



Universidad de
San Andrés

Universidad de San Andrés
Departamento de Economía
Maestría en Economía

**Nowcasting economic activity in Argentina
using newspaper articles**

Juan Pablo ROMERO
DNI: 39.347.171

Mentor: Daniel AROMÍ

Buenos Aires, Argentina
15 de marzo, 2024

Juan Pablo ROMERO

“Nowcasting economic activity in Argentina using newspaper articles”

Abstract

This paper explores leveraging unstructured textual data from Argentinian newspapers to nowcast the monthly Economic Activity Estimator (EMAE) of Argentina. Various economic uncertainty indexes are constructed by applying natural language processing techniques on local newspaper articles. The association between these indexes and the EMAE is analyzed through correlation analysis, Granger causality tests, and out-of-sample nowcasting exercises. The results suggest that the proposed indexes exhibit predictive value for nowcasting the EMAE across diverse data splits. While limitations exist regarding generalization across economic cycles, this study contributes in the viability of extracting valuable signals from news content to gain timely insights into economic trends, and highlights the potential for nowcasting key indicators from unstructured data as text mining capabilities and data availability continue expanding.

Keywords: Nowcasting, Unstructured data, Natural language processing, Economic indicators

JEL codes: C53, E27, C45

“Nowcasting de la actividad económica en Argentina con artículos periodísticos”

Resumen

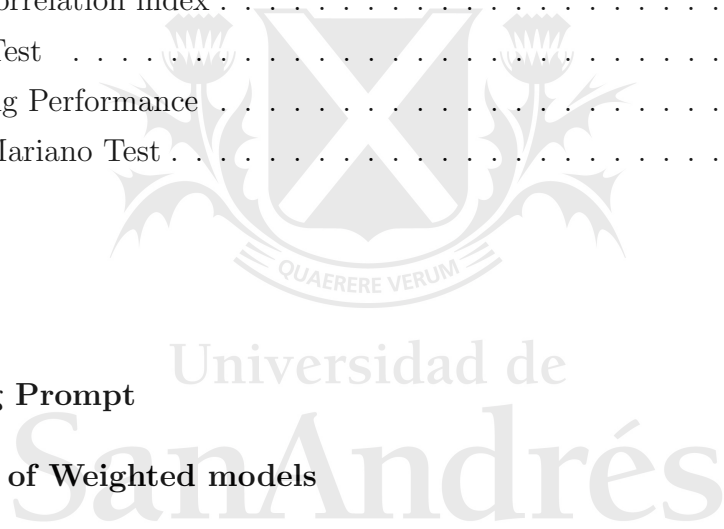
Este trabajo utiliza los datos no estructurados de periódicos argentinos para predecir el Estimador de Actividad Económica (EMAE) mensual de Argentina. Se construyen varios índices de incertidumbre económica aplicando técnicas de procesamiento del lenguaje natural en artículos de periódicos locales. La asociación entre estos índices y el EMAE se analiza mediante análisis de correlación, pruebas de causalidad de Granger y ejercicios de *Nowcasting* fuera de muestra. Los resultados sugieren que los índices propuestos poseen valor predictivo para el *Nowcasting* del EMAE a través de diversas divisiones de datos. Si bien existen limitaciones con respecto a la generalización entre ciclos económicos, este estudio contribuye a la viabilidad de extraer señales valiosas del contenido de las noticias para obtener información oportuna sobre las tendencias económicas, y destaca el potencial para predecir indicadores clave a partir de datos no estructurados a medida que las capacidades de minería de texto y la disponibilidad de datos continúan expandiéndose.

Palabras claves: Nowcasting, Datos no estructurados, Procesamiento del lenguaje natural, Indicadores económicos

Códigos JEL: C53, E27, C45

Contents

1	Introduction	3
2	Literature Review	4
3	Data	6
4	Methodology: Building uncertainty indexes	7
4.1	EPU Index	7
4.2	Customized Index	9
4.3	GPT Index	10
5	Evaluation	11
5.1	Pearson correlation index	12
5.2	Granger Test	13
5.3	Nowcasting Performance	13
5.4	Diebold-Mariano Test	16
6	Results	17
7	Conclusion	21
A	GPT Tagging Prompt	22
B	Development of Weighted models	22
C	Diebold-Mariano Test: Complete list	26
D	Results using MAE instead of RMSE	27



1 Introduction

In the ever-changing landscape of the global economy, having real-time knowledge of the evolution of the main macroeconomic aggregates such as the Gross Domestic Product (GDP) is relevant for making economic policy decisions. Traditional economic indicators often exhibit delays in reflecting the current economic reality since they are published several months after. The adoption of more agile and responsive forecasting methods is essential to address this issue. Nowcasting, with its emphasis on real-time data, emerges as a pivotal tool for policymakers, businesses, and investors to make informed decisions in the dynamic economic environment.

Seminal papers, such as Giannone et al. (2008) and Bańbura et al. (2013), stress the importance of nowcasting in enhancing forecasting precision and facilitating effective policy decisions. In the context of Argentina, where economic conditions can evolve rapidly, the urgency of nowcasting becomes even more pronounced. By providing anticipated insights, nowcasting gives stakeholders, such as central banks, the ability to respond swiftly to economic changes, mitigating risks and capitalizing on opportunities.

There is a large amount of unstructured information available in sources such as newspaper articles and social networks that have a much greater frequency than traditional economic indicators, which allows the generation of metrics on the evolution of economic opinions and the perception of uncertainty. It is worth noting that economic theory and empirical evidence suggest that this perception is a good indicator of economic activity, since it impacts the consumption decisions of families, the investment of companies and the conditions of financial markets.

Therefore, it is relevant to find a way to structure the information available to analyze the evolution of the uncertainty perception, which can contribute to anticipate economic variables and facilitate economic policy decision making.

This paper focuses on the use of unstructured data from Argentinian newspapers to construct nowcasting indicators that can contribute to predict GDP movements. Building upon the foundational works of Tetlock (2007) and Bholat et al. (2015), I aim to utilize natural language processing (NLP) techniques and Large Language Models (LLMs) to extract valuable information from Argentinian news articles. The linguistic nuances and sentiment embedded in these articles are analyzed to create indicators that capture the economic sentiment in Argentina. I find that these indicators are associated with the monthly Economic Activity Estimator (EMAE), which is one of the most consulted economic indicators in Ar-

I would like to thank Agustín Tau and Agustín Ortega for his outstanding comments.

gentina. Therefore, it is possible anticipate official publications through the information of these indicators.

In the subsequent sections of this paper, I present the existing body of literature on nowcasting and the use of unstructured data in economic forecasting. Following the literature review, Section 4 details the methodology for the selection and data collection of Argentinian newspapers. I present the indexes that have been explored and the analytical techniques employed, including the application of NLP techniques and LLMs to extract meaningful insights from the selected dataset. Section 5 presents the evaluation methods chosen and a “weighted” version of two of the indexes. Subsequently, the Results section showcases the outcomes of the nowcasting models, showing the relevance of the analyzed Argentinian newspaper data on EMAE movements. Finally, the conclusion summarizes the findings and the limitations of this approach. It discusses the implications for economic forecasting in Argentina and suggests possibilities for future research.

2 Literature Review

Several papers contribute to the understanding of methodologies, challenges, and applications in the dynamic field of economic nowcasting in Argentina. Cavallo (2013) collects price data from online retailers to construct inflation indexes at a time when official statistics were heavily criticized. Blanco et al. (2021) gathers high-frequency estimators such as financial indicators, tax collection figures, data on sales, industrial production, imports and consumer confidence surveys to nowcast two of the most relevant economic aggregates: Private Consumption and Investment.

There are also efforts to nowcast poverty rates in Argentina. Arakaki et al. (2020) use the availability of the Permanent Household Survey (EPH) micro-data to estimate the poverty ratios for the urban population every six months. Their methodological approach is based on the work of Martín González Rozada, who periodically updates the estimations to anticipate official publications.

On the other hand, there are several papers that aim to quantify non-structured information with the purpose of building indexes that can be associated to macroeconomic variables. Tetlock (2007) builds a measure of pessimism of the *Wall Street Journal*’s “Abreast of the Market” columns by counting the words included in the General Inquirer (GI), a quantitative content analysis program. He finds an association between his indicator and daily stock market activity from the United States. Bholat et al. (2015) also argue that “text mining” is useful for addressing topics of interest for central banks and policy makers. They present several techniques that can be applied to quantify the information contained in financial

contracts, newspapers, social media and various kinds of reports. Boolean and dictionary techniques are the most intuitive suggestions, but they also mention more elaborated approaches such as Latent Semantic Analysis and Descending Hierarchical Classification.

The seminal work of Baker et al. (2016) has inspired several works on this topic. The authors create the economic policy uncertainty (EPU) index that is based on newspaper coverage frequency. They find evidence that this index correlated with policy-related economic uncertainty movements such as tight presidential elections, Gulf Wars and the 9/11 attacks. They also find that peaks of uncertainty were aligned with declines in investment, output, and employment in the United States.

Ghirelli et al. (2019) replicate and extend this analysis for Spain by widening press and time coverage, and they also find significant negative responses of GDP, private consumption and private investment as an effect of uncertainty shocks measured from their new index. In addition, Avela and Lehmus (2023) and Burri (2023) have also constructed indexes for measuring negative economic sentiment using newspapers titles. By leveraging multiple nowcasting exercises, they identify correlations with the economic indicators of Finland and Switzerland, respectively.

While the work of Baker et al. (2016) has become widely accepted as a reasonable proxy of uncertainty, methods for using text as data have drastically improved and manual labeling based on keywords is no longer the only feasible way to construct an EPU index. LLMs and Machine Learning techniques are increasingly used across several fields, and authors such as Gillmann and Kim (2021) develop a deep learning algorithm to identify economic uncertainty in newspaper articles. They base their analysis in a recurrent neural network that uses pre-trained embeddings from GPT-2, a language model developed by OpenAI.

Aromí et al. (2022) aim to organize unstructured text information as well, but instead of looking at newspapers, they propose an index to describe the economic opinions transmitted by Argentinian users on Twitter. After economic messages are identified, they undergo a classification based on the frequency of words associated with uncertainty utilized within them. They find a strong link with relevant economic and political events.

To the best of my knowledge, there are no papers that use these techniques to structure information provided by Argentinian newspapers to nowcast economic indicators for the country. This paper aims to contribute to this topic.

3 Data

This paper utilizes textual data from newspapers to construct economic uncertainty indexes. The primary source for obtaining newspaper articles is the *Factiva* repository provided by Dow Jones. While the indexes presented in Section 4.1 use a combination of local and foreign newspapers developed by Andres-Escayola et al. (2022), the other indexes are built using the following local newspapers:

- El Economista
- La Gaceta
- Infobae
- La Nación
- El Cronista
- La Voz del Interior

The selection criteria prioritized local newspapers with consistent frequency from 2018 onwards¹. No specific filtering criteria was applied. Table 1 shows the number of articles available in each newspaper:

Table 1: Argentinian newspapers available in *Factiva*

Newspaper Name	Number of articles
Infobae	1,428,568
La Nación	250,151
La Gaceta	230,097
La Voz del Interior	176,055
El Cronista	81,041
El Economista	44,604

Note: articles available from January 2018 to April 2023

Like Avela and Lehmus (2023) and Burri (2023), I am leveraging newspaper titles, which tend to capture the main message of articles and can be considered as a good summary despite being brief.

¹Some newspapers such as Clarín are excluded since the frequency available in *Factiva* was intermittent.

4 Methodology: Building uncertainty indexes

In this section I present three indexes that are built using unstructured data from *Factiva*'s newspapers. I analyze whether they contain any relevant information that could be useful to anticipate the official publication of economic indicators and compare them to understand which one has a better predictive power.

4.1 EPU Index

Andres-Escayola et al. (2022) construct Economic Policy Uncertainty (EPU) indexes for some Latin American economies (Argentina, Brazil, Chile, Colombia, Mexico, Peru, Venezuela) following the spirit of Baker et al. (2016). The indexes are built following the approach of Ghirelli et al. (2019) for the computation of such an index for Spain, and they are periodically updated by the *Banco de España* (Bank of Spain)². I utilize Argentina's EPU index as the first indicator of economic uncertainty.

To access newspaper articles, Andres-Escayola et al. (2022) rely on the *Factiva* repository. For each country, they construct the following alternative EPU indicators: 1) an index based on available local sources; 2) an index based on the available foreign newspapers; 3) an index constructed by pooling together all available sources (local and foreign newspapers). They select the following sources of **foreign press**:

- Anglophone press: Los Angeles Times, The Boston Globe, The Globe and Mail, The New York Times, The Telegraph U.K., The Times U.K., Chicago Tribune, The Guardian U.K., The Wall Street Journal, The Washington Post, The Economist.
- Spanish press: El País, El Mundo, Expansión, ABC, Cinco Días, El Economista, La Vanguardia.

In the **local press** index, they consider the most-read local generalist newspapers available in *Factiva*. For the case of Argentina, they include Clarín, La Nación and Infobae.

For each country, the index is built by considering, in each newspaper, the number of articles containing at least one keyword related to the following concepts:

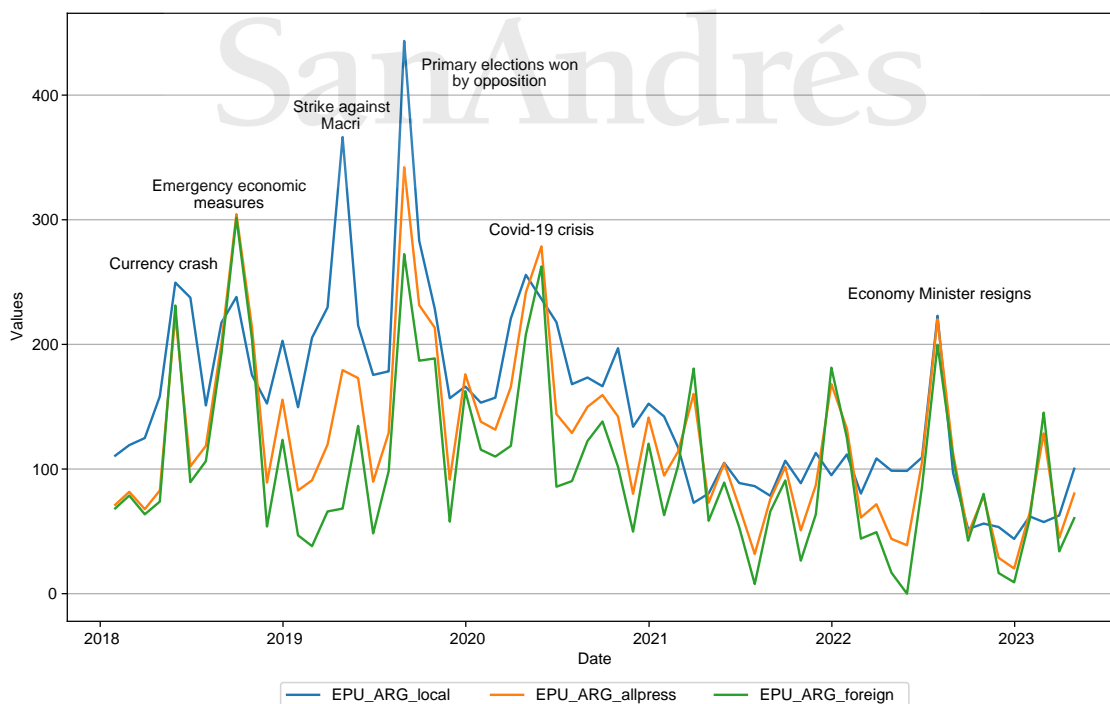
- Uncertainty: uncertain, uncertainty/uncertainties, unstable, instability/instabilities, risk(s).
- Economy: economic(s), economy.

²<https://www.bde.es/wbe/es/areas-actuacion/analisis-e-investigacion/america-latina/contenidos/publicaciones-latinoamerica/>

- Policy: the name of the central bank of the country, the name of the government's workplace in the country, parliament, government, federal reserve, treasury, tariff(s), deficit(s), budget(s), public spending, debt ceiling, exchange rate(s), currency crash(es), sovereign debt, public debt, fiscal policy/policies, monetary policy/policies, legislation, reform(s), tax(es), overhaul, rule(s), norm(s), normative, regulation(s), law(s).

As mentioned before, the construction of the indexes closely follows the procedure used by Baker et al. (2016). Firstly, they count the number of articles published each month in each newspaper that contains words in the mentioned blocks of keywords. Secondly, they express this count as a ratio of articles published by the newspaper in the month. When dealing with local press, they divide this count by the total number of articles published that month, while in the foreign case they use the number of articles published that month that were about the country of interest. Thirdly, they standardize each monthly series of scaled counts by dividing it by its standard deviation to make the volatility of the series comparable across newspapers. Fourthly, for each country they average the newspaper-based standardized series across newspapers to compute an aggregated index. Finally, they re-scale the resulting index to obtain a set of homogeneous country-based EPU indicators. They also average the country-specific EPU measures across countries (simple mean) to come up with aggregated EPU indexes for the Latin American region.

Figure 1: Foreign, local and all press-based EPU indexes for Argentina



Source: Banco de España

In Figure 1, it can be observed that the EPU indexes (in the “local”, “foreign”, and “allpress” versions) coincide with significant political events, such as the currency crash that occurred in Argentina in May 2018 and the extension of Covid-19 lockdowns in April–May 2020. Additionally, it appears that the “local” index exhibits higher volatility, as indicated by generally higher peaks.

4.2 Customized Index

In this section I present another approach to construct an uncertainty index built on Argentinian newspapers. Following the spirit of Andres-Escayola et al. (2022) and Baker et al. (2016), I construct a “Customized” index using Dow Jones’ *Factiva* as well, but leveraging a broader number of local newspapers and a different set of keywords.

For this specific index, I use a dictionary of negative words that was firstly introduced by Loughran and McDonald (2011) and then updated by the authors. The 2355 words that have been tagged as negative include the following examples: anomalies, instability, risky, turbulence, unexpected, unpredictable, volatility, deteriorated, tragic, worries, unsustainable and crisis³.

I am using this set of keywords since it is a well-established lexicon of negative financial terms, which allows for a straightforward replication of this work. Some of them may be similar to the keywords used by Andres-Escayola et al. (2022) since the notion of negativeness is related to the concept of uncertainty (Loughran and McDonald, 2011).

The index is constructed as follows: Firstly, I take the titles of all the mentioned newspapers that are available in *Factiva* from 2018 onwards. Following basic NLP pre-processing techniques, such as converting all text strings to lowercase and segmenting them into word sets, I compute the proportion of “negative” words present in each title relative to the total word count. Finally, these shares are aggregated on a monthly basis using the mean and then scaled to generate a score ranging from 0 to 100.

³Since I am using newspapers in Spanish, English negative words have been translated into Spanish using *Google Translate*.

Figure 2: Foreign, local and all press-based EPU indexes

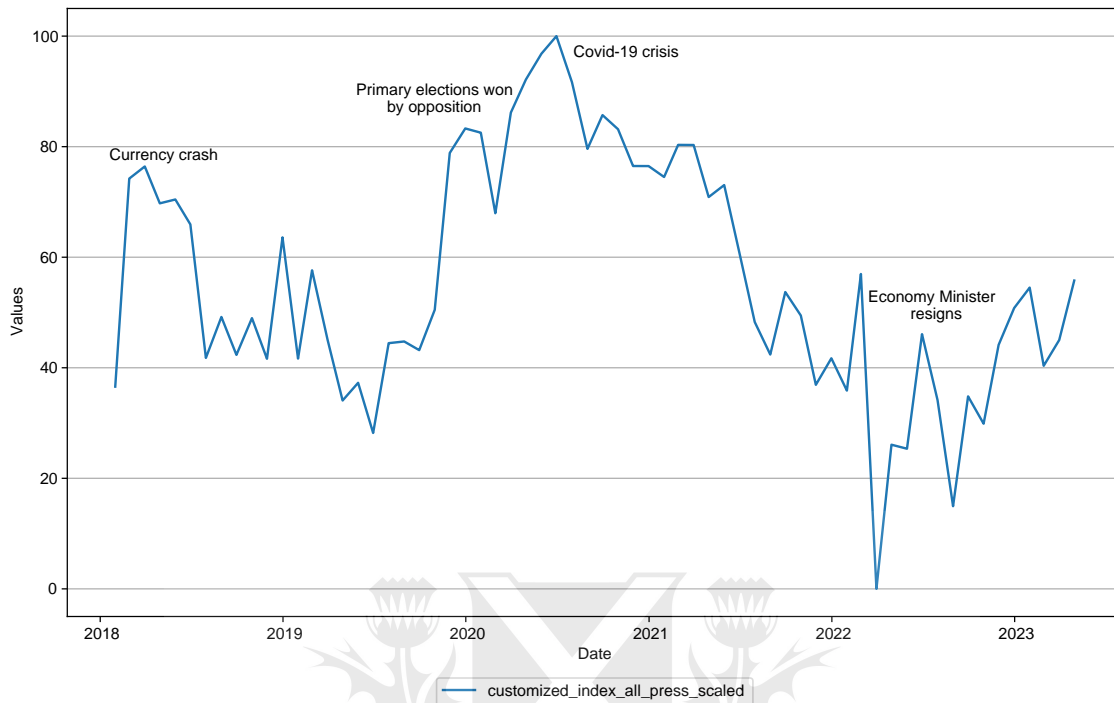


Figure 2 illustrates the evolution of the Customized index. It can be observed that this indicator also demonstrates spikes coinciding with significant political events, such as the currency crash, the opposition’s victory in the primary elections in August 2019, and the Covid-19 crisis.

4.3 GPT Index

In this section I present the third index proposed in this paper, which is based on a tagging using GPT-3.

GPT-3 (Generative Pre-trained Transformer 3) is an advanced artificial intelligence language model developed by OpenAI. It belongs to the family of transformer models, which are based on deep learning architectures primarily designed for NLP tasks. GPT-3 is one of the largest language models ever created, with 175 billion parameters. These parameters are learned from vast amounts of text data, enabling the model to understand and generate human-like text across a wide range of topics and styles.

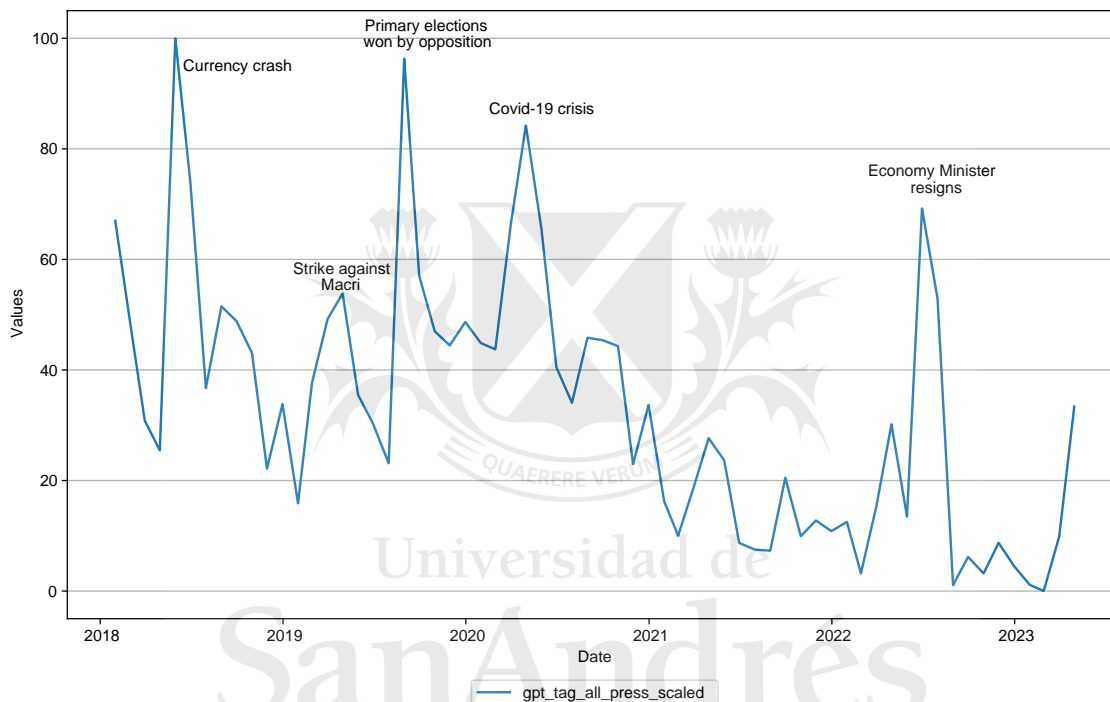
One of the remarkable features of GPT-3 is its ability to generate coherent and contextually relevant text given a prompt or input. This makes it highly versatile and applicable in various NLP tasks, including text generation, translation, summarization, question answering, etc. This is the feature that is leveraged to measure the economic uncertainty presented in Argentinian articles. Since researchers such as Gillmann and Kim (2021) and Trust et al. (2023) have explored LLMs for designing uncertainty indexes, I advocate for the inclusion

of an approach from this emerging field.

The index is constructed using the same input described in the preceding section. Specifically, it leverages local newspapers from Argentina accessible through *Factiva* from 2018 onwards. Each newspaper title is assessed for its relevance to economic uncertainty on a scale from 0 to 100 using a prompt provided to GPT-3⁴.

With the same logic of Section 4.2, the scores are aggregated on a monthly basis using the mean and then rescaled.

Figure 3: Foreign, local and all press-based EPU indexes



As observed in previous instances, Figure 3 illustrates that the peaks of the GPT index align with politically relevant events in Argentina. This alignment provides an indication that the model effectively captures the underlying uncertainty inherent in newspaper headlines.

5 Evaluation

As mentioned in Section 1, I am interested in understanding whether the uncertainty indexes I constructed are associated with the monthly Economic Activity Estimator of Argentina (EMAE) and if they can help anticipate official publications. The EMAE is an economic indicator used to measure the overall level of economic activity in the country. It is similar to other countries' gross domestic product (GDP) or gross national product (GNP) indicators.

⁴The specific GPT-3 model employed was gpt-3.5-turbo. The precise prompt utilized can be found in Section A of the appendix.

It is calculated and published by the Instituto Nacional de Estadística y Censos (INDEC), Argentina's national statistics institute, and serves as a key metric for monitoring economic trends and making policy decisions.

I contrast the proposed indexes with the seasonally adjusted version of the EMAE (*EMAE desestacionalizado*) to remove the seasonal fluctuations typically observed in economic data. To do so, I run some typical evaluation exercises:

- Pearson correlation index
- Granger Test
- Evaluation of the predictive capacity of the models
- Diebold-Mariano Test

These evaluation procedures cover the period from January 2018 to April 2023, ensuring consistency throughout the analysis.

5.1 Pearson correlation index

The Pearson correlation index is a statistical measure that quantifies the linear relationship between two continuous variables.

This index is denoted by the symbol r and ranges from -1 to 1. A value of $r = 1$ indicates a perfect positive linear relationship, meaning that as one variable increases, the other variable also increases proportionally. Conversely, a value of $r = -1$ indicates a perfect negative linear relationship, where as one variable increases, the other variable decreases proportionally. Additionally, a value of $r = 0$ suggests no linear relationship between the variables.

The formula to calculate the Pearson correlation coefficient between two variables X and Y with n data points is given by:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Where \bar{x} and \bar{y} are the means of variables x and y respectively.

The Pearson correlation coefficient is widely used in various fields including statistics, economics, psychology, and sociology to assess the strength and direction of the relationship between two variables. It is the first formal indicator presented for assessing the relationship between the proposed indexes and the EMAE, supplementing the visual observations drawn from the preceding section's plots.

5.2 Granger Test

The Granger causality test is a statistical hypothesis test used to determine whether one time series can be used to forecast another. Developed by Granger (1969), it assesses whether the past values of one variable improve the prediction of another variable beyond what can be predicted using only past values of the second variable itself.

Formally, let Y_t and X_t be two time series. The Granger causality test involves estimating two vector auto-regressive (VAR) models: one that includes only past values of Y_t and another that includes both past values of Y_t and X_t . The null hypothesis is that X_t does not Granger-cause Y_t . Rejection of this null hypothesis suggests that X_t provides significant information for forecasting Y_t beyond its own past values.

In the context of nowcasting, the Granger causality test can be employed to assess whether certain leading indicators or alternative data sources provide valuable information for forecasting economic variables of interest. Consequently, I employ this test to evaluate the informativeness of the proposed indexes regarding Argentinian economic activity.

5.3 Nowcasting Performance

I conduct linear regression analyses to assess the predictive capability of uncertainty indexes in nowcasting the EMAE. Each model includes one lag of the EMAE and the contemporaneous version of the corresponding index. Following the usual procedure in the literature, the results from an autoregressive (AR) model are also presented, which is considered as a benchmark model.

To ensure the reliability and generalizability of the findings, I implement a comprehensive approach that involves partitioning the dataset into multiple in-sample and out-of-sample windows.

In forecasting literature, the in-sample window represents historical data utilized for model training, while the out-of-sample window comprises data withheld for evaluating the model's predictive performance. By segregating the dataset into these distinct windows, I facilitate the assessment of the model's ability to generalize to unseen data, a critical aspect of model validation in predictive analytics. This methodological approach ensures that the model's predictive capabilities extend beyond the memorization of training data and thus enhancing its reliability and applicability in real-world forecasting scenarios.

The windows are defined based on different proportions of the total dataset, including segments representing 60%, 70%, 80%, and 90% of the observations. Gholamy et al. (2018) argue that 70% and 80% are the most effective proportions in empirical analyses, and I include two extra possibilities as a control check. By varying the size of the in-sample and

out-of-sample windows, I aim to capture a diverse range of scenarios, reflecting varying degrees of data availability for predictive modeling.

Additionally, I employ three distinct windowing techniques, which are widely used in the forecasting literature (West and McCracken, 1998), to evaluate the performance of the uncertainty indexes: fixed, recursive, and rolling.

- **Fixed Window:** In this method, a predetermined length of the dataset was designated for model training, while the subsequent observations constituted the testing set for evaluating predictive performance.
- **Recursive Window:** Unlike the fixed window approach, the recursive window technique involves iteratively expanding the training set with each new observation, enabling the model to adapt and update its parameters over time.
- **Rolling Window:** In the rolling window method, a moving window of fixed length is utilized for training the models. This scheme fixes the initial in-sample size and, with each iteration, the window shifts forward in time, incorporating new observations while discarding older ones.

By employing these diverse windowing techniques across different proportions of the dataset, I aim to comprehensively evaluate the predictive performance of the uncertainty indexes. Subsequently, I calculate the root mean square error (RMSE) for each out-of-sample window technique.

The RMSE is a widely used metric in statistics and machine learning to assess the accuracy of a predictive model. It measures the average magnitude of the errors between predicted and observed values, providing a single numerical value to quantify the overall performance of the model.

Mathematically, the RMSE is calculated by taking the square root of the average of the squared differences between predicted and observed values. Let n be the number of observations, y_i be the observed values, and \hat{y}_i be the predicted values. Then the RMSE is computed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

A lower RMSE indicates better agreement between predicted and observed values, implying higher predictive accuracy of the model. Conversely, a higher RMSE suggests greater discrepancy, indicating poorer model performance.

Therefore, I calculate the RMSE values for each proposed configuration to assess not only the individual forecasting abilities of the indexes but also their consistency and robustness across different evaluation contexts, thus providing valuable insights for informed decision-making in economic forecasting⁵.

Enhancing predictive power of Customized and GPT Indexes

In Sections 4.2 and 4.3, I elaborated on the construction of the Customized and GPT Indexes utilizing all available Argentinian sources in *Factiva*. However, it is reasonable to assume that not all sources contain equally relevant content for nowcasting economic activity. Some newspapers may exhibit biases toward topics that are not pertinent to the objectives of this study.

To improve predictive capability, I introduce two additional indexes: a “Weighted” Customized Index and a “Weighted” GPT Index. They assign greater importance to sources that contribute more significantly to accurate forecasting.

These new indexes are constructed following these steps: firstly, I create Customized and GPT Indexes by exhaustively exploring all conceivable combinations of available sources. With six newspapers available, this yields 63 potential combinations, encompassing scenarios like “El Economista” and “Infobae” exclusively, “La Nación”, “El Cronista”, and “La Voz del Interior” solely, or “Infobae” alone, among others⁶. Secondly, I conduct various forecast configurations: I build 60%, 70%, 80%, and 90% cuts for each of the 63 combinations of newspaper sources. I then take the RMSE of each in-sample estimate to build the weights. Those combinations demonstrating superior performance, as indicated by lower RMSE values, are accorded greater weight⁷.

Once the weights are settled, I develop a weighted average prediction leveraging the array of potential source combinations. It is computed as follows: under the proposed forecast combination strategy, I combine 63 models using the weight assigned to each combination given by

$$w_i = \frac{\frac{1}{RMSE_i}}{\sum_{i=1}^N \frac{1}{RMSE_i}}$$

⁵Alternative metrics for assessing predictive model accuracy include the mean absolute error (MAE). While an assessment based on MAE is included in Section D of the Appendix as a control check, the findings remain largely consistent with those derived from the RMSE.

⁶Section B of the Appendix shows the list of all possible combinations of newspaper sources.

⁷This exercise could also be run using the MAE (or any other measure of error) instead of the RMSE. Section D in the Appendix presents the prediction comparison using the MAE as a control check. However, results do not vary significantly.

where $RMSE_i$ is the in-sample RMSE of combination i , w_i is the settled weight and $N = 63$. The aggregation is done with the following expression:

$$\sum_{i=1}^N w_i f_{it} = F_t$$

where f_{it} is the forecast of combination i for the period t . Given that $\sum_{i=1}^N w_i = 1$ (the sum of all weights is 1), F_t is the weighted prediction for the period t .

For example, if the combination of sources “ k ” has a lower RMSE score than the combination “ j ”, then $w_k > w_j$, which means that f_{kt} has a higher impact in F_t .

By weighting the contributions of different newspaper sources based on their individual predictive performance, the weighted indexes aim to leverage the most informative and relevant sources, potentially improving the overall forecasting accuracy and capturing the most valuable signals for nowcasting economic activity. Consequently, the prediction comparison discussed in Section 6 includes a Weighted Customized Index and a Weighted GPT Index.

5.4 Diebold-Mariano Test

The Diebold-Mariano test is a statistical test used to assess the difference in accuracy between two competing forecasts. It was developed by Diebold and Mariano (1995) and is commonly used in the forecasting literature.

The test procedure involves several steps. Firstly, forecast errors are calculated for both models over a specified time period by comparing the predicted values with the actual observed values. These forecast errors serve as the basis for evaluating the performance of each model. Secondly, the mean squared forecast errors (MSFE) of each model are computed. The MSFE represents the average squared difference between the predicted and observed values, providing a measure of the overall forecast accuracy of each model. Next, the difference between the MSFEs of the two models is calculated. This difference serves as the statistic for the Diebold-Mariano test.

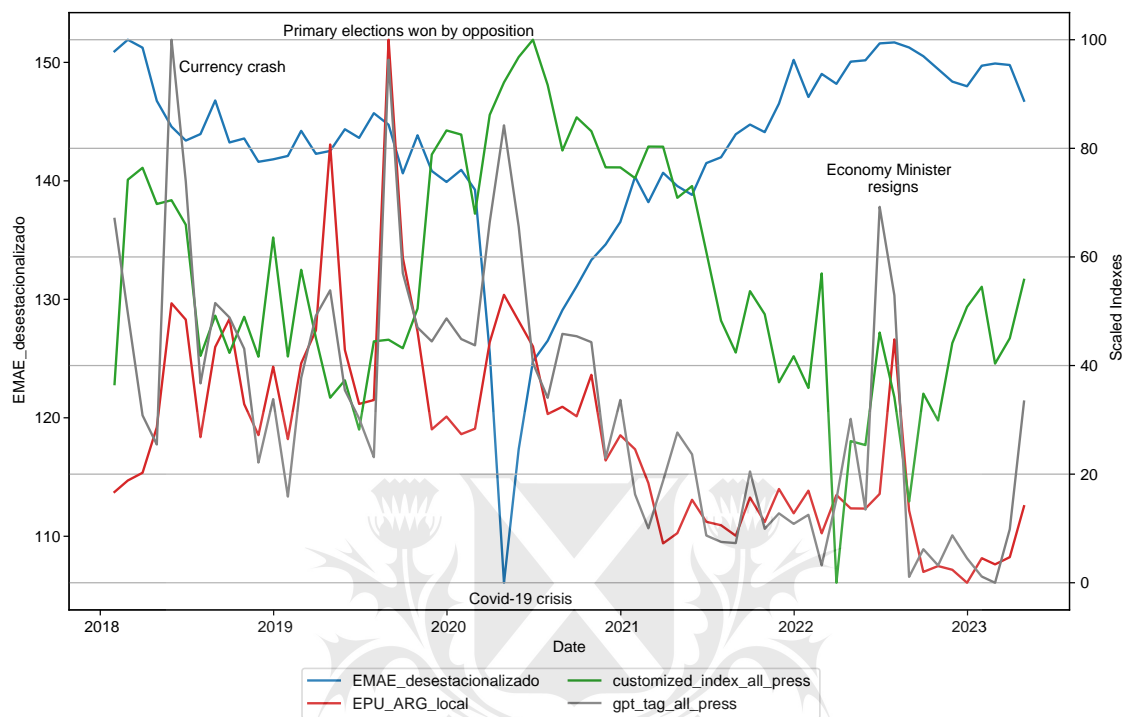
To determine whether the difference in forecast accuracy is statistically significant, the test statistic is compared to a critical value from the standard normal distribution. This critical value is derived based on the degrees of freedom, which depend on the sample size and the forecasting horizon.

The null hypothesis of the Diebold-Mariano test is that there is no difference in forecast accuracy between the two models. If the absolute value of the test statistic exceeds the critical value at a chosen significance level, then the null hypothesis is rejected, indicating that there is a statistically significant difference in forecast accuracy between the models.

I utilize this test to compare each of the proposed indexes against the autoregressive (AR) model with a single lag, which serves as a benchmark.

6 Results

Figure 4: Foreign, local and all press-based EPU indexes



Note: All indexes have been re-scaled from 0 to 100

Figure 4 illustrates the temporal evolution of the proposed indexes alongside the EMAE, which is the chosen proxy for economic activity in Argentina⁸. Visually, there appears to be a negative correlation between the indexes and the EMAE, particularly evident during significant events such as the Covid-19 crisis. However, discerning any superior predictive capacity among the indexes is challenging. To establish a more rigorous assessment, I compare the results of various evaluation exercises outlined in the preceding section.

Table 2: Pearson correlation index (monthly, 01/2018-04/2023)

Index	Contemporaneous EMAE	Anticipated EMAE (1 month ahead)
EPU_ARG_local	-0.38	-0.39
EPU_ARG_foreign	-0.30	-0.30
EPU_ARG_allpress	-0.40	-0.40
customized_index_local_press	-0.70	-0.71
gpt_tag_local_press	-0.38	-0.41

⁸To simplify the visualization, the “All press” and “Foreign” versions of the EPU index have been omitted. I maintain this approach in subsequent figures.

Table 2 presents the correlation coefficients between the different indexes proposed and the EMAE. Correlations are calculated both contemporaneously (i.e., considering the data points released in the same period as the index) and with a 1-month lead (i.e., considering the data points released one month after the index period).

The observed coefficients between all the indexes and the EMAE are negative, indicating a negative relationship between the uncertainty sentiment portrayed in newspapers and the economic activity indicator. This negative association is stronger in the Customized index.

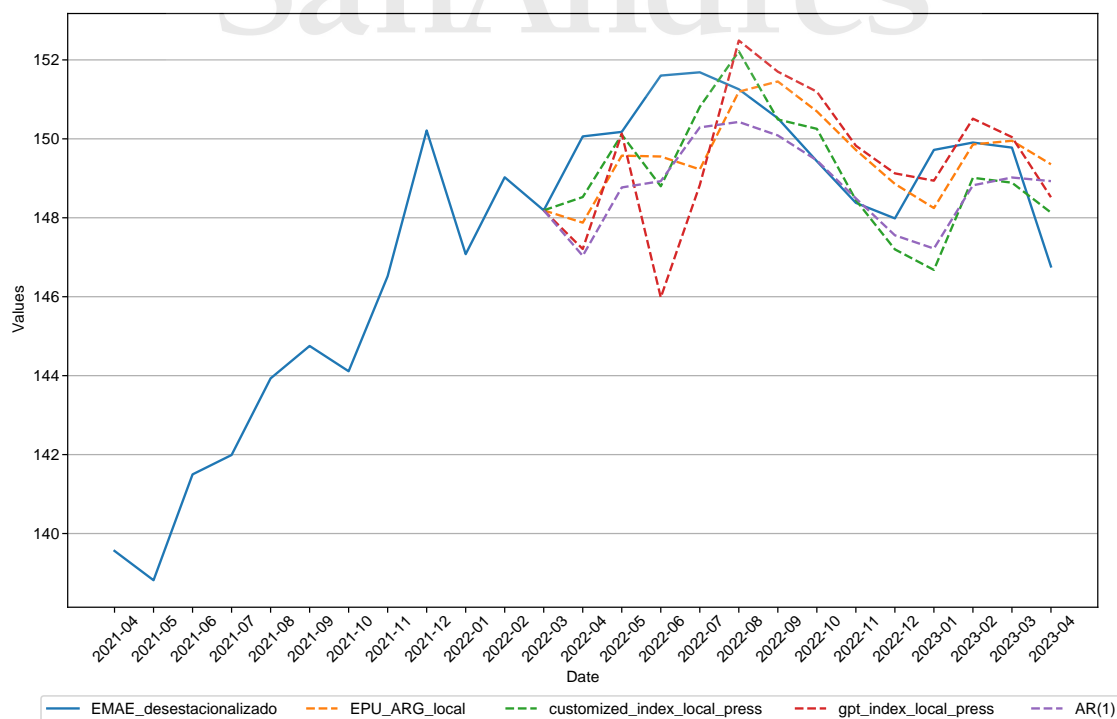
Table 3: Granger Causality Test

Index	Contemporaneous EMAE
EPU_ARG_allpress	0.417
EPU_ARG_local	0.307
EPU_ARG_foreign	0.520
customized_index_local_press	0.047
gpt_tag_local_press	0.180

Table 3 shows the p-value associated to the Granger Causality Test between each index and the EMAE. As mentioned in Section 5.2, the null hypothesis is that an index does not contain information that can help anticipate the dynamic of the economic cycle (EMAE). This hypothesis is rejected for the “Customized” index at the 5% level.

Nowcasting Performance and Diebold-Mariano Test

Figure 5: Nowcast of Economic Activity in Argetina



As a first overview of the nowcasting exercise, figure 5 illustrates the projections generated from the three proposed models and the autoregressive model chosen as benchmark for one of the several windows tested: 80% in-sample data and a recursive window. In this particular case, it seems that the GPT Index shows a higher volatility when making forecasts of the EMAE, whereas the Customized Index seems to be closer to the actual values.

As mentioned in Section 5.3, two additional models are introduced for evaluating the nowcasting performance of the proposed indexes: a Weighted Customized Index and a Weighted GPT Index.

Table 4: Top weights used for weighted models (Sample size: 70%)

Customized Index			GPT Index		
Source Codes	RMSE	Weights	Source Codes	RMSE	Weights
["NACION"]	3.82	1.80%	["GACARG", "INFOB"]	3.77	1.77%
["NACION", "ECONOM"]	4.01	1.71%	["NACION", "GACARG", "INFOB"]	3.84	1.73%
["ECONOM"]	4.23	1.62%	["VOZINT", "GACARG", "INFOB"]	3.87	1.72%
["NACION", "ECONOM", "SABICR", "INFOB"]	4.23	1.62%	["NACION", "VOZINT", "GACARG", "INFOB"]	3.93	1.70%
["ECONOM", "SABICR", "INFOB"]	4.24	1.62%	["ECONOM", "GACARG", "INFOB"]	3.94	1.69%
["NACION", "SABICR", "INFOB"]	4.25	1.61%	["GACARG", "SABICR", "INFOB"]	3.99	1.67%
["SABICR", "INFOB"]	4.25	1.61%	["VOZINT", "GACARG", "SABICR", "INFOB"]	3.99	1.67%
["NACION", "ECONOM", "INFOB"]	4.26	1.61%	["VOZINT", "INFOB"]	3.99	1.67%

Table 4 shows the largest weights for each of the weighted versions of the indexes when the sample size is 70%. They are calculated using the root mean square error (RMSE) of the in-sample sets and apply for the three types of window (fixed, recursive and rolling). The complete list is available in Section B. In this case, it seems that the most relevant combinations of sources of include "NACION" (*La Nación*) and "ECONOM" (*El Economista*) for the Customized Index, whereas "GACARG" (*La Gazeta*) and "INFOB" (*Infobae*) are important in combinations for the GPT Index.

Table 5: Forecasting results measured by RMSE and Diebold-Mariano Test for fixed window

Sample size: 60%			Sample size: 70%		
Model	RMSE	p-value of DM Test	Model	RMSE	p-value of DM Test
AR(1)	2.42		AR(1)	2.35	
EPU_ARG_foreign	2.25	0.068	EPU_ARG_foreign	2.25	0.399
EPU_ARG_local	1.74	0.000	EPU_ARG_local	1.79	0.000
EPU_ARG_allpress	2.09	0.006	EPU_ARG_allpress	2.13	0.117
customized_index_local_press	2.02	0.012	customized_index_local_press	1.97	0.086
gpt_index_local_press	2.04	0.281	gpt_index_local_press	2.12	0.586
customized_weighted_model_rmse	1.72	0.002	customized_weighted_model_rmse	1.62	0.010
gpt_weighted_model_rmse	1.37	0.012	gpt_weighted_model_rmse	1.47	0.201

Sample size: 80%			Sample size: 90%		
Model	RMSE	p-value of DM Test	Model	RMSE	p-value of DM Test
AR(1)	1.76		AR(1)	1.35	
EPU_ARG_foreign	1.68	0.870	EPU_ARG_foreign	1.40	0.862
EPU_ARG_local	1.48	0.220	EPU_ARG_local	1.40	0.526
EPU_ARG_allpress	1.66	0.831	EPU_ARG_allpress	1.38	0.573
customized_index_local_press	1.45	0.240	customized_index_local_press	1.41	0.098
gpt_index_local_press	2.24	0.420	gpt_index_local_press	1.31	0.612
customized_weighted_model_rmse	1.18	0.042	customized_weighted_model_rmse	1.09	0.199
gpt_weighted_model_rmse	1.28	0.869	gpt_weighted_model_rmse	0.85	0.256

Table 5 provides the prediction errors, quantified using the RMSE, and the result of the Diebold-Mariano test when compared with AR(1) models of the proposed out-of-sample fixed window of the nowcasting exercise. The full list is available in Section C from the Appendix⁹.

When evaluating the predictive efficacy of the proposed indexes, it becomes evident that several specifications outperform the benchmark AR(1) model. This observation indicates that the information contained in the indexes holds significance for contemporaneous EMAE forecasting. Notably, the “Weighted” versions of the Customized and GPT Indexes exhibit the lowest RMSE, which highlights their superior performance in nowcasting accuracy.

The ability of the proposed indexes to outperform the AR(1) model emphasizes their potential to provide valuable insights into economic trends and fluctuations. By surpassing the performance of a widely-used benchmark, these indexes demonstrate their capacity to capture nuanced patterns and dynamics in economic data.

Regarding the Diebold-Mariano Test, the varying p-values indicate that the proposed indexes exhibit statistical significance at the 10%, 5% or even 1% level for certain in-sample

⁹Section D in the Appendix shows the results using MAE for comparing prediction errors and the Diebold-Mariano Test using MAE for the “Weighted” Customized and GPT Indexes instead of RMSE.

windows, while not for others. This discrepancy suggests that the uncertainty indexes may offer valuable insights for nowcasting in specific scenarios, while their utility may vary in others.

7 Conclusion

This paper explores the use of unstructured data from Argentinian newspapers to construct economic uncertainty indexes to nowcast the monthly Economic Activity Estimator (EMAE) of Argentina. By leveraging Natural Language processing techniques and Large Language Models, I extract valuable signals from newspaper articles to anticipate movements in this key economic indicator.

The findings suggest that the proposed indexes, especially the weighted versions of the Customized and GPT indexes, exhibit predictive value for forecasting the EMAE. They present evidence of containing a negative association in the correlation analysis and that they could be useful for forecasting in the Granger causality tests. Additionally, most models outperformed the autoregressive benchmark in most of the configurations of training and testing data splits.

However, some limitations persist with the current approach. The analysis could be extended by incorporating more diverse sources beyond just local Argentinian press. Additionally, other large language models such as GloVe could be explored for sentiment analysis.

There is also room to refine the modeling of lags in EMAE publication dates. An underlying assumption of these exercises is that there is only one month lag in the publication of the EMAE. An improved nowcasting exercise should take into consideration that EMAE publications may be lagged more than one month.

Nonetheless, this study highlights the viability of leveraging unstructured textual data to nowcast economic activity. The uncertainty signals extracted from news articles can provide policymakers and businesses with valuable early indicators to support data-driven decision making.

Further research can build on these foundations to develop more sophisticated models, optimize predictive lead times, and establish the robustness of findings across different economic cycles. As availability of textual data expands alongside advances in Natural Language Processing, the opportunities for nowcasting key economic trends from unstructured sources will only grow.

Appendix

A GPT Tagging Prompt

The following prompt was used for tagging the titles in the “GPT Tagging” index:

Eres un economista. Evalúe el nivel de incertidumbre o negatividad de la economía argentina transmitido por cada uno de los siguientes titulares de periódicos argentinos y asigne una puntuación de 0 a 100 a cada título. Utilice una escala en la que 0 representa una incertidumbre muy baja, 100 indica una incertidumbre o negatividad muy alta y valores intermedios como 10, 20, 30, 40, 50, 60, 70, 80 o 90 también son aceptables. Si el título no se vincula a la economía en ningún sentido, evalúe con 0. Evalúe los títulos en función de su contenido y contexto económico para determinar su relevancia para la incertidumbre económica. Busque palabras clave o indicadores de riesgo económico, como “crisis económica”, “devaluación”, “volatilidad”, “cambios en la política gubernamental relacionados con la economía” y “turbulencias financieras”. Quiero que tu respuesta sea únicamente una lista de Python (es decir, entre corchetes), donde cada valor corresponda a la puntuación asignada a cada título. Por ejemplo, si hay tres títulos, la respuesta sería “[0,20,100]”.

Aquí hay algunos ejemplos de títulos que podrían indicar mucha incertidumbre o negatividad económica: “El Banco Central volvió a vender reservas, que perforaron un nuevo piso” “Inversionistas preocupados por la volatilidad de la moneda local.” “Cristina Kirchner anticipó una devaluación y criticó el aumento de las tarifas” “Crisis Cambiaria: Más medidas para controlar al dólar”

Recuerde: sólo interesan noticias de la economía argentina, y no de otros países

Título: “”

B Development of Weighted models

In this section I present the complete array of combinations of *Factiva*’s Argentinian newspapers that are consistently available throughout the period of interest and that are used for the weighted version of the Customized and GPT indexes. The corresponding source codes for these combinations are provided as references:

- El Economista: ECONOM
- La Gaceta: GACARG
- Infobae: INFOB
- La Nación: NACION

- El Cronista: SABICR
- La Voz del Interior: VOZINT

Table 6: Combinations of *Factiva*'s sources applied in weighted models

	Source codes
['NACION']	['ECONOM', 'VOZINT', 'SABICR']
['ECONOM']	['ECONOM', 'VOZINT', 'INFOB']
['VOZINT']	['ECONOM', 'GACARG', 'SABICR']
['GACARG']	['ECONOM', 'GACARG', 'INFOB']
['SABICR']	['ECONOM', 'SABICR', 'INFOB']
['INFOB']	['VOZINT', 'GACARG', 'SABICR']
['NACION', 'ECONOM']	['VOZINT', 'GACARG', 'INFOB']
['NACION', 'VOZINT']	['VOZINT', 'SABICR', 'INFOB']
['NACION', 'GACARG']	['GACARG', 'SABICR', 'INFOB']
['NACION', 'SABICR']	['NACION', 'ECONOM', 'VOZINT', 'GACARG']
['NACION', 'INFOB']	['NACION', 'ECONOM', 'VOZINT', 'SABICR']
['ECONOM', 'VOZINT']	['NACION', 'ECONOM', 'VOZINT', 'INFOB']
['ECONOM', 'GACARG']	['NACION', 'ECONOM', 'GACARG', 'SABICR']
['ECONOM', 'SABICR']	['NACION', 'ECONOM', 'GACARG', 'INFOB']
['ECONOM', 'INFOB']	['NACION', 'ECONOM', 'SABICR', 'INFOB']
['VOZINT', 'GACARG']	['NACION', 'VOZINT', 'GACARG', 'SABICR']
['VOZINT', 'SABICR']	['NACION', 'VOZINT', 'GACARG', 'INFOB']
['VOZINT', 'INFOB']	['NACION', 'VOZINT', 'SABICR', 'INFOB']
['GACARG', 'SABICR']	['NACION', 'GACARG', 'SABICR', 'INFOB']
['GACARG', 'INFOB']	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']
['SABICR', 'INFOB']	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']
['NACION', 'ECONOM', 'VOZINT']	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']
['NACION', 'ECONOM', 'GACARG']	['ECONOM', 'GACARG', 'SABICR', 'INFOB']
['NACION', 'ECONOM', 'SABICR']	['VOZINT', 'GACARG', 'SABICR', 'INFOB']
['NACION', 'ECONOM', 'INFOB']	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']
['NACION', 'VOZINT', 'GACARG']	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']
['NACION', 'VOZINT', 'SABICR']	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']
['NACION', 'VOZINT', 'INFOB']	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']
['NACION', 'GACARG', 'SABICR']	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']
['NACION', 'GACARG', 'INFOB']	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']
['NACION', 'SABICR', 'INFOB']	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']
['ECONOM', 'VOZINT', 'GACARG']	

It should be noted that the last combination (['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']) is equal to the original index.

Tables 7 and 8 contain the weights that are used for the “weighted” version of the Customized and GPT models:

Table 7: Weights used for the Weighted Customized model

Source Codes	Sample Size	RMSE	Weights	Source Codes	Sample Size	RMSE	Weights	Source Codes	Sample Size	RMSE	Weights
[NACION]	60%	4.04	1.80%	[NACION, 'ECONOM, 'VOZINT']	70%	4.33	1.58%	[NACION, 'ECONOM, 'VOZINT, 'SABICR']	80%	4.15	1.58%
[ECONOM]	60%	4.51	1.61%	[NACION, 'ECONOM, 'GACARG']	70%	4.33	1.58%	[NACION, 'ECONOM, 'VOZINT, 'INFOB']	80%	4.07	1.61%
[VOZINT]	60%	4.65	1.57%	[NACION, 'ECONOM, 'SABICR']	70%	4.33	1.58%	[NACION, 'ECONOM, 'GACARG, 'SABICR']	80%	4.17	1.57%
[GACARG]	60%	4.69	1.55%	[NACION, 'ECONOM, 'INFOB']	70%	4.26	1.61%	[NACION, 'ECONOM, 'GACARG, 'INFOB']	80%	4.08	1.60%
[SABICR]	60%	4.68	1.56%	[NACION, 'VOZINT, 'GACARG']	70%	4.39	1.56%	[NACION, 'ECONOM, 'SABICR, 'INFOB']	80%	4.04	1.61%
[INFOB]	60%	4.76	1.53%	[NACION, 'VOZINT, 'SABICR']	70%	4.41	1.55%	[NACION, 'VOZINT, 'GACARG, 'SABICR']	80%	4.20	1.56%
[NACION, 'ECONOM']	60%	4.30	1.70%	[NACION, 'VOZINT, 'INFOB']	70%	4.29	1.60%	[NACION, 'VOZINT, 'GACARG, 'INFOB']	80%	4.09	1.60%
[NACION, 'VOZINT']	60%	4.69	1.55%	[NACION, 'GACARG, 'SABICR']	70%	4.41	1.56%	[NACION, 'VOZINT, 'SABICR, 'INFOB']	80%	4.08	1.60%
[NACION, 'GACARG']	60%	4.67	1.56%	[NACION, 'VOZINT, 'INFOB']	70%	4.29	1.60%	[NACION, 'GACARG, 'SABICR, 'INFOB']	80%	4.09	1.60%
[NACION, 'SABICR']	60%	4.69	1.55%	[NACION, 'SABICR, 'INFOB']	70%	4.25	1.61%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR']	80%	4.20	1.55%
[NACION, 'INFOB']	60%	4.56	1.60%	[ECONOM, 'VOZINT, 'GACARG']	70%	4.40	1.56%	[ECONOM, 'VOZINT, 'GACARG, 'INFOB']	80%	4.10	1.59%
[ECONOM, 'VOZINT']	60%	4.69	1.56%	[ECONOM, 'VOZINT, 'SABICR']	70%	4.41	1.56%	[ECONOM, 'VOZINT, 'SABICR, 'INFOB']	80%	4.08	1.60%
[ECONOM, 'GACARG']	60%	4.67	1.56%	[ECONOM, 'VOZINT, 'INFOB']	70%	4.29	1.60%	[ECONOM, 'GACARG, 'SABICR, 'INFOB']	80%	4.10	1.59%
[ECONOM, 'SABICR']	60%	4.66	1.56%	[ECONOM, 'GACARG, 'SABICR']	70%	4.40	1.56%	[VOZINT, 'GACARG, 'SABICR, 'INFOB']	80%	4.11	1.59%
[ECONOM, 'INFOB']	60%	4.58	1.59%	[ECONOM, 'VOZINT, 'INFOB']	70%	4.29	1.60%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR']	80%	4.17	1.56%
[VOZINT, 'GACARG']	60%	4.69	1.55%	[ECONOM, 'SABICR, 'INFOB']	70%	4.24	1.62%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'INFOB']	80%	4.08	1.60%
[VOZINT, 'SABICR']	60%	4.65	1.57%	[VOZINT, 'GACARG, 'SABICR']	70%	4.41	1.55%	[NACION, 'ECONOM, 'VOZINT, 'SABICR, 'INFOB']	80%	4.06	1.61%
[VOZINT, 'INFOB']	60%	4.58	1.59%	[VOZINT, 'GACARG, 'INFOB']	70%	4.31	1.59%	[NACION, 'ECONOM, 'GACARG, 'SABICR, 'INFOB']	80%	4.08	1.60%
[GACARG, 'SABICR']	60%	4.69	1.55%	[VOZINT, 'SABICR, 'INFOB']	70%	4.29	1.60%	[NACION, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	80%	4.09	1.60%
[GACARG, 'INFOB']	60%	4.57	1.60%	[GACARG, 'SABICR, 'INFOB']	70%	4.30	1.59%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	80%	4.09	1.60%
[SABICR, 'INFOB']	60%	4.49	1.62%	[NACION, 'ECONOM, 'VOZINT, 'GACARG']	70%	4.36	1.57%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	80%	4.08	1.60%
[NACION, 'ECONOM, 'VOZINT']	60%	4.64	1.57%	[NACION, 'ECONOM, 'VOZINT, 'SABICR']	70%	4.38	1.56%	[NACION]	90%	3.44	1.81%
[NACION, 'ECONOM, 'GACARG']	60%	4.62	1.58%	[NACION, 'ECONOM, 'VOZINT, 'INFOB']	70%	4.27	1.61%	[ECONOM]	90%	3.84	1.62%
[NACION, 'ECONOM, 'SABICR']	60%	4.63	1.57%	[NACION, 'ECONOM, 'GACARG, 'SABICR']	70%	4.37	1.57%	[VOZINT]	90%	4.02	1.55%
[NACION, 'ECONOM, 'INFOB']	60%	4.53	1.61%	[NACION, 'ECONOM, 'GACARG, 'INFOB']	70%	4.27	1.61%	[GACARG]	90%	4.01	1.55%
[NACION, 'VOZINT, 'GACARG']	60%	4.69	1.55%	[NACION, 'ECONOM, 'SABICR, 'INFOB']	70%	4.23	1.62%	[SABICR]	90%	4.01	1.55%
[NACION, 'VOZINT, 'SABICR']	60%	4.69	1.55%	[NACION, 'VOZINT, 'GACARG, 'SABICR']	70%	4.41	1.55%	[INFOB]	90%	4.06	1.53%
[NACION, 'VOZINT, 'INFOB']	60%	4.57	1.59%	[NACION, 'VOZINT, 'GACARG, 'INFOB']	70%	4.29	1.60%	[NACION, 'ECONOM']	90%	3.61	1.72%
[NACION, 'GACARG, 'SABICR']	60%	4.69	1.55%	[NACION, 'VOZINT, 'SABICR, 'INFOB']	70%	4.28	1.60%	[NACION, 'VOZINT']	90%	3.95	1.58%
[NACION, 'GACARG, 'INFOB']	60%	4.55	1.60%	[NACION, 'GACARG, 'SABICR, 'INFOB']	70%	4.28	1.60%	[NACION, 'GACARG']	90%	4.00	1.55%
[NACION, 'SABICR, 'INFOB']	60%	4.51	1.62%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR']	70%	4.41	1.56%	[NACION, 'SABICR']	90%	3.98	1.57%
[ECONOM, 'VOZINT, 'GACARG']	60%	4.69	1.55%	[ECONOM, 'VOZINT, 'GACARG, 'INFOB']	70%	4.29	1.60%	[NACION, 'INFOB']	90%	3.89	1.60%
[ECONOM, 'VOZINT, 'SABICR']	60%	4.69	1.55%	[ECONOM, 'VOZINT, 'SABICR, 'INFOB']	70%	4.28	1.60%	[ECONOM, 'VOZINT']	90%	3.98	1.56%
[ECONOM, 'VOZINT, 'INFOB']	60%	4.56	1.60%	[ECONOM, 'GACARG, 'SABICR, 'INFOB']	70%	4.28	1.60%	[ECONOM, 'GACARG']	90%	4.02	1.55%
[ECONOM, 'GACARG, 'SABICR']	60%	4.68	1.56%	[VOZINT, 'GACARG, 'SABICR, 'INFOB']	70%	4.30	1.59%	[ECONOM, 'SABICR']	90%	3.93	1.58%
[ECONOM, 'GACARG, 'INFOB']	60%	4.54	1.61%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR']	70%	4.39	1.56%	[ECONOM, 'INFOB']	90%	3.93	1.58%
[ECONOM, 'SABICR, 'INFOB']	60%	4.48	1.63%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'INFOB']	70%	4.27	1.60%	[VOZINT, 'GACARG']	90%	4.02	1.55%
[VOZINT, 'GACARG, 'SABICR']	60%	4.68	1.56%	[NACION, 'ECONOM, 'VOZINT, 'SABICR, 'INFOB']	70%	4.26	1.61%	[VOZINT, 'SABICR']	90%	4.01	1.55%
[VOZINT, 'GACARG, 'INFOB']	60%	4.57	1.59%	[NACION, 'ECONOM, 'GACARG, 'SABICR, 'INFOB']	70%	4.26	1.61%	[VOZINT, 'INFOB']	90%	3.90	1.59%
[GACARG, 'SABICR, 'INFOB']	60%	4.57	1.60%	[NACION, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	70%	4.29	1.60%	[GACARG, 'SABICR']	90%	4.02	1.55%
[GACARG, 'SABICR, 'INFOB']	60%	4.55	1.60%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	70%	4.29	1.60%	[GACARG, 'INFOB']	90%	3.93	1.58%
[NACION, 'ECONOM, 'VOZINT, 'GACARG']	60%	4.67	1.56%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	70%	4.27	1.60%	[SABICR, 'INFOB']	90%	3.88	1.61%
[NACION, 'ECONOM, 'VOZINT, 'SABICR']	60%	4.68	1.56%	[NACION]	80%	3.62	1.81%	[NACION, 'ECONOM, 'VOZINT']	90%	3.89	1.60%
[NACION, 'ECONOM, 'VOZINT, 'INFOB']	60%	4.55	1.60%	[ECONOM]	80%	4.03	1.62%	[NACION, 'ECONOM, 'GACARG']	90%	3.96	1.57%
[NACION, 'ECONOM, 'GACARG, 'SABICR']	60%	4.66	1.56%	[VOZINT]	80%	4.21	1.55%	[NACION, 'ECONOM, 'SABICR']	90%	3.90	1.60%
[NACION, 'ECONOM, 'GACARG, 'INFOB']	60%	4.52	1.61%	[GACARG]	80%	4.20	1.55%	[NACION, 'ECONOM, 'INFOB']	90%	3.87	1.61%
[NACION, 'ECONOM, 'SABICR, 'INFOB']	60%	4.49	1.62%	[SABICR]	80%	4.20	1.55%	[NACION, 'VOZINT, 'GACARG']	90%	3.99	1.56%
[NACION, 'VOZINT, 'GACARG, 'SABICR']	60%	4.69	1.55%	[INFOB]	80%	4.28	1.53%	[NACION, 'VOZINT, 'SABICR']	90%	3.98	1.56%
[NACION, 'VOZINT, 'GACARG, 'INFOB']	60%	4.56	1.60%	[NACION, 'ECONOM']	80%	3.80	1.72%	[NACION, 'VOZINT, 'INFOB']	90%	3.88	1.60%
[NACION, 'VOZINT, 'SABICR, 'INFOB']	60%	4.56	1.60%	[NACION, 'VOZINT']	80%	4.15	1.58%	[NACION, 'GACARG, 'INFOB']	90%	4.01	1.55%
[NACION, 'GACARG, 'SABICR, 'INFOB']	60%	4.54	1.61%	[NACION, 'GACARG']	80%	4.19	1.56%	[NACION, 'GACARG, 'SABICR']	90%	3.91	1.59%
[ECONOM, 'VOZINT, 'GACARG, 'SABICR']	60%	4.69	1.55%	[NACION, 'SABICR']	80%	4.18	1.56%	[NACION, 'SABICR, 'INFOB']	90%	3.86	1.61%
[ECONOM, 'VOZINT, 'GACARG, 'INFOB']	60%	4.55	1.60%	[NACION, 'INFOB']	80%	4.08	1.60%	[ECONOM, 'VOZINT, 'GACARG']	90%	4.01	1.55%
[ECONOM, 'VOZINT, 'SABICR, 'INFOB']	60%	4.55	1.60%	[ECONOM, 'VOZINT']	80%	4.18	1.56%	[ECONOM, 'VOZINT, 'SABICR']	90%	3.98	1.56%
[ECONOM, 'VOZINT, 'SABICR, 'INFOB']	60%	4.53	1.61%	[ECONOM, 'GACARG']	80%	4.20	1.55%	[ECONOM, 'VOZINT, 'INFOB']	90%	3.89	1.60%
[VOZINT, 'GACARG, 'SABICR, 'INFOB']	60%	4.57	1.59%	[ECONOM, 'SABICR']	80%	4.13	1.58%	[ECONOM, 'GACARG, 'SABICR']	90%	4.01	1.55%
[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR']	60%	4.68	1.56%	[ECONOM, 'INFOB']	80%	4.12	1.58%	[ECONOM, 'GACARG, 'INFOB']	90%	3.92	1.59%
[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'INFOB']	60%	4.54	1.60%	[VOZINT, 'GACARG']	80%	4.21	1.55%	[ECONOM, 'SABICR, 'INFOB']	90%	3.87	1.61%
[NACION, 'ECONOM, 'VOZINT, 'SABICR, 'INFOB']	60%	4.54	1.61%	[VOZINT, 'SABICR']	80%	4.21	1.55%	[VOZINT, 'GACARG, 'SABICR']	90%	4.02	1.55%
[NACION, 'ECONOM, 'GACARG, 'SABICR, 'INFOB']	60%	4.52	1.61%	[VOZINT, 'INFOB']	80%	4.10	1.59%	[VOZINT, 'GACARG, 'INFOB']	90%	3.91	1.59%
[NACION, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	60%	4.56	1.60%	[GACARG, 'SABICR']	80%	4.21	1.55%	[VOZINT, 'SABICR, 'INFOB']	90%	3.89	1.60%
[ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	60%	4.55	1.60%	[GACARG, 'INFOB']	80%	4.12	1.58%	[GACARG, 'SABICR, 'INFOB']	90%	3.92	1.59%
[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	60%	4.54	1.60%	[SABICR, 'INFOB']	80%	4.07	1.60%	[NACION, 'ECONOM, 'VOZINT, 'GACARG']	90%	3.96	1.57%
[NACION]	70%	3.82	1.80%	[NACION, 'ECONOM, 'VOZINT']	80%	4.09	1.60%	[NACION, 'ECONOM, 'VOZINT, 'SABICR']	90%	3.94	1.58%
[ECONOM]	70%	4.23	1.62%	[NACION, 'ECONOM, 'GACARG']	80%	4.14	1.58%	[NACION, 'ECONOM, 'VOZINT, 'INFOB']	90%	3.87	1.61%
[VOZINT]	70%	4.41	1.55%	[NACION, 'ECONOM, 'SABICR']	80%	4.10	1.59%	[NACION, 'ECONOM, 'GACARG, 'SABICR']	90%	3.98	1.56%
[GACARG]	70%	4.41	1.55%	[NACION, 'ECONOM, 'INFOB']	80%	4.07	1.61%	[NACION, 'ECONOM, 'GACARG, 'INFOB']	90%	3.89	1.60%
[SABICR]	70%	4.41	1.55%	[NACION, 'VOZINT, 'GACARG']	80%	4.19	1.56%	[NACION, 'ECONOM, 'SABICR, 'INFOB']	90%	3.85	1.62%
[INFOB]	70%	4.49	1.53%	[NACION, 'VOZINT, 'SABICR']	80%	4.18	1.56%	[NACION, 'VOZINT, 'GACARG, 'SABICR']	90%	4.00	1.56%
[NACION, 'ECONOM']	70%	4.01	1.71%	[NACION, 'VOZINT, 'INFOB']	80%	4.08	1.60%	[NACION, 'VOZINT, 'GACARG, 'INFOB']	90%	3.89	1.60%
[NACION, 'VOZINT']	70%	4.38	1.56%	[NACION, 'GACARG, 'SABICR']	80%	4.20	1.55%	[NACION, 'VOZINT, 'SABICR, 'INFOB']	90%	3.87	1.61%
[NACION, 'GACARG']	70%	4.39	1.56%	[NACION, 'VOZINT, 'INFOB']	80%	4.10	1.59%	[NACION, 'GACARG, 'SABICR, 'INFOB']	90%	3.90	1.60%
[NACION, 'SABICR']	70%	4.40	1.56%	[ECONOM, 'VOZINT, 'GACARG']	80%	4.06	1.61%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR']	90%	4.01	1.55%
[NACION, 'INFOB']	70%	4.29	1.60%	[ECONOM, 'VOZINT, 'SABICR']	80%	4.20	1.55%	[ECONOM, 'VOZINT, 'GACARG, 'INFOB']	90%	3.90	1.60%
[ECONOM, 'VOZINT']	70%	4.39	1.56%	[ECONOM, 'VOZINT, 'INFOB']	80%	4.18	1.56%	[ECONOM, 'VOZINT, 'SABICR, 'INFOB']	90%	3.88	1.60%
[ECONOM, 'GACARG']	70%	4.39	1.56%	[ECONOM, 'GACARG, 'SABICR']	80%	4.09	1.60%	[ECONOM, 'GACARG, 'SABICR, 'INFOB']	90%	3.90	1.59%
[ECONOM, 'SABICR']	70%	4.36	1.57%	[ECONOM, 'GACARG, 'INFOB']	80%	4.20	1.55%	[VOZINT, 'GACARG, 'SABICR, 'INFOB']	90%	3.91	1.59%
[ECONOM, 'INFOB']	70%	4.32	1.59%	[ECONOM, 'SABICR, 'INFOB']	80%	4.11	1.59%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR']	90%	3.97	1.57%
[VOZINT, 'GACARG']	70%	4.41	1.55%	[ECONOM, 'GACARG, 'INFOB']	80%	4.06	1.61%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'INFOB']	90%	3.88	1.60%
[VOZINT, 'SABICR']	70%	4.41	1.55%	[VOZINT, 'GACARG, 'SABICR']	80%	4.21	1.55%	[NACION, 'ECONOM, 'VOZINT, 'SABICR, 'INFOB']	90%	3.86	1.61%
[VOZINT, 'INFOB']	70%	4.31	1.59%	[VOZINT, 'SABICR, 'INFOB']	80%	4.11	1.59%	[NACION, 'ECONOM, 'GACARG, 'SABICR, 'INFOB']	90%	3.88	1.60%
[GACARG, 'SABICR']	70%	4.41	1.55%	[GACARG, 'SABICR, 'INFOB']	80%	4.09	1.60%	[NACION, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	90%	3.89	1.60%
[GACARG, 'INFOB']	70%	4.32	1.59%	[GACARG, 'SABICR, 'INFOB']	80%	4.11	1.59%	[ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	90%	3.89	1.60%
[SABICR, 'INFOB']	70%	4.25	1.61%	[NACION, 'ECONOM, 'VOZINT, 'GACARG']	80%	4.16	1.57%	[NACION, 'ECONOM, 'VOZINT, 'GACARG, 'SABICR, 'INFOB']	90%	3.88	1.60%

Table 8: Weights used for the Weighted GPT model

Source Codes	Sample Size	RMSE	Weights	Source Codes	Sample Size	RMSE	Weights	Source Codes	Sample Size	RMSE	Weights
[NACION]	60%	4.60	1.54%	[NACION, ECONOM, VOZINT]	70%	4.41	1.51%	[NACION, ECONOM, VOZINT, SABICR]	80%	4.16	1.52%
[ECONOM]	60%	4.67	1.52%	[NACION, ECONOM, GACARG]	70%	4.25	1.57%	[NACION, ECONOM, VOZINT, INFOB]	80%	3.96	1.59%
[VOZINT]	60%	4.69	1.52%	[NACION, ECONOM, SABICR]	70%	4.28	1.56%	[NACION, ECONOM, GACARG, SABICR]	80%	3.96	1.60%
[GACARG]	60%	4.50	1.58%	[NACION, ECONOM, INFOB]	70%	4.35	1.53%	[NACION, ECONOM, GACARG, INFOB]	80%	3.78	1.67%
[SABICR]	60%	4.57	1.56%	[NACION, VOZINT, GACARG]	70%	4.29	1.55%	[NACION, ECONOM, SABICR, INFOB]	80%	4.11	1.54%
[INFOB]	60%	4.25	1.67%	[NACION, VOZINT, SABICR]	70%	4.34	1.53%	[NACION, VOZINT, GACARG, SABICR]	80%	4.05	1.56%
[NACION, ECONOM]	60%	4.67	1.52%	[NACION, VOZINT, INFOB]	70%	4.07	1.64%	[NACION, VOZINT, GACARG, INFOB]	80%	3.70	1.71%
[NACION, VOZINT]	60%	4.69	1.52%	[NACION, GACARG, SABICR]	70%	4.12	1.62%	[NACION, VOZINT, SABICR, INFOB]	80%	3.95	1.60%
[NACION, GACARG]	60%	4.52	1.57%	[NACION, GACARG, INFOB]	70%	3.84	1.73%	[NACION, GACARG, SABICR, INFOB]	80%	3.80	1.66%
[NACION, SABICR]	60%	4.48	1.59%	[NACION, SABICR, INFOB]	70%	4.32	1.54%	[ECONOM, VOZINT, GACARG, SABICR]	80%	4.12	1.53%
[NACION, INFOB]	60%	4.51	1.58%	[ECONOM, VOZINT, GACARG]	70%	4.32	1.54%	[ECONOM, VOZINT, GACARG, INFOB]	80%	3.77	1.67%
[ECONOM, VOZINT]	60%	4.69	1.52%	[ECONOM, VOZINT, SABICR]	70%	4.40	1.51%	[ECONOM, VOZINT, SABICR, INFOB]	80%	4.02	1.57%
[ECONOM, GACARG]	60%	4.56	1.56%	[ECONOM, VOZINT, INFOB]	70%	4.19	1.59%	[ECONOM, GACARG, SABICR, INFOB]	80%	3.88	1.63%
[ECONOM, SABICR]	60%	4.67	1.52%	[ECONOM, GACARG, SABICR]	70%	4.18	1.59%	[VOZINT, GACARG, SABICR, INFOB]	80%	3.77	1.68%
[ECONOM, INFOB]	60%	4.69	1.52%	[ECONOM, GACARG, INFOB]	70%	3.94	1.69%	[NACION, ECONOM, VOZINT, GACARG, SABICR]	80%	4.07	1.55%
[VOZINT, GACARG]	60%	4.58	1.55%	[ECONOM, SABICR, INFOB]	70%	4.38	1.52%	[NACION, ECONOM, VOZINT, GACARG, INFOB]	80%	3.81	1.66%
[VOZINT, SABICR]	60%	4.64	1.53%	[VOZINT, GACARG, SABICR]	70%	4.25	1.57%	[NACION, ECONOM, VOZINT, SABICR, INFOB]	80%	4.01	1.58%
[VOZINT, INFOB]	60%	4.26	1.67%	[VOZINT, GACARG, INFOB]	70%	3.87	1.72%	[NACION, ECONOM, GACARG, SABICR, INFOB]	80%	3.89	1.62%
[GACARG, SABICR]	60%	4.40	1.62%	[VOZINT, SABICR, INFOB]	70%	4.18	1.59%	[NACION, VOZINT, GACARG, SABICR, INFOB]	80%	3.79	1.67%
[GACARG, INFOB]	60%	4.00	1.78%	[GACARG, SABICR, INFOB]	70%	3.99	1.67%	[ECONOM, VOZINT, GACARG, SABICR, INFOB]	80%	3.86	1.64%
[SABICR, INFOB]	60%	4.66	1.53%	[NACION, ECONOM, VOZINT, GACARG]	70%	4.31	1.54%	[NACION, ECONOM, VOZINT, GACARG, SABICR, INFOB]	80%	3.87	1.63%
[NACION, ECONOM, VOZINT]	60%	4.69	1.52%	[NACION, ECONOM, VOZINT, SABICR]	70%	4.36	1.53%	[NACION]	90%	3.87	1.57%
[NACION, ECONOM, GACARG]	60%	4.55	1.56%	[NACION, ECONOM, VOZINT, INFOB]	70%	4.20	1.58%	[ECONOM]	90%	4.02	1.51%
[NACION, ECONOM, SABICR]	60%	4.59	1.55%	[NACION, ECONOM, GACARG, SABICR]	70%	4.17	1.60%	[VOZINT]	90%	4.02	1.52%
[NACION, ECONOM, INFOB]	60%	4.65	1.53%	[NACION, ECONOM, GACARG, INFOB]	70%	4.01	1.66%	[GACARG]	90%	3.89	1.57%
[NACION, VOZINT, GACARG]	60%	4.59	1.55%	[NACION, ECONOM, SABICR, INFOB]	70%	4.35	1.53%	[SABICR]	90%	4.02	1.52%
[NACION, VOZINT, SABICR]	60%	4.63	1.54%	[NACION, VOZINT, GACARG, SABICR]	70%	4.23	1.57%	[INFOB]	90%	3.56	1.71%
[NACION, VOZINT, INFOB]	60%	4.35	1.64%	[NACION, VOZINT, GACARG, INFOB]	70%	3.93	1.70%	[NACION, ECONOM]	90%	3.92	1.55%
[NACION, GACARG, SABICR]	60%	4.40	1.62%	[NACION, VOZINT, SABICR, INFOB]	70%	4.19	1.59%	[NACION, VOZINT]	90%	4.01	1.52%
[NACION, GACARG, INFOB]	60%	4.09	1.74%	[NACION, GACARG, SABICR, INFOB]	70%	4.02	1.65%	[NACION, GACARG]	90%	3.88	1.57%
[NACION, SABICR, INFOB]	60%	4.62	1.54%	[ECONOM, VOZINT, GACARG, SABICR]	70%	4.28	1.56%	[NACION, SABICR]	90%	3.86	1.58%
[ECONOM, VOZINT, GACARG]	60%	4.62	1.54%	[ECONOM, VOZINT, GACARG, INFOB]	70%	4.00	1.66%	[NACION, INFOB]	90%	3.78	1.61%
[ECONOM, VOZINT, SABICR]	60%	4.67	1.52%	[ECONOM, VOZINT, SABICR, INFOB]	70%	4.26	1.56%	[ECONOM, VOZINT]	90%	4.00	1.52%
[ECONOM, VOZINT, INFOB]	60%	4.49	1.58%	[ECONOM, GACARG, SABICR, INFOB]	70%	4.11	1.62%	[ECONOM, GACARG]	90%	3.91	1.56%
[ECONOM, GACARG, SABICR]	60%	4.48	1.59%	[VOZINT, GACARG, SABICR, INFOB]	70%	3.99	1.67%	[ECONOM, SABICR]	90%	4.02	1.52%
[ECONOM, GACARG, INFOB]	60%	4.18	1.70%	[NACION, ECONOM, VOZINT, GACARG, SABICR]	70%	4.26	1.56%	[ECONOM, INFOB]	90%	3.99	1.53%
[ECONOM, SABICR, INFOB]	60%	4.67	1.52%	[NACION, ECONOM, VOZINT, GACARG, INFOB]	70%	4.04	1.65%	[VOZINT, GACARG]	90%	3.95	1.54%
[VOZINT, GACARG, SABICR]	60%	4.52	1.57%	[NACION, ECONOM, VOZINT, SABICR, INFOB]	70%	4.25	1.57%	[VOZINT, SABICR]	90%	4.02	1.52%
[VOZINT, GACARG, INFOB]	60%	4.12	1.73%	[NACION, ECONOM, GACARG, SABICR, INFOB]	70%	4.13	1.61%	[VOZINT, INFOB]	90%	3.59	1.69%
[VOZINT, SABICR, INFOB]	60%	4.48	1.59%	[NACION, VOZINT, GACARG, SABICR, INFOB]	70%	4.02	1.66%	[GACARG, SABICR]	90%	3.96	1.54%
[GACARG, SABICR, INFOB]	60%	4.26	1.67%	[ECONOM, VOZINT, GACARG, SABICR, INFOB]	70%	4.09	1.63%	[GACARG, INFOB]	90%	3.40	1.79%
[NACION, ECONOM, VOZINT, GACARG]	60%	4.62	1.54%	[NACION, ECONOM, VOZINT, GACARG, SABICR, INFOB]	70%	4.10	1.62%	[SABICR, INFOB]	90%	3.94	1.55%
[NACION, ECONOM, VOZINT, SABICR]	60%	4.66	1.53%	[NACION]	80%	4.06	1.56%	[NACION, ECONOM, VOZINT]	90%	4.02	1.52%
[NACION, ECONOM, VOZINT, INFOB]	60%	4.50	1.58%	[ECONOM]	80%	4.21	1.50%	[NACION, ECONOM, GACARG]	90%	3.87	1.57%
[NACION, ECONOM, GACARG, SABICR]	60%	4.47	1.59%	[VOZINT]	80%	4.21	1.50%	[NACION, ECONOM, SABICR]	90%	3.91	1.56%
[NACION, ECONOM, GACARG, INFOB]	60%	4.27	1.67%	[GACARG]	80%	4.04	1.56%	[NACION, ECONOM, INFOB]	90%	3.91	1.56%
[NACION, ECONOM, SABICR, INFOB]	60%	4.65	1.53%	[SABICR]	80%	4.19	1.51%	[NACION, VOZINT, GACARG]	90%	3.94	1.54%
[NACION, VOZINT, GACARG, SABICR]	60%	4.52	1.57%	[INFOB]	80%	3.75	1.68%	[NACION, VOZINT, SABICR]	90%	4.01	1.52%
[NACION, VOZINT, GACARG, INFOB]	60%	4.19	1.70%	[NACION, ECONOM]	80%	4.11	1.54%	[NACION, VOZINT, INFOB]	90%	3.66	1.66%
[NACION, VOZINT, SABICR, INFOB]	60%	4.48	1.59%	[NACION, VOZINT]	80%	4.19	1.51%	[NACION, GACARG, SABICR]	90%	3.91	1.56%
[NACION, GACARG, SABICR, INFOB]	60%	4.30	1.65%	[NACION, GACARG]	80%	4.04	1.65%	[NACION, GACARG, INFOB]	90%	3.46	1.76%
[ECONOM, VOZINT, GACARG, SABICR]	60%	4.57	1.56%	[NACION, SABICR]	80%	3.96	1.60%	[NACION, SABICR, INFOB]	90%	3.90	1.56%
[ECONOM, VOZINT, GACARG, INFOB]	60%	4.27	1.67%	[ECONOM, VOZINT]	80%	3.97	1.59%	[ECONOM, VOZINT, GACARG]	90%	3.98	1.53%
[ECONOM, VOZINT, SABICR, INFOB]	60%	4.56	1.56%	[ECONOM, GACARG]	80%	4.20	1.50%	[ECONOM, VOZINT, SABICR]	90%	4.02	1.52%
[ECONOM, GACARG, SABICR, INFOB]	60%	4.40	1.61%	[NACION, VOZINT, GACARG]	80%	4.05	1.56%	[ECONOM, VOZINT, INFOB]	90%	3.77	1.62%
[VOZINT, GACARG, SABICR, INFOB]	60%	4.26	1.67%	[ECONOM, SABICR]	80%	4.19	1.51%	[ECONOM, GACARG, SABICR]	90%	3.97	1.54%
[NACION, ECONOM, VOZINT, GACARG, SABICR]	60%	4.56	1.56%	[ECONOM, INFOB]	80%	4.19	1.51%	[ECONOM, GACARG, INFOB]	90%	3.55	1.71%
[NACION, ECONOM, VOZINT, GACARG, INFOB]	60%	4.31	1.65%	[VOZINT, GACARG]	80%	4.10	1.54%	[ECONOM, SABICR, INFOB]	90%	3.96	1.54%
[NACION, ECONOM, VOZINT, SABICR, INFOB]	60%	4.56	1.56%	[VOZINT, SABICR]	80%	4.21	1.50%	[VOZINT, GACARG, SABICR]	90%	4.01	1.52%
[NACION, ECONOM, GACARG, SABICR, INFOB]	60%	4.42	1.61%	[VOZINT, INFOB]	80%	3.77	1.68%	[VOZINT, GACARG, INFOB]	90%	3.49	1.75%
[NACION, VOZINT, GACARG, SABICR, INFOB]	60%	4.30	1.65%	[GACARG, SABICR]	80%	4.00	1.58%	[VOZINT, SABICR, INFOB]	90%	3.82	1.60%
[ECONOM, VOZINT, GACARG, SABICR, INFOB]	60%	4.37	1.63%	[GACARG, INFOB]	80%	3.55	1.78%	[GACARG, SABICR, INFOB]	90%	3.69	1.65%
[NACION, ECONOM, VOZINT, GACARG, SABICR, INFOB]	60%	4.39	1.62%	[SABICR, INFOB]	80%	4.13	1.53%	[NACION, ECONOM, VOZINT, GACARG]	90%	3.96	1.54%
[NACION]	70%	4.30	1.55%	[NACION, ECONOM, VOZINT]	80%	4.19	1.51%	[NACION, ECONOM, VOZINT, SABICR]	90%	4.01	1.52%
[ECONOM]	70%	4.41	1.51%	[NACION, ECONOM, GACARG]	80%	4.02	1.57%	[NACION, ECONOM, VOZINT, INFOB]	90%	3.78	1.61%
[VOZINT]	70%	4.41	1.51%	[NACION, ECONOM, SABICR]	80%	4.04	1.56%	[NACION, ECONOM, GACARG, SABICR]	90%	3.92	1.55%
[GACARG]	70%	4.23	1.57%	[NACION, ECONOM, INFOB]	80%	4.11	1.54%	[NACION, ECONOM, GACARG, INFOB]	90%	3.61	1.69%
[SABICR]	70%	4.34	1.53%	[NACION, VOZINT, GACARG]	80%	4.10	1.54%	[NACION, ECONOM, SABICR, INFOB]	90%	3.93	1.55%
[INFOB]	70%	4.00	1.67%	[NACION, VOZINT, SABICR]	80%	4.15	1.52%	[NACION, VOZINT, GACARG, SABICR]	90%	3.99	1.53%
[NACION, ECONOM]	70%	4.36	1.53%	[NACION, VOZINT, INFOB]	80%	3.84	1.65%	[NACION, VOZINT, GACARG, INFOB]	90%	3.54	1.72%
[NACION, VOZINT]	70%	4.40	1.51%	[NACION, GACARG, SABICR]	80%	3.93	1.61%	[NACION, VOZINT, SABICR, INFOB]	90%	3.82	1.60%
[NACION, GACARG]	70%	4.24	1.57%	[NACION, GACARG, INFOB]	80%	3.62	1.75%	[NACION, GACARG, SABICR, INFOB]	90%	3.70	1.64%
[NACION, SABICR]	70%	4.18	1.59%	[NACION, SABICR, INFOB]	80%	4.08	1.55%	[ECONOM, VOZINT, GACARG, SABICR]	90%	4.01	1.52%
[NACION, INFOB]	70%	4.21	1.58%	[ECONOM, VOZINT, GACARG]	80%	4.13	1.53%	[ECONOM, VOZINT, GACARG, INFOB]	90%	3.61	1.69%
[ECONOM, VOZINT]	70%	4.41	1.51%	[VOZINT, GACARG]	80%	4.21	1.50%	[ECONOM, VOZINT, SABICR, INFOB]	90%	3.87	1.58%
[ECONOM, GACARG]	70%	4.25	1.57%	[ECONOM, VOZINT, INFOB]	80%	3.95	1.60%	[ECONOM, GACARG, SABICR, INFOB]	90%	3.76	1.62%
[ECONOM, SABICR]	70%	4.38	1.52%	[ECONOM, GACARG, SABICR]	80%	4.02	1.57%	[VOZINT, GACARG, SABICR, INFOB]	90%	3.71	1.64%
[ECONOM, INFOB]	70%	4.41	1.51%	[ECONOM, GACARG, INFOB]	80%	3.72	1.70%	[NACION, ECONOM, VOZINT, GACARG, SABICR]	90%	3.99	1.53%
[VOZINT, GACARG]	70%	4.29	1.55%	[ECONOM, SABICR, INFOB]	80%	4.15	1.52%	[NACION, ECONOM, VOZINT, GACARG, INFOB]	90%	3.64	1.67%
[VOZINT, SABICR]	70%	4.39	1.52%	[VOZINT, GACARG, SABICR]	80%	4.10	1.54%	[NACION, ECONOM, VOZINT, SABICR, INFOB]	90%	3.86	1.58%
[VOZINT, INFOB]	70%	3.99	1.67%	[VOZINT, GACARG, INFOB]	80%	3.65	1.73%	[NACION, ECONOM, VOZINT, GACARG, SABICR, INFOB]	90%	3.77	1.62%
[GACARG, SABICR]	70%	4.13	1.61%	[VOZINT, SABICR, INFOB]	80%	3.95	1.60%	[NACION, VOZINT, GACARG, SABICR, INFOB]	90%	3.72	1.64%
[GACARG, INFOB]	70%	3.77	1.77%	[GACARG, SABICR, INFOB]	80%	3.76	1.68%	[ECONOM, VOZINT, GACARG, SABICR, INFOB]	90%	3.76	1.62%
[SABICR, INFOB]	70%	4.36	1.53%	[NACION, ECONOM, VOZINT, GACARG]	80%	4.10	1.54%	[NACION, ECONOM, VOZINT, GACARG, SABICR, INFOB]	90%	3.77	1.62%

C Diebold-Mariano Test: Complete list

Table 9: RMSE and Diebold-Mariano Test for all windows proposed

Window Type	In-sample size	Model	RMSE	p-value of DM Test	Window Type	In-sample size	Model	RMSE	p-value of DM Test
fixed	60%	AR(1)	2.42		fixed	60%	gpt_weighted_model_rmse	1.37	0.012
fixed	70%	AR(1)	2.35		fixed	60%	gpt_index_local_press	2.04	0.281
fixed	80%	AR(1)	1.76		fixed	70%	gpt_weighted_model_rmse	1.47	0.201
fixed	90%	AR(1)	1.35		fixed	70%	gpt_index_local_press	2.12	0.586
recursive	60%	AR(1)	2.03		fixed	80%	gpt_weighted_model_rmse	1.28	0.869
recursive	70%	AR(1)	2.04		fixed	80%	gpt_index_local_press	2.24	0.420
recursive	80%	AR(1)	1.62		fixed	90%	gpt_weighted_model_rmse	0.85	0.256
recursive	90%	AR(1)	1.36		fixed	90%	gpt_index_local_press	1.31	0.612
rolling	60%	AR(1)	2.07		recursive	60%	gpt_weighted_model_rmse	1.37	0.230
rolling	70%	AR(1)	2.06		recursive	60%	gpt_index_local_press	1.95	0.991
rolling	80%	AR(1)	1.64		recursive	70%	gpt_weighted_model_rmse	1.43	0.536
rolling	90%	AR(1)	1.40		recursive	70%	gpt_index_local_press	2.10	0.801
fixed	60%	EPU_ARG_allpress	2.09	0.006	recursive	80%	gpt_weighted_model_rmse	1.26	0.592
fixed	60%	EPU_ARG_local	1.74	0.000	recursive	80%	gpt_index_local_press	2.17	0.282
fixed	60%	EPU_ARG_foreign	2.25	0.068	recursive	90%	gpt_weighted_model_rmse	0.84	0.254
fixed	70%	EPU_ARG_allpress	2.13	0.117	recursive	90%	gpt_index_local_press	1.23	0.539
fixed	70%	EPU_ARG_local	1.79	0.000	rolling	60%	gpt_weighted_model_rmse	1.43	0.792
fixed	70%	EPU_ARG_foreign	2.25	0.399	rolling	60%	gpt_index_local_press	2.38	0.548
fixed	80%	EPU_ARG_allpress	1.66	0.831	rolling	70%	gpt_weighted_model_rmse	1.57	0.901
fixed	80%	EPU_ARG_local	1.48	0.220	rolling	70%	gpt_index_local_press	2.54	0.454
fixed	80%	EPU_ARG_foreign	1.68	0.870	rolling	80%	gpt_weighted_model_rmse	1.32	0.474
fixed	90%	EPU_ARG_allpress	1.38	0.573	rolling	80%	gpt_index_local_press	2.28	0.225
fixed	90%	EPU_ARG_local	1.40	0.526	rolling	90%	gpt_weighted_model_rmse	0.81	0.248
fixed	90%	EPU_ARG_foreign	1.40	0.862	rolling	90%	gpt_index_local_press	1.18	0.523
recursive	60%	EPU_ARG_allpress	1.94	0.403	fixed	60%	customized_weighted_model_rmse	1.72	0.002
recursive	60%	EPU_ARG_local	1.69	0.040	fixed	60%	customized_index_local_press	2.02	0.012
recursive	60%	EPU_ARG_foreign	2.03	0.669	fixed	70%	customized_weighted_model_rmse	1.62	0.010
recursive	70%	EPU_ARG_allpress	2.00	0.758	fixed	70%	customized_index_local_press	1.97	0.086
recursive	70%	EPU_ARG_local	1.76	0.065	fixed	80%	customized_weighted_model_rmse	1.18	0.042
recursive	70%	EPU_ARG_foreign	2.06	0.933	fixed	80%	customized_index_local_press	1.45	0.240
recursive	80%	EPU_ARG_allpress	1.62	0.583	fixed	90%	customized_weighted_model_rmse	1.09	0.199
recursive	80%	EPU_ARG_local	1.50	0.754	fixed	90%	customized_index_local_press	1.41	0.098
recursive	80%	EPU_ARG_foreign	1.61	0.585	recursive	60%	customized_weighted_model_rmse	1.46	0.109
recursive	90%	EPU_ARG_allpress	1.39	0.702	recursive	60%	customized_index_local_press	1.94	0.511
recursive	90%	EPU_ARG_local	1.37	0.508	recursive	70%	customized_weighted_model_rmse	1.47	0.300
recursive	90%	EPU_ARG_foreign	1.41	0.969	recursive	70%	customized_index_local_press	1.95	0.865
rolling	60%	EPU_ARG_allpress	2.07	0.982	recursive	80%	customized_weighted_model_rmse	1.09	0.452
rolling	60%	EPU_ARG_local	1.83	0.426	recursive	80%	customized_index_local_press	1.41	0.637
rolling	60%	EPU_ARG_foreign	2.16	0.681	recursive	90%	customized_weighted_model_rmse	1.09	0.164
rolling	70%	EPU_ARG_allpress	2.08	0.882	recursive	90%	customized_index_local_press	1.41	0.108
rolling	70%	EPU_ARG_local	1.79	0.116	rolling	60%	customized_weighted_model_rmse	1.58	0.725
rolling	70%	EPU_ARG_foreign	2.14	0.599	rolling	60%	customized_index_local_press	2.19	0.724
rolling	80%	EPU_ARG_allpress	1.67	0.414	rolling	70%	customized_weighted_model_rmse	1.60	0.982
rolling	80%	EPU_ARG_local	1.53	0.839	rolling	70%	customized_index_local_press	2.22	0.542
rolling	80%	EPU_ARG_foreign	1.66	0.377	rolling	80%	customized_weighted_model_rmse	1.17	0.719
rolling	90%	EPU_ARG_allpress	1.42	0.645	rolling	80%	customized_index_local_press	1.58	0.213
rolling	90%	EPU_ARG_local	1.38	0.490	rolling	90%	customized_weighted_model_rmse	1.14	0.039
rolling	90%	EPU_ARG_foreign	1.44	0.975	rolling	90%	customized_index_local_press	1.58	0.051

D Results using MAE instead of RMSE

Table 10: Weights used for the Weighted Customized model (MAE version)

Source Codes	Sample Size	MAE	Weights	Source Codes	Sample Size	MAE	Weights	Source Codes	Sample Size	MAE	Weights
['NACION']	60%	2.75	1.80%	['NACION', 'ECONOM', 'VOZINT']	60%	2.63	1.57%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	60%	2.63	1.56%
['NACION']	70%	2.50	1.80%	['NACION', 'ECONOM', 'VOZINT']	70%	2.41	1.58%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	70%	2.42	1.56%
['NACION']	80%	2.37	1.81%	['NACION', 'ECONOM', 'VOZINT']	80%	2.30	1.60%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	80%	2.33	1.58%
['NACION']	90%	2.22	1.81%	['NACION', 'ECONOM', 'VOZINT']	90%	2.16	1.60%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	90%	2.18	1.58%
['ECONOM']	60%	2.71	1.61%	['NACION', 'ECONOM', 'GACARG']	60%	2.69	1.58%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	60%	2.67	1.60%
['ECONOM']	70%	2.51	1.62%	['NACION', 'ECONOM', 'GACARG']	70%	2.47	1.58%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	70%	2.44	1.61%
['ECONOM']	80%	2.46	1.62%	['NACION', 'ECONOM', 'GACARG']	80%	2.43	1.58%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	80%	2.37	1.61%
['ECONOM']	90%	2.33	1.62%	['NACION', 'ECONOM', 'GACARG']	90%	2.33	1.57%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	90%	2.23	1.61%
['VOZINT']	60%	2.78	1.57%	['NACION', 'ECONOM', 'SABICR']	60%	2.67	1.57%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	60%	2.66	1.56%
['VOZINT']	70%	2.52	1.55%	['NACION', 'ECONOM', 'SABICR']	70%	2.47	1.58%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	70%	2.43	1.57%
['VOZINT']	80%	2.41	1.55%	['NACION', 'ECONOM', 'SABICR']	80%	2.39	1.59%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	80%	2.40	1.57%
['VOZINT']	90%	2.31	1.55%	['NACION', 'ECONOM', 'SABICR']	90%	2.24	1.60%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	90%	2.30	1.56%
['GACARG']	60%	2.66	1.55%	['NACION', 'ECONOM', 'INFOB']	60%	2.64	1.61%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	60%	2.66	1.61%
['GACARG']	70%	2.48	1.55%	['NACION', 'ECONOM', 'INFOB']	70%	2.42	1.61%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	70%	2.46	1.61%
['GACARG']	80%	2.39	1.55%	['NACION', 'ECONOM', 'INFOB']	80%	2.37	1.61%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	80%	2.41	1.60%
['GACARG']	90%	2.28	1.55%	['NACION', 'ECONOM', 'INFOB']	90%	2.24	1.61%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	90%	2.30	1.60%
['SABICR']	60%	2.71	1.56%	['NACION', 'VOZINT', 'GACARG']	60%	2.65	1.55%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	60%	2.65	1.62%
['SABICR']	70%	2.49	1.55%	['NACION', 'VOZINT', 'GACARG']	70%	2.42	1.56%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	70%	2.43	1.62%
['SABICR']	80%	2.37	1.55%	['NACION', 'VOZINT', 'GACARG']	80%	2.37	1.56%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	80%	2.38	1.61%
['SABICR']	90%	2.25	1.55%	['NACION', 'VOZINT', 'GACARG']	90%	2.27	1.56%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	90%	2.24	1.62%
['INFOB']	60%	2.92	1.53%	['NACION', 'VOZINT', 'SABICR']	60%	2.69	1.55%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	60%	2.66	1.55%
['INFOB']	70%	2.63	1.53%	['NACION', 'VOZINT', 'SABICR']	70%	2.44	1.55%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	70%	2.44	1.55%
['INFOB']	80%	2.54	1.53%	['NACION', 'VOZINT', 'SABICR']	80%	2.33	1.56%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	80%	2.37	1.56%
['INFOB']	90%	2.38	1.53%	['NACION', 'VOZINT', 'SABICR']	90%	2.20	1.56%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	90%	2.25	1.56%
['NACION', 'ECONOM']	60%	2.80	1.70%	['NACION', 'VOZINT', 'INFOB']	60%	2.65	1.59%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	60%	2.64	1.60%
['NACION', 'ECONOM']	70%	2.54	1.71%	['NACION', 'VOZINT', 'INFOB']	70%	2.43	1.60%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	70%	2.41	1.60%
['NACION', 'ECONOM']	80%	2.47	1.72%	['NACION', 'VOZINT', 'INFOB']	80%	2.37	1.60%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	80%	2.37	1.60%
['NACION', 'ECONOM']	90%	2.32	1.72%	['NACION', 'VOZINT', 'INFOB']	90%	2.22	1.60%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	90%	2.24	1.60%
['NACION', 'VOZINT']	60%	2.64	1.55%	['NACION', 'GACARG', 'SABICR']	60%	2.65	1.55%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	60%	2.65	1.60%
['NACION', 'VOZINT']	70%	2.39	1.56%	['NACION', 'GACARG', 'SABICR']	70%	2.45	1.56%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	70%	2.42	1.60%
['NACION', 'VOZINT']	80%	2.29	1.58%	['NACION', 'GACARG', 'SABICR']	80%	2.39	1.55%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	80%	2.37	1.60%
['NACION', 'VOZINT']	90%	2.18	1.58%	['NACION', 'GACARG', 'SABICR']	90%	2.29	1.55%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	90%	2.21	1.61%
['NACION', 'GACARG']	60%	2.67	1.56%	['NACION', 'GACARG', 'INFOB']	60%	2.68	1.60%	['NACION', 'GACARG', 'SABICR', 'INFOB']	60%	2.65	1.61%
['NACION', 'GACARG']	70%	2.46	1.56%	['NACION', 'GACARG', 'INFOB']	70%	2.47	1.60%	['NACION', 'GACARG', 'SABICR', 'INFOB']	70%	2.45	1.60%
['NACION', 'GACARG']	80%	2.40	1.56%	['NACION', 'GACARG', 'INFOB']	80%	2.42	1.59%	['NACION', 'GACARG', 'SABICR', 'INFOB']	80%	2.41	1.60%
['NACION', 'GACARG']	90%	2.31	1.55%	['NACION', 'GACARG', 'INFOB']	90%	2.31	1.59%	['NACION', 'GACARG', 'SABICR', 'INFOB']	90%	2.29	1.60%
['NACION', 'SABICR']	60%	2.64	1.55%	['NACION', 'SABICR', 'INFOB']	60%	2.65	1.62%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	60%	2.65	1.55%
['NACION', 'SABICR']	70%	2.45	1.56%	['NACION', 'SABICR', 'INFOB']	70%	2.42	1.61%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	70%	2.43	1.56%
['NACION', 'SABICR']	80%	2.38	1.56%	['NACION', 'SABICR', 'INFOB']	80%	2.37	1.61%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	80%	2.38	1.55%
['NACION', 'SABICR']	90%	2.25	1.57%	['NACION', 'SABICR', 'INFOB']	90%	2.24	1.61%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	90%	2.26	1.55%
['NACION', 'INFOB']	60%	2.64	1.60%	['ECONOM', 'VOZINT', 'GACARG']	60%	2.65	1.55%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	60%	2.65	1.60%
['NACION', 'INFOB']	70%	2.42	1.60%	['ECONOM', 'VOZINT', 'GACARG']	70%	2.42	1.56%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	70%	2.42	1.60%
['NACION', 'INFOB']	80%	2.36	1.60%	['ECONOM', 'VOZINT', 'GACARG']	80%	2.38	1.55%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	80%	2.36	1.59%
['NACION', 'INFOB']	90%	2.23	1.60%	['ECONOM', 'VOZINT', 'GACARG']	90%	2.28	1.55%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	90%	2.24	1.60%
['ECONOM', 'VOZINT']	60%	2.64	1.56%	['ECONOM', 'VOZINT', 'SABICR']	60%	2.66	1.55%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	60%	2.66	1.60%
['ECONOM', 'VOZINT']	70%	2.40	1.56%	['ECONOM', 'VOZINT', 'SABICR']	70%	2.44	1.56%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	70%	2.43	1.60%
['ECONOM', 'VOZINT']	80%	2.34	1.56%	['ECONOM', 'VOZINT', 'SABICR']	80%	2.34	1.56%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	80%	2.37	1.60%
['ECONOM', 'VOZINT']	90%	2.21	1.56%	['ECONOM', 'VOZINT', 'SABICR']	90%	2.19	1.56%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	90%	2.22	1.60%
['ECONOM', 'GACARG']	60%	2.66	1.56%	['ECONOM', 'VOZINT', 'INFOB']	60%	2.67	1.60%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	60%	2.63	1.61%
['ECONOM', 'GACARG']	70%	2.45	1.56%	['ECONOM', 'VOZINT', 'INFOB']	70%	2.44	1.60%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	70%	2.45	1.60%
['ECONOM', 'GACARG']	80%	2.40	1.55%	['ECONOM', 'VOZINT', 'INFOB']	80%	2.38	1.60%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	80%	2.40	1.59%
['ECONOM', 'GACARG']	90%	2.30	1.55%	['ECONOM', 'VOZINT', 'INFOB']	90%	2.23	1.60%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	90%	2.28	1.59%
['ECONOM', 'SABICR']	60%	2.66	1.56%	['ECONOM', 'GACARG', 'SABICR']	60%	2.65	1.56%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.63	1.59%
['ECONOM', 'SABICR']	70%	2.48	1.57%	['ECONOM', 'GACARG', 'SABICR']	70%	2.44	1.56%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.41	1.59%
['ECONOM', 'SABICR']	80%	2.40	1.58%	['ECONOM', 'GACARG', 'SABICR']	80%	2.39	1.55%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.37	1.59%
['ECONOM', 'SABICR']	90%	2.25	1.58%	['ECONOM', 'GACARG', 'SABICR']	90%	2.29	1.55%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.23	1.59%
['ECONOM', 'INFOB']	60%	2.63	1.59%	['ECONOM', 'GACARG', 'INFOB']	60%	2.66	1.61%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	60%	2.64	1.56%
['ECONOM', 'INFOB']	70%	2.44	1.59%	['ECONOM', 'GACARG', 'INFOB']	70%	2.47	1.60%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	70%	2.42	1.56%
['ECONOM', 'INFOB']	80%	2.39	1.58%	['ECONOM', 'GACARG', 'INFOB']	80%	2.42	1.59%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	80%	2.37	1.56%
['ECONOM', 'INFOB']	90%	2.26	1.58%	['ECONOM', 'GACARG', 'INFOB']	90%	2.31	1.59%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	90%	2.25	1.57%
['VOZINT', 'GACARG']	60%	2.67	1.55%	['ECONOM', 'SABICR', 'INFOB']	60%	2.64	1.63%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	60%	2.66	1.60%
['VOZINT', 'GACARG']	70%	2.47	1.55%	['ECONOM', 'SABICR', 'INFOB']	70%	2.44	1.62%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	70%	2.42	1.60%
['VOZINT', 'GACARG']	80%	2.40	1.55%	['ECONOM', 'SABICR', 'INFOB']	80%	2.38	1.61%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	80%	2.36	1.60%
['VOZINT', 'GACARG']	90%	2.29	1.55%	['ECONOM', 'SABICR', 'INFOB']	90%	2.24	1.61%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	90%	2.23	1.60%
['VOZINT', 'SABICR']	60%	2.76	1.57%	['VOZINT', 'GACARG', 'SABICR']	60%	2.68	1.56%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	60%	2.67	1.61%
['VOZINT', 'SABICR']	70%	2.51	1.55%	['VOZINT', 'GACARG', 'SABICR']	70%	2.48	1.55%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	70%	2.43	1.61%
['VOZINT', 'SABICR']	80%	2.39	1.55%	['VOZINT', 'GACARG', 'SABICR']	80%	2.40	1.55%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	80%	2.37	1.61%
['VOZINT', 'SABICR']	90%	2.25	1.55%	['VOZINT', 'GACARG', 'SABICR']	90%	2.28	1.55%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	90%	2.22	1.61%
['VOZINT', 'INFOB']	60%	2.64	1.59%	['VOZINT', 'GACARG', 'INFOB']	60%	2.63	1.59%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	60%	2.64	1.61%
['VOZINT', 'INFOB']	70%	2.43	1.59%	['VOZINT', 'GACARG', 'INFOB']	70%	2.42	1.59%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	70%	2.43	1.61%
['VOZINT', 'INFOB']	80%	2.37	1.59%	['VOZINT', 'GACARG', 'INFOB']	80%	2.38	1.59%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	80%	2.39	1.60%
['VOZINT', 'INFOB']	90%	2.22	1.59%	['VOZINT', 'GACARG', 'INFOB']	90%	2.25	1.59%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	90%	2.28	1.60%
['GACARG', 'SABICR']	60%	2.67	1.55%	['VOZINT', 'SABICR', 'INFOB']	60%	2.64	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.64	1.60%
['GACARG', 'SABICR']	70%	2.49	1.55%	['VOZINT', 'SABICR', 'INFOB']	70%	2.43	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.40	1.60%
['GACARG', 'SABICR']	80%	2.40	1.55%	['VOZINT', 'SABICR', 'INFOB']	80%	2.37	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.36	1.60%
['GACARG', 'SABICR']	90%	2.29	1.55%	['VOZINT', 'SABICR', 'INFOB']	90%	2.22	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.23	1.60%
['GACARG', 'INFOB']	60%	2.68	1.60%	['GACARG', 'SABICR', 'INFOB']	60%	2.65	1.60%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.65	1.60%
['GACARG', 'INFOB']	70%	2.49	1.59%	['GACARG', 'SABICR', 'INFOB']	70%	2.46	1.59%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.41	1.60%
['GACARG', 'INFOB']	80%	2.43	1.58%	['GACARG', 'SABICR', 'INFOB']	80%	2.41	1.59%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.36	1.60%
['GACARG', 'INFOB']	90%	2.32	1.58%	['GACARG', 'SABICR', 'INFOB']	90%	2.30	1.59%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.22	1.60%
['SABICR', 'INFOB']	60%	2.63	1.62%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	60%	2.64	1.56%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.66	1.60%
['SABICR', 'INFOB']	70%	2.44	1.61%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	70%	2.42	1.57%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.41	1.60%
['SABICR', 'INFOB']	80%	2.38	1.60%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	80%	2.38	1.57%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.36	1.60%
['SABICR', 'INFOB']	90%	2.24	1.61%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	90%	2.26	1.57%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.22	1.60%

Table 11: Weights used for the Weighted GPT model (MAE version)

Source Codes	Sample Size	MAE	Weights	Source Codes	Sample Size	MAE	Weights	Source Codes	Sample Size	MAE	Weights
['NACION']	60%	2.70	1.54%	['NACION', 'ECONOM', 'VOZINT']	60%	2.66	1.52%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	60%	2.68	1.53%
['NACION']	70%	2.50	1.55%	['NACION', 'ECONOM', 'VOZINT']	70%	2.46	1.51%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	70%	2.46	1.53%
['NACION']	80%	2.38	1.56%	['NACION', 'ECONOM', 'VOZINT']	80%	2.38	1.51%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	80%	2.39	1.52%
['NACION']	90%	2.26	1.57%	['NACION', 'ECONOM', 'VOZINT']	90%	2.27	1.52%	['NACION', 'ECONOM', 'VOZINT', 'SABICR']	90%	2.28	1.52%
['ECONOM']	60%	2.66	1.52%	['NACION', 'ECONOM', 'GACARG']	60%	2.72	1.56%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	60%	2.77	1.58%
['ECONOM']	70%	2.48	1.51%	['NACION', 'ECONOM', 'GACARG']	70%	2.50	1.57%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	70%	2.53	1.58%
['ECONOM']	80%	2.40	1.50%	['NACION', 'ECONOM', 'GACARG']	80%	2.40	1.57%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	80%	2.41	1.59%
['ECONOM']	90%	2.29	1.51%	['NACION', 'ECONOM', 'GACARG']	90%	2.32	1.57%	['NACION', 'ECONOM', 'VOZINT', 'INFOB']	90%	2.28	1.61%
['VOZINT']	60%	2.67	1.52%	['NACION', 'ECONOM', 'SABICR']	60%	2.75	1.55%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	60%	2.72	1.59%
['VOZINT']	70%	2.48	1.51%	['NACION', 'ECONOM', 'SABICR']	70%	2.52	1.56%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	70%	2.50	1.60%
['VOZINT']	80%	2.39	1.50%	['NACION', 'ECONOM', 'SABICR']	80%	2.40	1.56%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	80%	2.40	1.60%
['VOZINT']	90%	2.28	1.52%	['NACION', 'ECONOM', 'SABICR']	90%	2.31	1.56%	['NACION', 'ECONOM', 'GACARG', 'SABICR']	90%	2.32	1.55%
['GACARG']	60%	2.74	1.58%	['NACION', 'ECONOM', 'INFOB']	60%	2.71	1.53%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	60%	2.69	1.67%
['GACARG']	70%	2.49	1.57%	['NACION', 'ECONOM', 'INFOB']	70%	2.50	1.53%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	70%	2.47	1.66%
['GACARG']	80%	2.37	1.56%	['NACION', 'ECONOM', 'INFOB']	80%	2.39	1.54%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	80%	2.33	1.67%
['GACARG']	90%	2.30	1.57%	['NACION', 'ECONOM', 'INFOB']	90%	2.25	1.56%	['NACION', 'ECONOM', 'GACARG', 'INFOB']	90%	2.22	1.69%
['SABICR']	60%	2.76	1.56%	['NACION', 'VOZINT', 'GACARG']	60%	2.74	1.55%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	60%	2.72	1.53%
['SABICR']	70%	2.51	1.53%	['NACION', 'VOZINT', 'GACARG']	70%	2.51	1.55%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	70%	2.52	1.53%
['SABICR']	80%	2.38	1.51%	['NACION', 'VOZINT', 'GACARG']	80%	2.40	1.54%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	80%	2.41	1.54%
['SABICR']	90%	2.28	1.52%	['NACION', 'VOZINT', 'GACARG']	90%	2.32	1.54%	['NACION', 'ECONOM', 'SABICR', 'INFOB']	90%	2.28	1.55%
['INFOB']	60%	2.75	1.67%	['NACION', 'VOZINT', 'SABICR']	60%	2.70	1.54%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	60%	2.74	1.57%
['INFOB']	70%	2.60	1.67%	['NACION', 'VOZINT', 'SABICR']	70%	2.47	1.53%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	70%	2.51	1.57%
['INFOB']	80%	2.41	1.68%	['NACION', 'VOZINT', 'SABICR']	80%	2.38	1.52%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	80%	2.40	1.56%
['INFOB']	90%	2.27	1.71%	['NACION', 'VOZINT', 'SABICR']	90%	2.28	1.52%	['NACION', 'VOZINT', 'GACARG', 'SABICR']	90%	2.30	1.53%
['NACION', 'ECONOM']	60%	2.67	1.52%	['NACION', 'VOZINT', 'INFOB']	60%	2.77	1.64%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	60%	2.74	1.70%
['NACION', 'ECONOM']	70%	2.47	1.53%	['NACION', 'VOZINT', 'INFOB']	70%	2.52	1.64%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	70%	2.52	1.70%
['NACION', 'ECONOM']	80%	2.36	1.54%	['NACION', 'VOZINT', 'INFOB']	80%	2.38	1.65%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	80%	2.36	1.71%
['NACION', 'ECONOM']	90%	2.25	1.55%	['NACION', 'VOZINT', 'INFOB']	90%	2.26	1.66%	['NACION', 'VOZINT', 'GACARG', 'INFOB']	90%	2.25	1.72%
['NACION', 'VOZINT']	60%	2.67	1.52%	['NACION', 'GACARG', 'SABICR']	60%	2.68	1.62%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	60%	2.82	1.59%
['NACION', 'VOZINT']	70%	2.47	1.51%	['NACION', 'GACARG', 'SABICR']	70%	2.47	1.62%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	70%	2.58	1.59%
['NACION', 'VOZINT']	80%	2.38	1.51%	['NACION', 'GACARG', 'SABICR']	80%	2.36	1.61%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	80%	2.45	1.60%
['NACION', 'VOZINT']	90%	2.28	1.52%	['NACION', 'GACARG', 'SABICR']	90%	2.30	1.56%	['NACION', 'VOZINT', 'SABICR', 'INFOB']	90%	2.36	1.60%
['NACION', 'GACARG']	60%	2.72	1.57%	['NACION', 'GACARG', 'INFOB']	60%	2.68	1.74%	['NACION', 'GACARG', 'SABICR', 'INFOB']	60%	2.73	1.65%
['NACION', 'GACARG']	70%	2.49	1.57%	['NACION', 'GACARG', 'INFOB']	70%	2.48	1.73%	['NACION', 'GACARG', 'SABICR', 'INFOB']	70%	2.52	1.65%
['NACION', 'GACARG']	80%	2.35	1.56%	['NACION', 'GACARG', 'INFOB']	80%	2.32	1.75%	['NACION', 'GACARG', 'SABICR', 'INFOB']	80%	2.37	1.66%
['NACION', 'GACARG']	90%	2.27	1.57%	['NACION', 'GACARG', 'INFOB']	90%	2.22	1.76%	['NACION', 'GACARG', 'SABICR', 'INFOB']	90%	2.32	1.64%
['NACION', 'SABICR']	60%	2.79	1.59%	['NACION', 'SABICR', 'INFOB']	60%	2.72	1.54%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	60%	2.74	1.56%
['NACION', 'SABICR']	70%	2.57	1.59%	['NACION', 'SABICR', 'INFOB']	70%	2.52	1.54%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	70%	2.51	1.56%
['NACION', 'SABICR']	80%	2.43	1.60%	['NACION', 'SABICR', 'INFOB']	80%	2.42	1.55%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	80%	2.40	1.53%
['NACION', 'SABICR']	90%	2.33	1.58%	['NACION', 'SABICR', 'INFOB']	90%	2.29	1.56%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR']	90%	2.29	1.52%
['NACION', 'INFOB']	60%	2.69	1.58%	['ECONOM', 'VOZINT', 'GACARG']	60%	2.72	1.54%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	60%	2.76	1.67%
['NACION', 'INFOB']	70%	2.48	1.58%	['ECONOM', 'VOZINT', 'GACARG']	70%	2.50	1.54%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	70%	2.53	1.66%
['NACION', 'INFOB']	80%	2.37	1.59%	['ECONOM', 'VOZINT', 'GACARG']	80%	2.41	1.53%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	80%	2.39	1.67%
['NACION', 'INFOB']	90%	2.23	1.61%	['ECONOM', 'VOZINT', 'GACARG']	90%	2.32	1.53%	['ECONOM', 'VOZINT', 'GACARG', 'INFOB']	90%	2.28	1.69%
['ECONOM', 'VOZINT']	60%	2.66	1.52%	['ECONOM', 'VOZINT', 'SABICR']	60%	2.67	1.52%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	60%	2.79	1.56%
['ECONOM', 'VOZINT']	70%	2.48	1.51%	['ECONOM', 'VOZINT', 'SABICR']	70%	2.47	1.51%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	70%	2.57	1.56%
['ECONOM', 'VOZINT']	80%	2.40	1.50%	['ECONOM', 'VOZINT', 'SABICR']	80%	2.39	1.50%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	80%	2.45	1.57%
['ECONOM', 'VOZINT']	90%	2.28	1.52%	['ECONOM', 'VOZINT', 'SABICR']	90%	2.27	1.52%	['ECONOM', 'VOZINT', 'SABICR', 'INFOB']	90%	2.35	1.58%
['ECONOM', 'GACARG']	60%	2.72	1.56%	['ECONOM', 'VOZINT', 'INFOB']	60%	2.78	1.58%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	60%	2.78	1.61%
['ECONOM', 'GACARG']	70%	2.48	1.57%	['ECONOM', 'VOZINT', 'INFOB']	70%	2.54	1.59%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	70%	2.55	1.62%
['ECONOM', 'GACARG']	80%	2.42	1.56%	['ECONOM', 'VOZINT', 'INFOB']	80%	2.42	1.60%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	80%	2.41	1.63%
['ECONOM', 'GACARG']	90%	2.33	1.56%	['ECONOM', 'VOZINT', 'INFOB']	90%	2.29	1.62%	['ECONOM', 'GACARG', 'SABICR', 'INFOB']	90%	2.33	1.62%
['ECONOM', 'SABICR']	60%	2.69	1.52%	['ECONOM', 'GACARG', 'SABICR']	60%	2.75	1.59%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.78	1.67%
['ECONOM', 'SABICR']	70%	2.48	1.52%	['ECONOM', 'GACARG', 'SABICR']	70%	2.51	1.59%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.56	1.67%
['ECONOM', 'SABICR']	80%	2.39	1.51%	['ECONOM', 'GACARG', 'SABICR']	80%	2.42	1.57%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.41	1.68%
['ECONOM', 'SABICR']	90%	2.29	1.52%	['ECONOM', 'GACARG', 'SABICR']	90%	2.31	1.54%	['VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.34	1.64%
['ECONOM', 'INFOB']	60%	2.66	1.52%	['ECONOM', 'GACARG', 'INFOB']	60%	2.68	1.70%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	60%	2.73	1.56%
['ECONOM', 'INFOB']	70%	2.48	1.51%	['ECONOM', 'GACARG', 'INFOB']	70%	2.50	1.69%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	70%	2.50	1.56%
['ECONOM', 'INFOB']	80%	2.37	1.51%	['ECONOM', 'GACARG', 'INFOB']	80%	2.35	1.70%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	80%	2.42	1.55%
['ECONOM', 'INFOB']	90%	2.24	1.53%	['ECONOM', 'GACARG', 'INFOB']	90%	2.24	1.71%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR']	90%	2.30	1.53%
['VOZINT', 'GACARG']	60%	2.76	1.55%	['ECONOM', 'SABICR', 'INFOB']	60%	2.70	1.52%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	60%	2.76	1.65%
['VOZINT', 'GACARG']	70%	2.53	1.55%	['ECONOM', 'SABICR', 'INFOB']	70%	2.51	1.52%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	70%	2.52	1.65%
['VOZINT', 'GACARG']	80%	2.40	1.54%	['ECONOM', 'SABICR', 'INFOB']	80%	2.40	1.52%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	80%	2.37	1.66%
['VOZINT', 'GACARG']	90%	2.33	1.54%	['ECONOM', 'SABICR', 'INFOB']	90%	2.26	1.54%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'INFOB']	90%	2.27	1.67%
['VOZINT', 'SABICR']	60%	2.69	1.53%	['VOZINT', 'GACARG', 'SABICR']	60%	2.76	1.57%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	60%	2.79	1.56%
['VOZINT', 'SABICR']	70%	2.48	1.52%	['VOZINT', 'GACARG', 'SABICR']	70%	2.53	1.57%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	70%	2.57	1.57%
['VOZINT', 'SABICR']	80%	2.39	1.50%	['VOZINT', 'GACARG', 'SABICR']	80%	2.39	1.54%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	80%	2.45	1.58%
['VOZINT', 'SABICR']	90%	2.27	1.52%	['VOZINT', 'GACARG', 'SABICR']	90%	2.29	1.52%	['NACION', 'ECONOM', 'VOZINT', 'SABICR', 'INFOB']	90%	2.34	1.58%
['VOZINT', 'INFOB']	60%	2.76	1.67%	['VOZINT', 'GACARG', 'INFOB']	60%	2.72	1.73%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	60%	2.77	1.61%
['VOZINT', 'INFOB']	70%	2.53	1.67%	['VOZINT', 'GACARG', 'INFOB']	70%	2.52	1.72%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	70%	2.53	1.61%
['VOZINT', 'INFOB']	80%	2.38	1.68%	['VOZINT', 'GACARG', 'INFOB']	80%	2.37	1.73%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	80%	2.40	1.62%
['VOZINT', 'INFOB']	90%	2.26	1.69%	['VOZINT', 'GACARG', 'INFOB']	90%	2.25	1.75%	['NACION', 'ECONOM', 'GACARG', 'SABICR', 'INFOB']	90%	2.32	1.62%
['GACARG', 'SABICR']	60%	2.70	1.62%	['VOZINT', 'SABICR', 'INFOB']	60%	2.83	1.59%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.77	1.65%
['GACARG', 'SABICR']	70%	2.48	1.61%	['VOZINT', 'SABICR', 'INFOB']	70%	2.59	1.59%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.54	1.66%
['GACARG', 'SABICR']	80%	2.39	1.58%	['VOZINT', 'SABICR', 'INFOB']	80%	2.46	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.39	1.67%
['GACARG', 'SABICR']	90%	2.30	1.54%	['VOZINT', 'SABICR', 'INFOB']	90%	2.36	1.60%	['NACION', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.34	1.64%
['GACARG', 'INFOB']	60%	2.60	1.78%	['GACARG', 'SABICR', 'INFOB']	60%	2.74	1.67%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.81	1.63%
['GACARG', 'INFOB']	70%	2.44	1.77%	['GACARG', 'SABICR', 'INFOB']	70%	2.53	1.67%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.57	1.63%
['GACARG', 'INFOB']	80%	2.30	1.78%	['GACARG', 'SABICR', 'INFOB']	80%	2.39	1.68%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.42	1.64%
['GACARG', 'INFOB']	90%	2.19	1.79%	['GACARG', 'SABICR', 'INFOB']	90%	2.33	1.65%	['ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.34	1.62%
['SABICR', 'INFOB']	60%	2.71	1.53%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	60%	2.72	1.54%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	60%	2.79	1.62%
['SABICR', 'INFOB']	70%	2.51	1.53%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	70%	2.50	1.54%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	70%	2.55	1.62%
['SABICR', 'INFOB']	80%	2.40	1.53%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	80%	2.41	1.54%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	80%	2.41	1.63%
['SABICR', 'INFOB']	90%	2.27	1.55%	['NACION', 'ECONOM', 'VOZINT', 'GACARG']	90%	2.33	1.54%	['NACION', 'ECONOM', 'VOZINT', 'GACARG', 'SABICR', 'INFOB']	90%	2.34	1.62%

Table 12: MAE and Diebold-Mariano Test for all windows proposed

Window Type	In-sample size	Model	MAE	p-value of DM Test	Window Type	In-sample size	Model	MAE	p-value of DM Test
fixed	60%	AR(1)	2.10		fixed	60%	gpt_weighted_model_mae	1.37	0.012
fixed	70%	AR(1)	2.04		fixed	60%	gpt_index_local_press	1.61	0.281
fixed	80%	AR(1)	1.53		fixed	70%	gpt_weighted_model_mae	1.47	0.201
fixed	90%	AR(1)	1.02		fixed	70%	gpt_index_local_press	1.66	0.586
recursive	60%	AR(1)	1.63		fixed	80%	gpt_weighted_model_mae	1.28	0.869
recursive	70%	AR(1)	1.61		fixed	80%	gpt_index_local_press	1.69	0.420
recursive	80%	AR(1)	1.30		fixed	90%	gpt_weighted_model_mae	0.85	0.256
recursive	90%	AR(1)	1.01		fixed	90%	gpt_index_local_press	1.19	0.612
rolling	60%	AR(1)	1.63		recursive	60%	gpt_weighted_model_mae	1.37	0.230
rolling	70%	AR(1)	1.59		recursive	60%	gpt_index_local_press	1.51	0.991
rolling	80%	AR(1)	1.31		recursive	70%	gpt_weighted_model_mae	1.43	0.536
rolling	90%	AR(1)	1.03		recursive	70%	gpt_index_local_press	1.64	0.801
fixed	60%	EPU_ARG_allpress	1.67	0.006	recursive	80%	gpt_weighted_model_mae	1.26	0.592
fixed	60%	EPU_ARG_local	1.42	0.000	recursive	80%	gpt_index_local_press	1.66	0.282
fixed	60%	EPU_ARG_foreign	1.82	0.068	recursive	90%	gpt_weighted_model_mae	0.84	0.254
fixed	70%	EPU_ARG_allpress	1.67	0.117	recursive	90%	gpt_index_local_press	1.11	0.539
fixed	70%	EPU_ARG_local	1.43	0.000	rolling	60%	gpt_weighted_model_mae	1.43	0.792
fixed	70%	EPU_ARG_foreign	1.80	0.399	rolling	60%	gpt_index_local_press	1.65	0.548
fixed	80%	EPU_ARG_allpress	1.33	0.831	rolling	70%	gpt_weighted_model_mae	1.57	0.901
fixed	80%	EPU_ARG_local	1.21	0.220	rolling	70%	gpt_index_local_press	1.85	0.454
fixed	80%	EPU_ARG_foreign	1.38	0.870	rolling	80%	gpt_weighted_model_mae	1.32	0.474
fixed	90%	EPU_ARG_allpress	1.15	0.573	rolling	80%	gpt_index_local_press	1.73	0.225
fixed	90%	EPU_ARG_local	1.15	0.526	rolling	90%	gpt_weighted_model_mae	0.81	0.248
fixed	90%	EPU_ARG_foreign	1.10	0.862	rolling	90%	gpt_index_local_press	1.08	0.523
recursive	60%	EPU_ARG_allpress	1.56	0.403	fixed	60%	customized_weighted_model_mae	1.72	0.002
recursive	60%	EPU_ARG_local	1.41	0.040	fixed	60%	customized_index_local_press	1.58	0.012
recursive	60%	EPU_ARG_foreign	1.57	0.669	fixed	70%	customized_weighted_model_mae	1.62	0.010
recursive	70%	EPU_ARG_allpress	1.56	0.758	fixed	70%	customized_index_local_press	1.57	0.086
recursive	70%	EPU_ARG_local	1.43	0.065	fixed	80%	customized_weighted_model_mae	1.18	0.042
recursive	70%	EPU_ARG_foreign	1.56	0.933	fixed	80%	customized_index_local_press	1.13	0.240
recursive	80%	EPU_ARG_allpress	1.32	0.583	fixed	90%	customized_weighted_model_mae	1.09	0.199
recursive	80%	EPU_ARG_local	1.23	0.754	fixed	90%	customized_index_local_press	1.13	0.098
recursive	80%	EPU_ARG_foreign	1.28	0.585	recursive	60%	customized_weighted_model_mae	1.46	0.109
recursive	90%	EPU_ARG_allpress	1.15	0.702	recursive	60%	customized_index_local_press	1.49	0.511
recursive	90%	EPU_ARG_local	1.11	0.508	recursive	70%	customized_weighted_model_mae	1.47	0.300
recursive	90%	EPU_ARG_foreign	1.10	0.969	recursive	70%	customized_index_local_press	1.54	0.865
rolling	60%	EPU_ARG_allpress	1.68	0.982	recursive	80%	customized_weighted_model_mae	1.09	0.452
rolling	60%	EPU_ARG_local	1.55	0.426	recursive	80%	customized_index_local_press	1.09	0.637
rolling	60%	EPU_ARG_foreign	1.65	0.681	recursive	90%	customized_weighted_model_mae	1.09	0.164
rolling	70%	EPU_ARG_allpress	1.65	0.882	recursive	90%	customized_index_local_press	1.12	0.108
rolling	70%	EPU_ARG_local	1.45	0.116	rolling	60%	customized_weighted_model_mae	1.58	0.725
rolling	70%	EPU_ARG_foreign	1.62	0.599	rolling	60%	customized_index_local_press	1.63	0.724
rolling	80%	EPU_ARG_allpress	1.38	0.414	rolling	70%	customized_weighted_model_mae	1.60	0.982
rolling	80%	EPU_ARG_local	1.25	0.839	rolling	70%	customized_index_local_press	1.75	0.542
rolling	80%	EPU_ARG_foreign	1.32	0.377	rolling	80%	customized_weighted_model_mae	1.17	0.719
rolling	90%	EPU_ARG_allpress	1.16	0.645	rolling	80%	customized_index_local_press	1.25	0.213
rolling	90%	EPU_ARG_local	1.11	0.490	rolling	90%	customized_weighted_model_mae	1.14	0.039
rolling	90%	EPU_ARG_foreign	1.10	0.975	rolling	90%	customized_index_local_press	1.22	0.051

References

- Andres-Escayola, E., Ghirelli, C., Molina, L., Pérez, J. J., & Vidal Muñoz, E. (2022). Using newspapers for textual indicators: Which and how many?
- Arakaki, A., Chamussy, L., & Vezza, E. (2020). Nowcasting poverty in argentina: A methodological note. *Washington, DC World Bank Group*. <http://documents.worldbank.org/curated/en/283681593599582508/Nowcasting-Poverty-in-Argentina-A-Methodological-Note>.
- Aromí, J. D., et al. (2022). Medición de incertidumbre económica en redes sociales en base a modelos de procesamiento de lenguaje natural (tech. rep.).
- Avela, A., & Lehmus, M. (2023). Negative economic sentiment index based on finnish news titles. *Journal of the Finnish Economic Association*, 4(1).
- Baker, S., Bloom, N., & Davis, S. (2016). Measuring economic policy uncertainty. *the quarterly journal of economics*, 131(4), 1593–1636.
- Bañbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting* (pp. 195–237, Vol. 2). Elsevier.
- Bholat, D., Hansen, S., Santos, P., & Schonhardt-Bailey, C. (2015). Text mining for central banks. *Available at SSRN 2624811*.
- Blanco, E., D’Amato, L., Dogliolo, F., & Garegnani, M. L. (2021). Nowcast de agregados macroeconómicos en argentina: Comparando la capacidad predictiva de diferentes modelos (tech. rep.). Economic Research Working Papers.
- Burri, M. (2023). Do daily lead texts help nowcasting gdp growth? (Tech. rep.). IRENE Working Paper.
- Cavallo, A. (2013). Online and official price indexes: Measuring argentina’s inflation. *Journal of Monetary Economics*, 60(2), 152–165.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Ghirelli, C., Pérez, J. J., & Urtasun, A. (2019). A new economic policy uncertainty index for spain. *Economics Letters*, 182, 64–67.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation.
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of monetary economics*, 55(4), 665–676.
- Gillmann, N., & Kim, A. (2021). Quantification of economic uncertainty: A deep learning approach.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

- 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.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, *62*(3), 1139–1168.
- Trust, P., Zahran, A., & Minghim, R. (2023). Understanding the influence of news on society decision making: Application to economic policy uncertainty. *Neural Computing and Applications*, *35*(20), 14929–14945.
- West, K. D., & McCracken, M. W. (1998). Regression-based tests of predictive ability.



Universidad de
San Andrés