependent variable. An interesting empirical exercise consists in considering a model in which the estimated associations can take different values according to the economic conditions. In this sense, a conjecture is that volatility levels are related to information volume. During turbulent periods, the dissemination and scope of news is greater, since economic agents are more willing to incorporate information, review their beliefs and modify behaviors (Gorodnichenko, 2008, Coibion and Gorodnickenko, 2012), which induces an effect expected higher on volatility levels. More specifically, we are interested in estimating a model in which the estimated coefficients can take different values in periods of low volatility and periods of high volatility. This exercise serves as a robustness check of the exercises developed in section 5.1.

The estimation methodology involves computing a dummy variable (D_t) that takes a value of one if quarter t was classified as higher volatility period (the level of volatility in period t is greater than one standard deviation of the mean volatility level in period $t-1$) and takes a value of zero otherwise. This variable is used to estimate a flexible model in which different slopes are allowed for depending on the level of volatility level in period $t-1$. Thus, 13 quarters result in periods of high volatility (13% of the sample quarters), while the rest of the quarters correspond to periods of low volatility. This variable is used to estimate the following model:

$$\sigma_{\Delta r,t+1} = \alpha + \beta_1 \sigma_{\Delta r,t} + \beta_2 I_t + \beta_3 D_t + \beta_4 D_t I_t + \mu_{t+1}$$

In other words, in this section we allow the coefficient associated with the non-traditional indicators of the model proposed in section 5.1 to take different values according to the volatility levels of the past period. The results are shown in table 7. The number of lags is selected minimizing the Bayesian Information Criterion (BIC).

Table 7 shows the results for in-sample forecast exercises, proposed in section 5.1, incorporating a term that captures the interaction of non-traditional indicators with a dummy of high country risk volatility in quarter $t-1$. In all cases, the estimated coefficients associated with the interaction variable are positive and not significant. That is, the effects of non-traditional indicators on future country risk volatility do not intensify during periods of high volatility. Additionally, the coefficient that captures the levels of volatility conditional on the value of the independent variables being zero is not higher in periods of high volatility than the level observed when the level of volatility is low.

In summary, the results reported in this subsection are in line with the patterns documented throughout this work. In particular, this subsection shows that indices based on press content have valuable information regarding future

levels of the country risk volatility of Argentina. In turn, there is no statistically significant evidence to conclude that these effects intensify during periods of high volatility. Finally, an analysis of the robustness of this exercise indicates that the results are not sensitive to variations in the definition of high volatility period.³³

TABLE 7
NON-LINEAR MODELS

Sample period is 1998.Q1-2019.Q1. Data frequency is quarterly. *vol_embiarg_t*: country risk volatility of Argentina. *fear_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. *uncertainty_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. *pessimism_weighted_t*: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. *combined_indices_t*: average of indexes *uncertainty_weighted_t*, *pessimism_weighted_t* and *fear_weighted_t*. *dummy_t*: variable that takes a value one if quarter *t* was classified as higher volatility period and zero otherwise. *interaction_t*: variable that allow a non-traditional indicator takes different value according to the volatility levels of the past period.

	(1)	(2)	(3)	(4)
<i>vol_embiarg_t</i>	0.005*** (0.002)	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
<i>fear_weighted_t</i>	-0.0003 (0.002)			
<i>pessimism_weighted_t</i>		0.002** (0.001)		
<i>uncertainty_weighted_t</i>			0.002* (0.001)	
<i>combined_indices_t</i>				0.001 (0.001)
β_{dummy}	-0.0001 (0.007)	-0.001 (0.004)	-0.002 (0.004)	-0.002 (0.005)
$\beta_{interaction}$	-0.003 (0.005)	0.0001 (0.004)	0.0002 (0.004)	0.0002 (0.004)
Constant	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)
Observations	84	84	84	84
R^2	0.279	0.309	0.296	0.280
Adjusted R^2	0.242	0.274	0.261	0.244
F Statistic	7.635***	8.839***	8.314***	7.697***

Note: nonparametric bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

³³ The model proposed in this subsection was estimated by alternatively defining the dummy variable that captures periods of high volatility. First, periods of high volatility were considered those quarters in which the country risk was greater than 1000 basis points. Second, periods of high volatility were considered those quarters in which the level of country risk volatility in quarter *t* was higher than the mean volatility level during period *t*-1. Third, the quarters in which Argentina was in default (2001Q-2005Q2 and 2014Q3-2016Q2) were considered as high volatility periods.

6.4. In-sample forecast incorporating new unstructured content

This work studies to what extent non-traditional economic indices based on natural language processing techniques capture valuable information regarding the future levels of the country risk volatility of Argentina. The indicators were built based on a corpus of articles regarding the country risk topic, which were identified through an unsupervised machine learning (LDA) method as mentioned in detail in subsection 2.1. The press articles correspond to the economics section of the newspaper *La Nación*. Therefore, an interesting exercise could be to incorporate content published in other media. This exercise provides an informative and coherent description of the patterns documented in this work.

The trained model in section 2.1 is applied to the economics section of the newspaper *Página 12* (available from 2002:Q1) and *Ámbito Financiero* (available from 2009:Q1) in order to identify those articles regarding the country risk topic (topics 7 and 47). Once the articles have been identified, the same set of quantitative indicators that captures different manifestations of subjective states is constructed by using the closely associated words to a keyword which are identified by the GloVe model, described in section 2.2.

Table 8 shows the results for in-sample forecast exercises³⁴, proposed in section 5.1, incorporating the indicators proposed in this work, which combine the information from the articles of the newspaper *La Nación*, *Ámbito Financiero* and *Página 12*. Therefore, the indicators developed in this section capture different manifestations of subjective states in the articles regarding the country risk topic published by three Argentine newspapers with national coverage. The information is combined by averaging the standardized newspaper-level series.³⁵ In consistency with the previously reported results, the estimated extended models indicate that press media content adds information regarding future levels of the country risk volatility of Argentina. These findings are robust to changes in the corpus used to develop the predictors.

³⁴ The number of lags is selected minimizing the Bayesian Information Criterion (BIC).

³⁵ More specifically, the indicators of interest are computed for each newspaper, then these are standardized (each regressor is standardized subtracting its sample mean and dividing the difference by its sample standard deviation) and, finally, the average is taken. This way of combining information responds to the overall volume of articles varies across newspapers and time. In particular, the average number of articles published in the economics section of the newspaper *Página 12* throughout the period analyzed reaches 533 articles (on average, 31 articles regarding the country risk topic), while this number rises to 1913 articles (80 deal with the topic of interest) for the newspaper *La Nación* and 1510 articles for the newspaper *Ámbito Financiero* (57 correspond to the country risk topic).

TABLE 8
FORECASTING MODELS INCORPORATING NEW UNSTRUCTURED CONTENT

Sample period is 1998:Q1-2019:Q1. Data frequency is quarterly. *vol_embiarg_t*: country risk volatility of Argentina. *fear_weighted_t*: is the “fear” indicator using new contents. *uncertainty_weighted_t* is the “uncertainty” indicator using new contents. *pessimism_weighted_t*: is the “pessimism” indicator using new contents. *combined_indices_t*: average of indexes *uncertainty_weighted_t*, *pessimism_weighted_t* and *fear_weighted_t*. The non-traditional indicators were generated using *Ámbito Financiero*, *La Nación*, and *Página 12* newspapers.

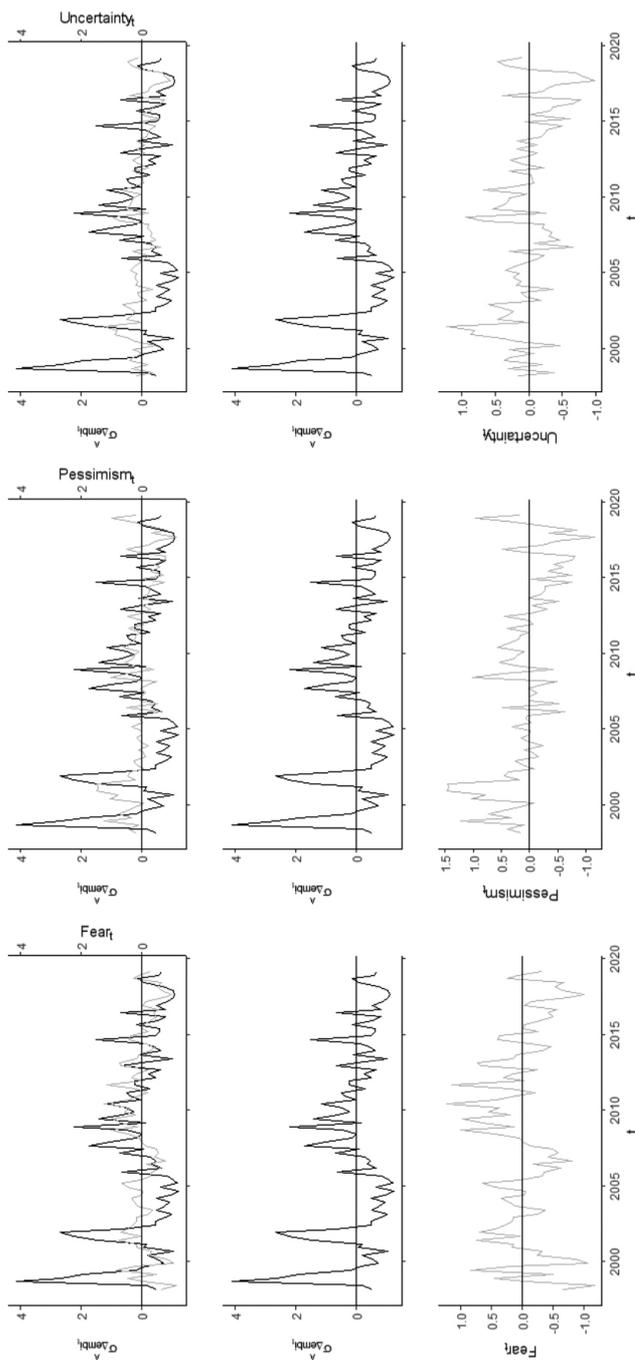
	(1)	(2)	(3)	(4)	(5)
<i>vol_embiarg_t</i>	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>fear_weighted_t</i>		-0.001 (0.001)			
<i>pessimism_weighted_t</i>			0.002* (0.001)		
<i>uncertainty_weighted_t</i>				0.002** (0.001)	
<i>combined_indices_t</i>					0.002 (0.003)
Constant	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)	0.029*** (0.001)
Observations	84	84	84	84	84
R^2	0.264	0.274	0.293	0.305	0.269
Adjusted R^2	0.255	0.256	0.276	0.288	0.251
F Statistic	29.456***	15.286***	16.808***	17.790***	14.905***

Note: nonparametric bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Finally, Figure 2 provides further evidence on the co-movement of the country risk volatility of Argentina and the non-traditional indicators based on the newspaper articles regarding country risk topic published by the newspapers *La Nación*, *Ámbito Financiero*, *Financiero* and *Página 12*. The incorporation of new press content can result in more precise non-traditional indicators. Figure 2 shows these non-traditional indicators based on economic information published in press media contain information regarding the evolution of the country risk volatility of Argentina.

According to the evidence reported above, non-traditional indicators that capture manifestations of subjective states are consistently associated with the country risk volatility of Argentina. More specifically, the increase in the indicators that capture manifestations of fear, pessimism and uncertainty are positively associated with future country risk volatility levels. The incorporation of new unstructured content not only allows reinforcing the regularities reported in previous sections but also makes these findings are robust to changes in the corpus used to develop the non-traditional indicators proposed in this work. These results are consistent with those results reported in table 3.

FIGURE 5
COUNTRY RISK VOLATILITY OF ARGENTINA INDEX AND NON-TRADITIONAL INDICATORS BASED ON PRESS ARTICLES PUBLISHED BY THREE NEWSPAPERS



Note: To facilitate comparisons, both time series were standardized.

6.5. In-sample forecast incorporating other countries: Brazil, Chile and Peru

This work studies to what extent non-traditional economic indices based on natural language processing techniques capture valuable information regarding the future levels of the country risk volatility of Argentina. However, an interesting exercise could be to incorporate other countries into the analysis. In this sense, three Latin American countries, Brazil, Chile and Peru, are incorporated into the analysis in order to assess whether the regularities reported in this work are found in economies with structural and socio-demographic characteristics similar to the Argentine economy. The choice of these countries is linked to the availability of historical digitized information.

The trained model in section 2.1 is applied to the economics section of the newspaper *El Mercurio* of Chile³⁶ (available from 2010: Q1) and *El Comercio* of Peru³⁷ (available from 2013: Q1) in order to identify those articles regarding the country risk topic (topics 7 and 47). Once the articles have been identified, the same set of quantitative indicators that captures different manifestations of subjective states is constructed using the closely associated words to a keyword which are identified by the GloVe model, described in section 2.2. On the other hand, a different methodology was followed for Brazil since its official language is Portuguese. In the Brazilian case, the indicators of uncertainty, pessimism and fear are based on economics section articles published in the newspaper *Folha do São Paulo*³⁸ (available from 1999:Q4) that contain at least one term associated with the country risk topic³⁹ in an article, following the methodology proposed in section 2.2.2.⁴⁰ In turn, the country risk volatility of each country is estimated by following the methodology proposed in section 2.3, where the sovereign spread is approximated through the Emerging Markets Bond Index⁴¹ (EMBI+), calculated and disclosed by J.P. Morgan, computed for Brazil, Chile and Peru. The EMBI+ data for Brazil was taken from the newspaper *Ámbito Financiero*⁴², while the data for Chile and Peru comes from the Central Bank of Chile⁴³.

³⁶ <https://www.emol.com/buscador/?query=econom%C3%ADa>

³⁷ <https://elcomercio.pe/archivo/>

³⁸ <https://acervo.folha.com.br/index.do>

³⁹ The corpus associated with the country risk topic is made up of those articles that contain at least one of the following terms “risco-país”, “risco país”, “Risco-país”, “Risco-País”, “Risco País”, “EMBI+” or “EMBI”.

⁴⁰ In this case, the list of words associated with the term uncertainty is: incerteza, confusão, dúvida, ambiguidade, dúvida, dubiez, hesitação, imprecisão, indecisão, indefinição, indeterminação, insegurança, interrogação, irresolução, oscilação, perplexidade, vacilação, incerto, imprevisibilidade, incertezas, incerta, incertas, incertos, incertamente, incertar, inseguro, duvidoso, desconfiar, diagnóstico, ruído, vulnerável. The list of words used for the term pessimism is: pessimismo. Finally, the list of words associated with fear is: medo, rumores, desconfiar, suposto, considerar, pânico, avisar, dúvida, negar.

⁴¹ Quarterly average of monthly EMBI+ values for Chile and Peru.

⁴² <https://www.ambito.com/contenidos/riesgo-pais-brasil-historico.html>

⁴³ <https://si3.bcentral.cl/siete>

Table 9, 10 and 11 show the results for in-sample forecast exercises⁴⁴, proposed in section 5.1, for each country, namely Brazil, Chile and Peru, respectively. On the one hand, the intercept is positive and statistically significant, which indicates that the country risk volatility reaches values of around 2.7% for Brazil, 2.4% for Chile and 2.7% for Peru even when the value of the covariates is equal to zero. These coefficients are also positive and statistically significant even when the model specification changes. On the other hand, the estimated models indicate that a one standard deviation increment in the indicator of volatility anticipates a mean increment of approximately 0.5, 0.2 and 0.1% in the volatility indicator over the next quarter for Brazil, Chile and Peru, respectively. The estimated coefficient associated with the indicator of lagged country risk volatility for Brazil is similar to that reported for Argentina, while this estimated coefficient for Chile and Peru is lower than the estimated coefficient for Argentina and non-statistically significant (see table 3). Finally, the indicators that capture different manifestations of subjective states have valuable information regarding future levels of country risk volatility for Brazil (whose estimated coefficients are similar to those reported for Argentina). However, non-traditional indicators do not seem to capture valuable information regarding future levels of country risk for Chile and Peru beyond what the autoregressive model communicates.

TABLE 9
FORECASTING MODEL FOR BRAZIL

Sample period is 1999.Q4-2019.Q1. Data frequency is quarterly. *vol_embiarg*: country risk volatility of Brazil. *fear_weighted*: is the “fear” indicator. *uncertainty_weighted*: is the “uncertainty” indicator. *pessimism_weighted*: is the “pessimism” indicator. *combined_indices*: average of indexes *uncertainty_weighted*, *pessimism_weighted* and *fear_weighted*.

	(1)	(2)	(3)	(4)	(5)
<i>vol_embibra_t</i>	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
<i>fear_weighted_t</i>		0.002** (0.001)			
<i>pessimism_weighted_t</i>			0.003*** (0.001)		
<i>uncertainty_weighted_t</i>				0.003*** (0.001)	
<i>combined_indices_t</i>					0.003*** (0.001)
Constant	0.026*** (0.001)	0.027*** (0.001)	0.027*** (0.001)	0.027*** (0.001)	0.027*** (0.001)
Observations	66	66	66	66	66
R^2	0.310	0.364	0.409	0.404	0.428
Adjusted R^2	0.299	0.344	0.391	0.385	0.410
F Statistic	28.776***	18.050***	21.823***	21.363***	23.583***

Note: bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

⁴⁴ In all cases, the number of lags is selected minimizing the Bayesian Information Criterion (BIC).

TABLE 10
FORECASTING MODEL FOR CHILE

Sample period is 2010.Q1-2019.Q1. Data frequency is quarterly. vol_embiarg: country risk volatility of Chile. fear_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. uncertainty_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. pessimism_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. combined_indices: average of indexes uncertainty_weighted, pessimism_weighted and fear_weighted.

	(1)	(2)	(3)	(4)	(5)
vol_embich_t	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)
fear_weighted_t		0.0001 (0.0006)			
pessimism_weighted_t			-0.0003 (0.0008)		
uncertainty_weighted_t				-0.0003 (0.0009)	
combined_indices_t					-0.0003 (0.0008)
Constant	0.024*** (0.0008)	0.024*** (0.0008)	0.024*** (0.0008)	0.024*** (0.0008)	0.024*** (0.0008)
Observations	34	34	34	34	34
R ²	0.103	0.103	0.107	0.106	0.106
Adjusted R ²	0.075	0.045	0.05	0.048	0.049
F Statistic	3.672*	1.784	1.860	1.834	1.844

Note: nonparametric bootstrapped standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

TABLE 11
FORECASTING MODEL FOR PERU

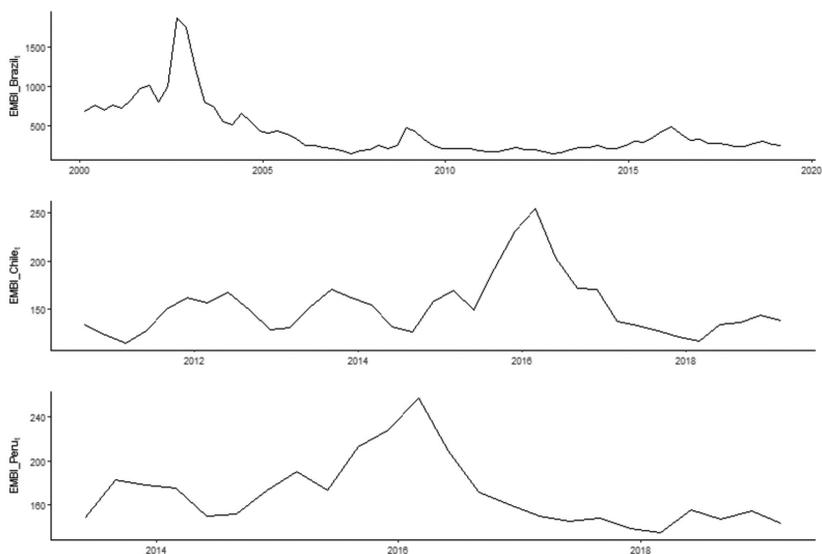
Sample period is 2013.Q1-2019.Q1. Data frequency is quarterly. vol_embiarg: country risk volatility of Peru. fear_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “fear”. uncertainty_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “uncertainty”. pessimism_weighted: is the average of the indices obtained using the 50, 500, and 1000 terms most strongly associated with “pessimism”. combined_indices: average of indexes uncertainty_weighted, pessimism_weighted and fear_weighted.

	(1)	(2)	(3)	(4)	(5)
vol_embipe_t	0.001 (0.0007)	0.001 (0.0009)	0.001 (0.0007)	0.001 (0.0007)	0.001 (0.0007)
fear_weighted_t		0.001 (0.0006)			
pessimism_weighted_t			0.001 (0.0007)		
uncertainty_weighted_t				0.0005 (0.0005)	
combined_indices_t					0.0006 (0.0006)
Constant	0.027*** (0.0005)	0.027*** (0.0005)	0.027*** (0.0005)	0.027*** (0.0005)	0.027*** (0.0005)
Observations	24	24	24	24	24
R ²	0.153	0.184	0.195	0.182	0.205
Adjusted R ²	0.114	0.108	0.119	0.104	0.130
F Statistic	3.971*	2.389	2.551	2.332	2.711

Note: nonparametric bootstrapped standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

These results are in line with the country risk variability of each country during the period analyzed (given by the availability of digital newspaper articles). Figure 6 shows the evolution of the country risk indicator for each country. The country spread of Brazil displayed a strong downward trend starting in the second quarter of 2002. By the end of the period, the Brazilian country risk oscillated around 500 basis points. In contrast, the country risk of Chile and Peru oscillates between 100 and 110 basis points during this period.

FIGURE 6
COUNTRY RISK OF BRAZIL, CHILE AND PERU



7. CONCLUSIONS

The unstructured information contents constitute a source of information that can help to understand the nature of certain economic phenomena. These new sources of information are potentially used by agents of the economy to formulate expectations and make decisions, which influence the dynamics of real economic and financial variables. Therefore, the study of links between indicators based on unstructured information and economic phenomena can shed light on our understanding of the nature of these phenomena.

This paper documents links between the dynamics of country risk volatility and several non-traditional indicators based on press content regarding the country risk topic. The estimated models provide a consistent picture of the association between these variables, that is, higher levels of manifestations of

subjective states -fear, uncertainty or pessimism- in articles regarding country risk topic are associated with a rise in the country risk volatility. Additionally, the evidence suggests that indices based on press media content have valuable information regarding future levels of country risk volatility. In particular, estimated forecasting models indicate that a one standard deviation increment in the indicator that captures manifestations of pessimism is associated to an increment of 0.21 standard deviations in next quarter expected country risk volatility of Argentina. The patterns documented in this work suggest that these non-traditional indicators allow for gains in forecast accuracy. These findings are robust to changes in the set of predictors, the specification of the model and the incorporation of new content published in the press. Also, the information provided by the non-traditional indicators is different from that provided by traditional macroeconomic indicators.

The patterns documented in this work suggest that there are gains associated with incorporating non-traditional sources of information in analyses that study the volatility of country risk. Taking into account the high influence that this variable exerts not only on the dynamics of the interest rate that emerging countries take debt, but also on the business cycle dynamics, we consider it relevant that resources are allocated towards a better understanding of these phenomena. This work shows that models that incorporate indicators that capture different manifestations of subjective states in the press articles describe more precisely the dynamics of country risk volatility.

There are several directions in which the exercises described above can be extended. In this sense, a possible direction is associated with the size of the sample. A higher sample will allow us to increase the precision level of the indicators. Another direction is related to the inclusion of variables that capture expectations of real and financial variables in the analysis. In this sense, several exercises could be estimated to evaluate whether these new variables display similar capacity to provide information regarding future levels country risk volatility. Finally, another possible direction is related to analyzing whether the patterns documented in this work keep when working with high frequency data.

REFERENCES

- Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2013). "Financial Risk Measurement for Financial Risk Management". *Handbook of the Economics of Finance* (pp. 1127-1220).
- Aromi, J. D. (2020). "Linking words in Economic Discourse: Implications for Macroeconomic Forecasts", *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2019.12.001>
- Azar, K., Oreiro, C., Tramontín, F. & Adler, G. (2007). "Determinantes del Riesgo Soberano en Uruguay", Banco Central del Uruguay, Documento de trabajo N°1.

- Bachmann, R., Steffen E. & Eric R. Sims (2013). "Uncertainty and Economic Activity: Evidence from Business Survey Data", *American Economic Journal: Macroeconomics*, 5(2): 217-49.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). "Measuring economic policy uncertainty", *The Quarterly Journal of Economics*, 131(4), 1593-1636.
- Bellas, D., Papaioannou, M. G., & Petrova, I. (2010). "Determinants of Emerging Market Sovereign Bond Spreads: Fundamentals vs Financial Stress", International Monetary Fund (IMF) Working paper, WP/10/281.
- Berkowitz, J., & Kilian, L. (2000). "Recent developments in bootstrapping time series", *Econometric Reviews*, 19(1), 1-48.
- Blei, D. M., Ng, A. Y. & Jordan, M. J. (2003). "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.
- Blei, D. M. (2012). "Probabilistic Topic Models", *Communication of the ACM*, 55(4), 77-84.
- Bloom, N. (2009). "The impact of uncertainty shocks", *Econometrica*, 77(3), 623-685.
- Bollerslev, T. (1986). "A Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31(3), 307-327.
- Calvo-González, O., Eizmendi, A., & Reyes, G. (2018). "Winners Never Quit, Quitters Never Grow: Using Text Mining to Measure Policy Volatility and Its Link with Long-Term Growth in Latin America", World Bank Policy Research Working Paper 8310. Washington: World Bank.
- Coibion, O. & Gorodnichenko, Y. (2015). "Information rigidity and the expectations formation process: A simple framework and new facts", *The American Economic Review*, 105(8), 2644-2678.
- Csonto, B. & Ivaschenko, I. (2013). "Determinants of Sovereign Bond Spreads in Emerging Markets: Local Fundamentals and Global Factors vs. Ever-Changing Misalignments", FMI Working Paper, WP/13/164.
- Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption", *International Journal of Remote Sensing*, 18(6), 1373-1379.
- Edwards, S. (1985). "The Pricing of Bonds and Bank Loans in International Markets: An Empirical Analysis of Developing Countries Foreign Borrowing", NBER Working Paper No. 1689.
- Faust, J., Gilchrist, S., Wright, H. J., & Zakrajsek, E. (2013). "Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach", *The Review of Economics and Statistics*, 95(5), 1501-1519.
- Faust, J., & Wright, J. H. (2013). "Forecasting inflation". *Handbook of Economic Forecasting*, 2, 2-56. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-53683-9.00001-3>
- Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F., & Uribe, M. (2011). "Risk Matters: The Real Effects of Volatility Shocks", *American Economic Review*, 101(6), 2530-2561.

- de Ferra, S., & Mallucci, E. (2020). "Sovereign Risk Matters: The Effects of Endogenous Default Risk on the Time-Varying Volatility of Interest Rate Spreads", Board of Governors of the Federal Reserve System", International Finance Discussion Paper No. 1276. <https://doi.org/10.17016/IFDP.2020.1276>
- Frenkel, R. (1982). "Mercado Financiero, Expectativas Cambiarias y Movimientos de Capital", *Desarrollo Económico*, 22(87), 307-336.
- Futoshii, N., & Yin, R. (2018). "In search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries", IMF Working Paper, WP/18/286.
- Garcia, D. (2013). "Sentiment during recessions", *The Journal of Finance*, 68(3), 1267-1300.
- Ghirelli, C., Pérez, J. J., & Urtasun, A. (2019). "A new economic policy uncertainty index for Spain", *Economics Letters*, Vol. 182, pp. 64-67.
- González-Rozada, M. & Levi-Yeyati, E. (2008). "Global Factors and Emerging Market Spreads", *The Economic Journal*, Vol. 118, pp. 1917-1936.
- Gorodnichenko, Y. (2008). "Endogenous Information, Menu Costs, and Inflation Persistence", NBER Working Paper 14184 Cambridge, Massachusetts: National Bureau of Economic Research.
- Grandes, M. (2007). "The Determinants of Sovereign Bond Spreads: Theory and Facts from Latin America", *Cuadernos de Economía*, 44(130), 151-181.
- Griffiths, T. L. & M. Steyvers (2004). "Finding scientific topics. Proceedings of the National", Academy of Sciences of the United States of America 101(Suppl. 1): 5228-5235.
- Henderson, J. V., Storeygard, A., & Weil, D. N. (2009). "Measuring economic growth from outer space", NBER Working Paper 15199. Cambridge, Massachusetts: National Bureau of Economic Research.
- Kieguel M. & Lopetegui G. (1997). "*Entendiendo el Riesgo País*", Argentine Banking Association (ADEBA) Conferences May 19-21, 1997.
- Jurado, K., Ludvigson, S. C., & Ng, S. (2015). "Measuring uncertainty", *American Economic Review*, 105(3), 1177-1216.
- Loughran, T., & McDonald, B. (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *The Journal of Finance*, 66(1), 35-65.
- Mpapalika, J. & Malikane, C. (2019). "The Determinants of Sovereign Risk Premium in African Countries", *Journal of Risk and Financial Management*, 12(1), 29.
- Newey W. K., & West K. D. (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, Vol. 55, pp. 703-708.
- Newey, W. K., & West, K. D. (1994). "Automatic lag selection in covariance matrix estimation", *The Review of Economic Studies*, 61(4), 631-653.
- Nogués, J. & Grandes, M. (2001). "Country risk: economic policy, contagion effect or political noise?", *Journal of Applied Economics*, 4(1), 125-162.

- Palić, P., Šimović, P.P., & Vizek, M. (2015). "The Determinants of Country's Risk Premium Volatility: Evidence from Panel VAR Model", *Croatian Economic Survey*, 19(1), 37-66.
- Pennington, J., Socher, R., & Manning, C. D. (2014). "Glove: Global vectors for word representation", *Empirical Methods in Natural Language Processing (EMNLP)*, Vol. 14, pp. 1532-1543.
- Rossi, B., & Sekhposyan, T. (2015). "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions", *American Economic Review*, 105(5), 650-655.
- Rowland, P. (2004). "The Colombian Sovereign Spread and its Determinants", Banco de la República de Colombia, Borradores de Economía 315.
- Tetlock, Paul, (2007). "Giving content to investor sentiment: The role of media in the stock market", *Journal of Finance*, pp. 1139-68.
- Uribe, M. & Yue, V. Z. (2006). "Country spreads and emerging countries: Who drives whom?", *Journal of International Economics*, Vol. 69, pp. 6-36.
- Zhao, N., Liu, Y., Hsu, F., Samson, E. L., Letu, H., Liang, D., & Cao, G. (2020). "Time series analysis of VIIRS-DNB nighttime lights imagery for change detection in urban areas: A case study of devastation in Puerto Rico from hurricanes Irma and Maria", *Applied Geography*, 20:102222.

APPENDIX A

The LDA model assumes that the observed documents were generated through a probabilistic generative process. Therefore, the key inferential task of LDA is estimating the latent parameter structure of this generative process. LDA performs this task using the words of each document to estimate the parameters of the generative process that are most likely to have generated the collection of documents observed. Following Calvo-González *et al.* (2018), the generative process that LDA follows after specifying the number of topics k , is:

1. For each topic k , draw a distribution over the words φ_k according to a Dirichlet distribution $\sim Dir(\beta)$, where β is a parameter that describes the prior knowledge about how the words are distributed in each topic.⁴⁵
2. For each document D :
 - a. Draw a vector of topic proportions θ_d according to a Dirichlet distribution $\sim Dir(\alpha)$, where α which describes the prior knowledge about how topics are distributed in documents.⁴⁶
 - b. For each word ω_n that belong to the vocabulary N :
 - i. Draw a topic assignment Z_n according to a multinomial distribution $\sim Multinomial(\theta)$, which depends on the topic proportion θ_d .
 - ii. Given the topic, the most probable word ω_n is selected from the multinomial probability conditional distribution $p(\omega_n | Z_n, \varphi)$.

As we mentioned previously, the key inferential task of LDA consist in estimating the latent structure (the distribution of the parameters θ , Z and φ)⁴⁷ that are most likely to have generated the observed document.

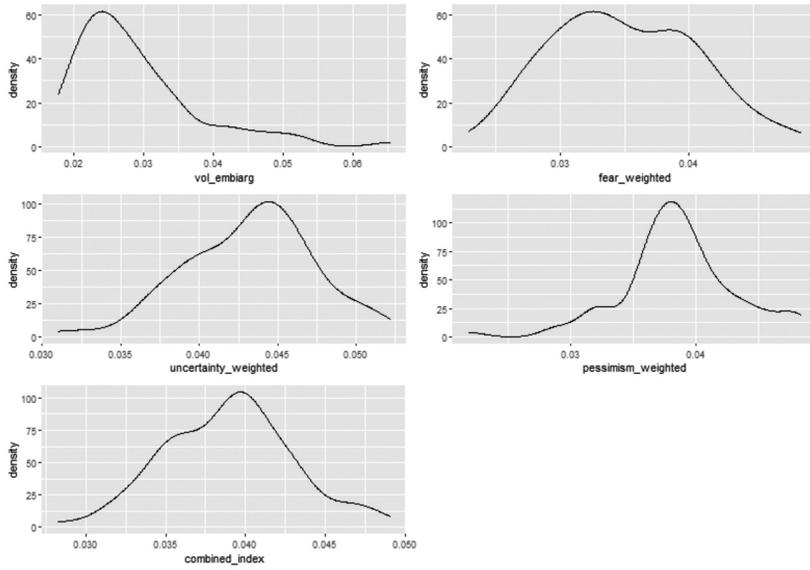
⁴⁵ A high beta value means that each topic is likely to be made up of most of the words in the corpus, while a low beta indicates that each topic will have fewer words.

⁴⁶ A high alpha means that each document is likely to contain a mixture of most of the topics, while a low alpha value means that each document is likely to contain fewer topics.

⁴⁷ Where Z represents the per-word topic assignments, θ represents the proportion of topics per-document (that is, the topic distribution of each document), which indicates the extent to which each document belongs to each topic, and φ represents the distribution of words in topic k , which is used to define the semantic content of each topic.

APPENDIX B

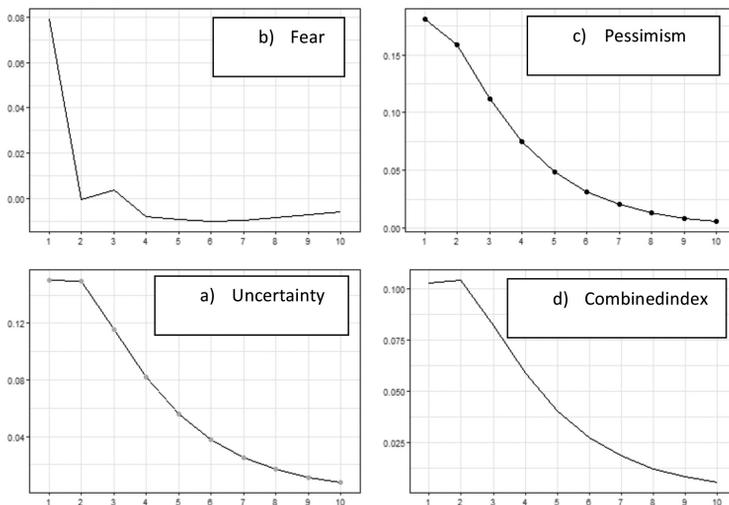
FIGURE B.1
ESTIMATED DENSITIES



APPENDIX C

FIGURE C.1

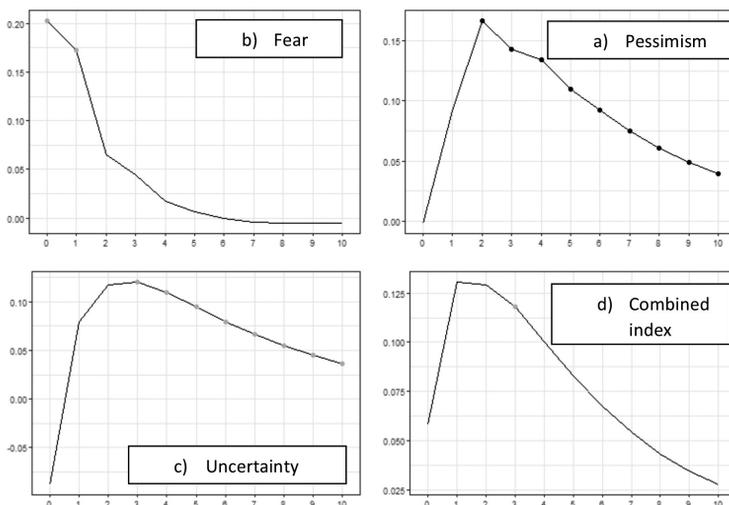
RESPONSE TO NON-TRADITIONAL INDICATORS. DIFFERENT VAR LAGS



Note: Each graph shows responses to a positive shock of one standard deviation in non-traditional indicator. Black (gray) circles indicate statistical significance at 5 (10)%; solid line, no statistical significance.

FIGURE C.2

RESPONSE TO NON-TRADITIONAL INDICATORS. DIFFERENT VARIABLES ORDER



Note: Each graph shows responses to a positive shock of one standard deviation in non-traditional indicator. Black (gray) circles indicate statistical significance at 5 (10)%; solid line, no statistical significance.