

**“MAKING TEXT TALK”: THE MINUTES
OF THE CENTRAL BANK OF BRAZIL
AND THE REAL ECONOMY**

2022

BANCO DE **ESPAÑA**
Eurosistema

Documentos de Trabajo
N.º 2240

Carlos Moreno Pérez and Marco Minozzo

**“MAKING TEXT TALK”: THE MINUTES OF THE CENTRAL BANK OF BRAZIL
AND THE REAL ECONOMY**

“MAKING TEXT TALK”: THE MINUTES OF THE CENTRAL BANK OF BRAZIL AND THE REAL ECONOMY

Carlos Moreno Pérez (*)

PABLO DE OLAVIDE UNIVERSITY

Marco Minozzo (**)

UNIVERSITY OF VERONA

(*) This paper was written when he was working at the Banco de España. Currently at the Economics Department, Pablo de Olavide University, Ctra. de Utrera, 1, 41013 Seville, Spain, orcid: 0000-0002-4725-3019. E-mail(s): cmorper1@upo.es.

(**) Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy, orcid: 0000-0002-4351-8980. E-mail(s): marco.minozzo@univr.it.

<https://doi.org/10.53479/23646>

Documentos de Trabajo. N.º 2240

November 2022

The Working Paper Series seeks to disseminate original research in economics and finance. All papers have been anonymously refereed. By publishing these papers, the Banco de España aims to contribute to economic analysis and, in particular, to knowledge of the Spanish economy and its international environment.

The opinions and analyses in the Working Paper Series are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

The Banco de España disseminates its main reports and most of its publications via the Internet at the following website: <http://www.bde.es>.

Reproduction for educational and non-commercial purposes is permitted provided that the source is acknowledged.

© BANCO DE ESPAÑA, Madrid, 2022

ISSN: 1579-8666 (on line)

Abstract

This paper investigates the relationship between the views expressed in the minutes of the meetings of the Central Bank of Brazil's Monetary Policy Committee (COPOM) and the real economy. It applies various computational linguistic machine learning algorithms to construct measures of the minutes of the COPOM. First, we create measures of the content of the paragraphs of the minutes using Latent Dirichlet Allocation (LDA). Second, we build an uncertainty index for the minutes using Word Embedding and K-Means. Then, we combine these indices to create two topic-uncertainty indices. The first one is constructed from paragraphs with a higher probability of topics related to “general economic conditions”. The second topic-uncertainty index is constructed from paragraphs that have a higher probability of topics related to “inflation” and the “monetary policy discussion”. Finally, we employ a structural VAR model to explore the lasting effects of these uncertainty indices on certain Brazilian macroeconomic variables. Our results show that greater uncertainty leads to a decline in inflation, the exchange rate, industrial production and retail trade in the period from January 2000 to July 2019.

Keywords: Central Bank of Brazil, monetary policy communication, Latent Dirichlet Allocation, monetary policy uncertainty, Structural Vector Autoregressive model, Word Embedding.

JEL classification: C32, C45, D83, E52.

Resumen

Este documento investiga la relación entre las opiniones expresadas en las minutas de las reuniones del Comité de Política Monetaria del Banco Central de Brasil (COPOM) y la economía real. Construimos medidas de las minutas del COPOM utilizando varios algoritmos de aprendizaje automático. En primer lugar, creamos medidas del contenido de los párrafos de las minutas, utilizando el algoritmo Asignación Latente de Dirichlet (LDA, por sus siglas en inglés). En segundo lugar, construimos un índice de incertidumbre para las minutas, utilizando los modelos Word Embedding y K-Medias. Combinando los anteriores índices, creamos dos índices de incertidumbre temáticos. El primero de estos se construye con los párrafos que tienen una mayor probabilidad de temas relacionados con las «condiciones económicas generales». El segundo índice de incertidumbre temático se construye con los párrafos que tienen una mayor probabilidad de temas relacionados con la «inflación» y la «discusión de política monetaria». Finalmente, a través de un modelo Estructural de Vectores Autorregresivos, estudiamos los efectos de estos índices de incertidumbre en algunas variables macroeconómicas de Brasil. Nuestros resultados muestran que una mayor incertidumbre conduce a un descenso de la inflación, del tipo de cambio, de la producción industrial y del comercio al por menor en el período comprendido entre enero de 2000 y julio de 2019.

Palabras clave: Banco Central de Brasil, comunicación de la política monetaria, Asignación Latente de Dirichlet, incertidumbre de la política monetaria, modelo Estructural de Vectores Autorregresivos, Word Embedding.

Códigos JEL: C32, C45, D83, E52.

1 Introduction

Central bank communication is an important instrument in the toolbox, able to influence the financial markets and the real economy, since it provides some sort of information on the risks to price stability and growth. The higher the risks, the greater the likelihood of monetary policy intervention (Rosa and Verga, 2007). In other words, most of the central bank communications provide relevant information to the markets with the aim of guiding their policy decisions. Moreover, with their communications, central banks provide accountability, which is a key aspect of independent central banks (Binder, 2017).

Central banks communicate with the markets in different ways such as press conferences, statements of monetary policy decisions, inflation reports and the minutes of monetary policy meetings. The US Federal Open Market Committee (FOMC) opts to publish its minutes some days after the meeting. Similarly, after adopting an inflation targeted monetary approach during the 90s and the early 2000s, several central banks in Latin America — such as the Central Banks of Colombia, Mexico, Chile and Brazil — also publish the minutes of their monetary policy meetings. In the last decades, several authors have investigated the effect of Latin American central banks communications on the markets. In particular for the Central Bank of Brazil (Banco Central do Brasil in Portuguese), several investigations processed the information contained in the text of the communications manually, categorizing them as dovish or hawkish (Costa-Fiho and Rocha, 2010; Cabral and Guimaraes, 2015; García-Herrero, Girandin and Dos Santos, 2017). However, this way of proceeding can introduce some bias due to personal misinterpretation and requires a huge amount of work. Some papers have attempted to overcome these issues by using dictionary methods, that is, lists of words related to a sentiment or a topic as in Chague, De-Losso, Giovannetti and Manoel (2015) that apply this methodology for the communications of the Central Bank of Brazil.

In this paper, we investigate the relationship between the views expressed in the minutes of the meetings of the Monetary Policy Committee (COPOM) of the Central Bank of Brazil and the real economy. The minutes of the COPOM contain relevant information about the state of the economy, inflation

expectations and the reasons behind monetary policy decisions (Costa-Filho and Rocha, 2010). The COPOM meets a fixed number of times a year and its minutes are released the week after the meeting. Our main objective is to construct new measures of communication for the COPOM minutes. For that purpose, we suggest simple measures of communication to identify the topic and tone (sentiment) of the minutes of the Central Bank of Brazil, using machine learning methods.

Some economic investigations use the unsupervised machine learning technique Latent Dirichlet Allocation (LDA) to deduce the content (the topics) of relevant documents. The basis of LDA is that documents are depicted as a random combination of latent topics, where each topic is represented by a distribution over words (Blei et al, 2003). Our first contribution is to apply LDA to the minutes of the COPOM to infer the content of each paragraph. In particular, we identify the paragraphs with higher probability on topics related to ‘general economic conditions’ and the paragraphs with higher probability on topics related to ‘inflation and the monetary policy decision’. To the best of our knowledge, this is one of the first papers to use LDA to investigate the communications of the Central Bank of Brazil after Fasolo, Graminho and Bastos (2022) who use hierarchical Latent Dirichlet Allocation (hLDA) and LDA to obtain topic structures of the minutes of the COPOM and use dictionary methods to create sentiment measures. These are combined to obtain topic-sentiment measures and, in particular, an Economic Uncertainty Index. Apart from this work, many other authors used LDA and other techniques to analyse textual data (Bholat et al., 2015; Hansen, McMahon and Prat, 2017; Hansen, McMahon and Tong, 2019; Larsen and Thorsrud, 2019; Gentzkow, Kelly, and Taddy, 2019; Ash and Hansen 2022).

Our second contribution is the application (for the first time to the best of our knowledge) of the Skip-Gram and K-Mean models, following Soto (2021), to the communications of the Central Bank of Brazil to construct an ‘uncertainty’ dictionary, that is, a list of words similar to ‘uncertain’, ‘uncertainty’, ‘uncertainties’ and ‘fears’. With this ‘uncertainty’ dictionary, we build an uncertainty index for the minutes of the Central Bank of Brazil by counting the relative frequency of the words in it. This ‘uncertainty’ dictionary should be less biased and better adapted to the text than pre-established sentiment

dictionaries, such as, that of Loughran and McDonald (2011), since it is constructed with the text to be analyzed. In particular, according to Hansen, McMahon, and Prat (2017), machine learning methods have an important advantage over dictionary methods since they use all terms of the text to depict paragraphs in low-dimensional space instead of using parts of them. These authors also argue that machine learning techniques detect the most significant words in the data instead of imposing them. Moreover, Hansen, McMahon, and Prat (2017) state that a cognizable trait of LDA compared to other algorithms for dimensionality reduction is that it is fully probabilistic. However, there is still some degree of discretionality depending on the parameters selected to apply the Skip-Gram model since this might change some of the words in the dictionary.

We then construct topic-uncertainty measures, following Soto (2021), by combining the results of LDA and the Skip-Gram model for a better understanding of the sentiment (uncertainty) contained in the paragraphs discussing different topics. Specifically, we create two topic-uncertainty indices, one with the paragraphs more likely to include a group of topics related to ‘general economic conditions’, and the other with the paragraphs more likely to include a group of topics related to ‘inflation’ and the ‘monetary policy decision’. Other papers in the literature that have proposed topic-sentiment measures are those by Hansen, McMahon and Prat (2016), Azqueta-Gavaldon, Hirschbühl, Onorante and Saiz (2020), Cieslak, Hansen, McMahon and Xiao (2021), Moreno-Pérez and Minozzo (2022a), Moreno-Pérez and Minozzo (2022b).

As our last contribution, we analyzed the effect of the minutes, that is, of our uncertainty measures on the Brazilian real economy through a Structural Vector Auto-Regression (SVAR) model. Our results, based on a time interval spanning from January 2000 to July 2019, show that higher uncertainty in the minutes of the COPOM leads, in the same period, to a decrease in industrial production, inflation and retail sales. Moreover, they show that a unit shock in the uncertainty of the minutes is associated with a depreciation of the exchange rate. And finally, that a unit shock in the two topic-uncertainty indices has diverse effects, in the period 2000-2016, on exchange rate, inflation and industrial production.

2 Minutes of the Central Bank of Brazil

Some decades ago, inflation in Brazil was a major economic issue. Brazil suffered hyperinflation for almost 15 years from 1980 to 1994, during which inflation racked up an astonishing 13,342,346,717,617.70 percent. It was stopped by the introduction of the ‘Real Plan’ (‘Plano Real’) which included the introduction of a new currency the ‘Real’ and the privatization of state monopolies. In the 15 years after the introduction of the ‘Real Plan’, inflation was significantly reduced, totaling 196.87 percent over the period (Corrado, 2013).

In 1999, an inflation targeting regimen was adopted which allowed the ‘Real’ to fluctuate in response to market foreign-exchange mechanisms. The same year, the Central Bank of Brazil’s Monetary Policy Committee (COPOM) was created to increase transparency and trust in the monetary policy decision-making process. The COPOM is responsible for setting the stance on monetary policy and the short-term interest rate. The main goal of the COPOM is to achieve the inflation target established by the National Monetary Council. Moreover, the Central Bank of Brazil releases four types of documents related to monetary policy. First, an inflation report is published at the end of every quarter. Second, a summary of the decision of the COPOM is published after each meeting. Third, a focus report is released weekly containing projections for inflation, economic activity, the Selic rate and other economic indicators. Finally, the minutes of the meetings of the COPOM are published the week after the meeting.

In this paper, to limit our efforts and the computational burden, we decided to analyze solely the minutes of the meetings of the COPOM. However, these minutes are the most important documents of the Central Bank of Brazil since they explain in detail all the reasons behind the decisions of the COPOM. Our sample comprises all the minutes of the COPOM from the last meeting in 1999 to September 2019, which are available on the website of the Central Bank of Brazil. Hence, we have 184 minutes of the COPOM. From the end of 1999 until 2005, the COPOM met once a month, with an additional meeting in 2002. In 2006, the COPOM reduced the number of yearly meetings to eight. The meetings last two days. On the first day, current economic and

financial conditions are illustrated by the various departments and discussed by the members of the COPOM. On the second day, the members and head of the Research Department discuss the updated projections for inflation. Then, the COPOM takes its monetary policy decision. Since the 200th meeting of the COPOM in 2016, the statement of the final decision of the COPOM has included a summary of the domestic risks for the baseline scenario. Hence, part of the information in the minutes is not new for economic agents.

We use the English version of the minutes of the COPOM as a proxy of the Portuguese version. The English version is published one or several days after the Portuguese version. Since the 94th meeting in 2004 until the 199th meeting in 2016, the Portuguese version of the minutes was released on Thursday at 8:30 a.m. the week after the meeting. Since the 200th meeting, in 2016, the Portuguese version of the minutes is released on Tuesday at 8:30 a.m. the week after the meeting. The minutes are made public before the Brazilian Stock Exchange (BM&FBOVESPA Exchange) opens at 9:30.

3 Topic Analysis: Latent Dirichlet Allocation

We use simple measures of communication to identify the topic and the tone of the minutes of the Central Bank of Brazil. First, we apply Latent Dirichlet Allocation to identify the content or tone of each paragraph. We identify the paragraphs of the minutes that have a higher probability of the group of topics related to the current state of the economy, as well as paragraphs that have a higher probability of the group of topics related to inflation and monetary policy decisions. We then compute the tone of each paragraph. By tone, what is meant is the sentiment or degree of uncertainty in each paragraph of the minutes. To compute the tone, we apply the Skip-Gram and K-means algorithms to create a list of words similar to ‘uncertain’, ‘uncertainty’, ‘uncertainties’ and ‘fears’. Later, we build an uncertainty index by counting the number of times words from our ‘uncertainty’ list appear in each paragraph. Finally, we combine both topic and tone measures to construct two topic-uncertainty measures. The first topic-uncertainty index is constructed from paragraphs with a higher probability of topics related to general economic conditions. The second topic-uncertainty index is constructed from paragraphs with a higher probability of topics related to inflation and monetary policy decisions.

3.1 Latent Dirichlet Allocation model

Latent Dirichlet Allocation (LDA) is a machine learning technique introduced by Blei, Ng and Jordan (2003) that can be used for textual analysis. It is an unsupervised machine learning technique that aims to identify the topics or content of the text of all the documents interest without a person needing to read the text. The capacity of LDA to produce easy interpretable topics is one of its advantages. In order to do that, a name is assigned to each topic, for instance, 'industrial production' since the words most likely to appear are 'industry', 'production', 'goods', 'workers' and 'supply'. This labelling does not affect the results.

3.2 Corpus pre-processing and LDA estimation

In order to apply LDA, we manually transform the PDF of each set of minutes into text files. We remove from the minutes the parts that are not relevant for the LDA model such as the cover, the introduction, the footnotes and acronyms. We also assign tags to each paragraph to identify the date, the number and section of the minutes. All the words are changed to lower case and the data are 'cleaned' before applying LDA. The 'cleaning' data process for LDA requires three steps eliminating non-relevant information from the text. The first step is to remove the punctuation and stop words such as 'the', 'all', 'because', 'this', not relevant since they provide no information about the theme of the paragraph. Second, we stem the remaining words. Stemming is a process that consists in reducing words into their word stem or base root. For instance, the words 'inflationary', 'inflation', 'consolidate' and 'consolidating' are transformed into their stem 'inflat' and 'consolid', respectively. Finally, we rank these stems according to the term frequency-inverse document frequency (tf-idf). This index grows proportionally with the number of times a stem appears in a document. However, it decreases by the number of documents that contain that stem. This index serves to eliminate common and unusual words. We disregard all stems that have a value of 3,000 or lower. This cutoff of 3,000 seems reasonable with the tf-idf ranking.

In our research, we apply LDA with 9 topics to the 9,484 paragraphs that comprise all the minutes from the end of 1999 to September 2019. In our analysis, each paragraph corresponds to a document of the corpus. Our corpus comprises 2,900 unique stems and the total number of stems is 450,174. After

several trials with a different number of topics (from 30 to 5), the optimal number of topics turns out to be 9. This number of topics is used to differentiate paragraphs discussing topics related to ‘general economic conditions’ and paragraphs discussing topics related to ‘inflation expectations’ and the ‘monetary policy decision’. A smaller number of topics do not allow this differentiation since topics mix with each other.

Furthermore, we follow the suggestions of Griffiths and Steyvers (2004) to set the two hyperparameters of the Dirichlet priors. First, we set the Dirichlet prior on topics to $200/V$, where V is the number of single or unique vocabulary items. Second, we set the hyperparameter of the Dirichlet prior on document-topic distributions equal to $50/K$ where K is the number of topics (Hansen, McMahon and Tong, 2019). We run 1000 iterations before running the sample. Then, we twice run 20 samples from points in the chain thinned with a thinning interval of 50.

3.3 First LDA output: words per topic

Table A.1 shows the first output of LDA, i.e. the word-topic matrix. We display the first twelve words with the highest probability for each topic. Word 1 is the word or stem with the highest probability in that topic. Word 2 is the word with the second highest probability and so on. Most of the topics are easily understandable. We can divide the topics into two groups, those that include words related to ‘current economic conditions’ and those that include words related to ‘inflation’ and the ‘monetary policy decision’. The aim of this division is to assign each paragraph of the minutes to one of the two previous groups of topics as in Hansen and McMahon (2016).

The first group of topics discusses ‘general economic conditions’ and comprises topics 2, 4, 6, 7 and 8. We assign a tag to each topic for mere interpretation. For instance, to topic 8 we assign the tag ‘industrial production’ since it comprises mainly stems related to industrial production such as ‘product’ with a probability of 0.081, also ‘industr’, ‘good’, etc. The topics related to ‘current economic conditions’ represent the first day of the COPOM meeting during which the various heads of department inform COPOM board members of the current economic and financial situation of Brazil and international markets.

The second group contains topics that are related to the ‘current situation of inflation and its expectations’ and the ‘monetary policy decision’. This group includes topics 0, 1, 3 and 5. Usually, the description of the ‘current state of inflation’ takes place on the first day of the meeting and discussions of ‘inflation expectations’ and the ‘monetary policy decision’ occur on the second day.

3.4 Second LDA output: topics per document

The second output of LDA is the distribution of probabilities of each topic per document represented by the term β_k . In our paper, we assign each paragraph to one of the two groups of topics. We determine that a paragraph is part of the ‘general economic conditions’ group of topics if the sum of the β_k probabilities of the topics of this group is higher than or equal to 0.555% since 5 topics over 9 belong to the ‘general economic conditions’ group of topics. However, if the value of the sum of β_k of the ‘general economic conditions’ group of topics is smaller than 0.555%, the paragraph is assigned to the group of topics related to ‘inflation’ and the ‘monetary policy decision’.

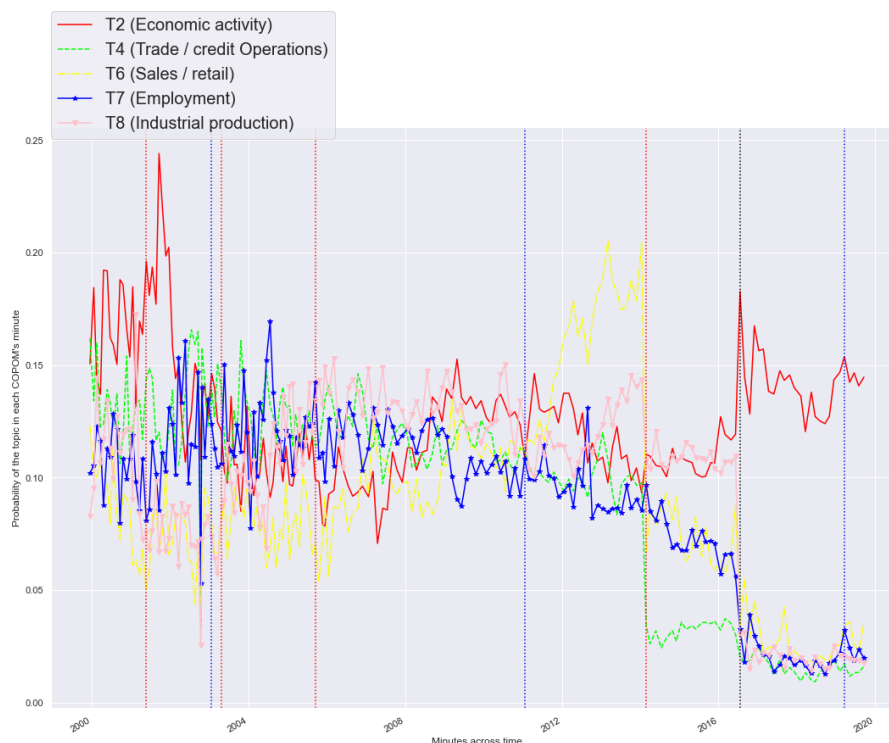


Fig. 1: Weights of topics 2, 4, 6, 7 and 8 in the minutes from December 1999 to 2019. Notes: The lines are the probabilities of each topic in each set of COPOM minutes. The vertical dotted blue lines represent a change in the Governor of the Central Bank of Brazil. The vertical dotted red lines represent a change in the format of the minutes. The vertical dotted black lines indicate a change in the format of the minutes and of the governor of the Central Bank of Brazil.

For illustrative purposes, we estimate the distribution of topics in the minutes. Figure 1 shows the probability of topics related to the ‘current economic situation’ in the minutes and Figure 2 shows the probability of topics related to ‘inflation’ and the ‘monetary policy decision’. In the figures there are events due to a change in the format of the minutes or to a change in the governor of the Central Bank of Brazil. Two events have a considerable effect. The first significant event occurs in the 181st minute in February 2014 due to a change in the format of the minutes. The 181st minute is represented in Figures 1 and 2 with a vertical dotted red line in 2014. The second event is in the 200th minute in 2016 where the format of the minutes is changed and the governor of the Central Bank of Brazil was replaced. Since the 200th minute in 2016 which is represented by the vertical dotted black line in Figures 1 and 2, topics related to ‘general economic conditions’ and ‘inflation’ have a lower probability than topics related to the ‘monetary policy decision’. The results of Fasolo, Graminho and Bastos (2022) capture also the same change in topic probabilities in these two events.

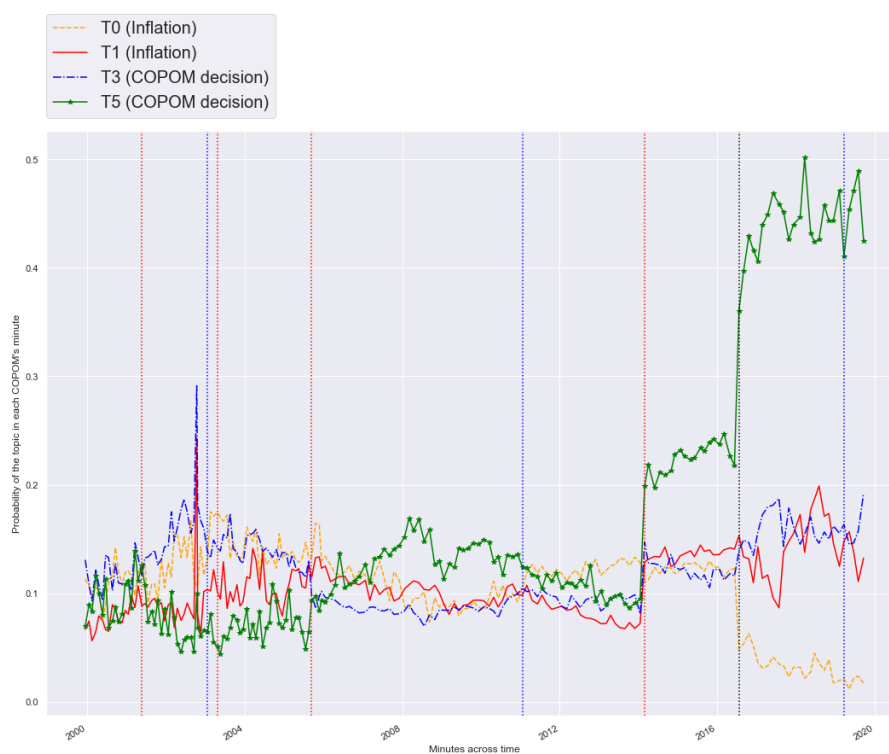


Fig. 2: Weights of topics 0, 1, 3 and 5 in the minutes from December 1999 to 2019. Notes: The lines are the probabilities of each topic in each set of COPOM minutes. The vertical dotted blue lines represent a change in the Governor of the Central Bank of Brazil. The vertical dotted red lines represent a change in the format of the minutes. The vertical dotted black lines indicate a change in the format of the minutes and of the governor of the Central Bank of Brazil.

4 Tone Analysis: Estimation of Uncertainty and Topic-Uncertainty Indices

Our next step is to determine the degree of uncertainty in each of the minutes. To measure the degree of uncertainty, we apply the Skip Gram model and K-Means following Soto (2021) to construct a list of words related to ‘uncertain’, ‘uncertainty’, ‘uncertainties’ and ‘fears’. We count the number of times that words from this ‘uncertainty’ list appear in each set of minutes compared to the total number of words in each set. Following the same procedure, we create two topic-uncertainty indices. First, we build an uncertainty index for the paragraphs more likely to contain topics related to the ‘current state of the economy’. Second, we construct an uncertainty index for the paragraphs more likely to contain topics related to ‘inflation’ and ‘monetary policy decisions’.

4.1 Word Embeddings theory and the Skip-Gram model

The Word Embeddings model was introduced by Mikolov et al. (2013). Word Embeddings are continuous vector representations of words with syntactical and semantic similarities between words in a Euclidean Space, decreasing the size of the text. The main idea of Word Embeddings is that we obtain a lot of meaning from a word by its context, i.e. the words around it or where it is embedded. For instance, consider the following documents:

Document 1: the economy experienced a period of growing **uncertainty** about the growth capacity

Document 2: the economy experienced a period of growing **concerns** about the growth capacity

The words ‘uncertainty’ and ‘concerns’ have similar meanings related to doubt and worry. In addition, the words ‘uncertainty’ and ‘concerns’ are preceded by ‘the economy experienced a period of growing’ and followed by ‘about the growth capacity’. The basic idea of Word Embeddings is to create a dense vector for each word type that is good at predicting the words that appear in their context and are also represented by a vector. In that case, we prefer a machine learning method that puts the vectors of words with similar mean-

ing such as ‘uncertainty’ and ‘concerns’ in the same part of the vector space since they appear in the same context. To create the Word Embeddings in this way, we utilize the Skip-Gram model introduced by Mikolov et al. (2013). The Skip-Gram model is a Neural Network machine learning method that tries to predict context words on the basis of a center word. This process is repeated for all the unique terms in the corpus, and for each term a vector of probabilities is created and placed in the vector space. For instance, uncertainty is the input or center word in document 1. The rest of the words are the output or context words.

$$\underbrace{\text{economy experienced growing}}_{\text{Output}} \quad \underbrace{\text{uncertainty}}_{\text{Input}} \quad \underbrace{\text{about the growth capacity}}_{\text{Output}}$$

In the previous example, the Skip-Gram model provides the probability distribution of each of the context words based on the word uncertainty, which is the center word. For instance, $P(\text{growing} \mid \text{uncertainty})$ or $P(\text{about} \mid \text{uncertainty})$. For each word ($t = 1, \dots, T$), the number of words in the context is given by the size of the window, m , that determines the number of context words before and after each center word. A window size of five means we estimate the probabilities of the five output words previous to the input word and the five output words following the input word.

4.2 K-Means Clustering

K-Means Clustering is a technique that attempts to link observations that are close to each other in the input space. In this paper, we use K-Means to cluster the Word Embeddings, which are vector representations constructed with the Skip-Gram model, into C disjoint groups or clusters. We then identify the cluster that encompass the words ‘uncertain’, ‘uncertainties’, ‘uncertainty’ and ‘fears’ as in Soto (2021).

K-Means is a centroid-base algorithm. This algorithm aims to find the cluster assignments of all m observations to C clusters that minimize the within cluster distances between each point x_i and its cluster centre μ_c (Chakraborty and Joseph, 2017). The within cluster distances is normally measured by the Euclidean distance. The corresponding cost function is:

$$ERR(X, C) = \frac{1}{m} \sum_{c=1}^C \sum_{x_i \in C_c} \|x_i - \mu_c\|^2. \quad (1)$$

Here, the sum of squares is normalized by the number of observations, as required to compare clusters of different sizes. In order to establish a fixed number of clusters C , we alternate steps of cluster assignment and centroid shifting. During clustering assignment, we assign each observation x_i to its closest centroid C_i . In centroid shifting, we compute the new position for each centroid. Moreover, highly correlated features must be avoided since they might cause spurious clustering. Finally, the number of clusters needs to be decided. Several evaluation methods can be used including the ‘silhouette coefficient’ and ‘elbow-method’ (Chakraborty and Joseph, 2017).

4.3 Estimation of Word Embeddings

The Skip-Gram model is applied to the same corpus of minutes of the Central Bank of Brazil. Nonetheless, there are some differences in the preprocessing of the corpus. First, the words in the Skip-Gram corpus are not stemmed because of the risk of losing information due to the semantic differences between words. Second, we identify bigrams or pairs of words that appear with a frequency higher than 10. The bigrams identify couples of words that represent the same term or idea. Finally, the text in the Skip-Gram model is a whole unique document instead of different documents comprising paragraphs as in LDA.

We attempt different combinations of the Hidden layer and the window size in the Skip-Gram model.¹ We select parameters that provide logical results. In particular, we estimate Skip-Gram with a Hidden layer (H) of 200 and a context window size (m) of 10. Furthermore, 140 clusters are selected for the application of K-Means.

After applying the Skip-Gram and K-Means models, we select all the words in the same clusters as ‘uncertainty’, ‘uncertain’, ‘uncertainties’ and ‘fears’ to construct a dictionary or list of words related to uncertainty. Though words in the same cluster do not share the same semantic meaning, they are related to similar contexts. The words in the same clusters as the words ‘uncertainty’, ‘uncertain’, ‘uncertainties’ and ‘fears’ are shown in Table 1. The list of ‘uncertainty’ words includes words such as ‘unstable’, ‘ambiguous_influence’, ‘turmoil’ and ‘risks’. Other words describe critical events such as ‘earthquake’,

¹We implement the Skip-Gram model with the Gensim library (Word2Vec) in Python.

Table 1: List of words in the same cluster as the words ‘uncertain’, ‘uncertainty’, ‘uncertainties’ and ‘fears’.

abrupt, absence, abundant, abundant_global, actually, adjust, adverse, affirm, africa, alternative, america, american, ample, another_concern, apparently_little, asian, asset, assign_low, assume, asymmetric, attack, attacks, band, benign_inflationary, brazilian_assets, brexit, capital_flows, causing, chances, chinese_economy, clear_identification, closely_monitored, committee_understands, commodities, commodity, complex, complexity, complexity_surrounding, comprise, concerns, concretization, consequences, consequent, considerable_degree, constitute, constraints, contaminate, could, could_affect, decades, deficits, degree, deleverage, depends, depreciating, derive, derived, deriving, despite_identifying, deteriorate, deterioration, devaluation, developed_countries, deviates, diagnosis, dollar_appreciation, dollar_depreciation, earlier, earthquake, ease, eased, eastern, economic_blocks, elections, electoral_process, emerging, emerging_countries, enable_natural, environment, episodes, equity_markets, european_countries, evaluates, existence, exporting_countries, extent_reflect, external_environment, external_financing, extraordinary, extreme_events, faced, facts, fashion, favoring, fear, fears, financial_markets, financing_conditions, fragility, fueled, generate, geopolitical, geopolitical_tensions, global_outlook, gradual_normalization, handling, heating, heightened, heterogeneous, highly_volatile, identifies, imply, impose, impose_adjustments, incidentally, industrialized_countries, industrialized_economies, inflationary, initially_localized, initiatives_taken, instability, international, international_financial, iraq, justified, latent, latin_america, less_likely, likelihood, localized, low_probability, major, major_advanced, major_economies, manifest, markets, markets_quotations, mechanisms, middle_east, midst, might, minor, mitigate, mortgage_crisis, movements, moves, nevertheless, news, normalization, north, northern_hemisphere, notably, nuclear, observes, ongoing_deleveraging, opec, originally, originated, particularly, persists, pessimism, political, pondered, pose, positive_spillovers, possible, potentially, predominantly, premature, pressuring, prevalence, pricing, problems, producing_countries, promptly_converges, prospectively, provoked, prudent, quotations_remains, reacting, reaction, reactions, realignments, reassessment, recently, recurrent_geopolitical, remain_tied, remains_complex, repercussions, risk, risk_appetite, risk_aversion, risks, risky_assets, satisfactory, scarcity, selected_commodities, shortage, show_resistance, significant_deterioration, since_mid, speculative, spillovers, stem, strongly_impacted, subdued, subsequent_years, substantial_share, suffer, surround, surrounded, surrounding, swings, tension, tensions, tensions_despite, tightened, towards_normality, traditionally, transition, transitory, turmoil, uncertain, uncertainties, uncertainty, uncertainty_concerning, unstable, valuation, venezuela, volatility, volatility_affecting, war, wave, weaken, wealth, widening, widespread_disinflation, winter, world, world_economy, worldwide, worries, would, yen.

‘brexit’, ‘mortgage_crisis’ or ‘war’. Besides, there are terms related to oil-producing countries that might be in trouble as ‘iraq’, ‘opec’ or ‘venezuela’. Some words are related to the business cycle such as ‘widespread_disinflation’, ‘devaluation’ or ‘dollar_appreciation’. Moreover, our results might not fully show the potential of the Skip-Gram model since the data available for the minutes of Brazil are limited compared to the size of current databases as in the case of social media.

4.4 Estimation of uncertainty and topic-uncertainty indices

An uncertainty index for the minutes of the Central Bank of Brazil is constructed by assigning an uncertainty score to each set of minutes. As we show in Equation (14), the uncertainty score of each set of minutes is computed as

the number of times any word in our ‘uncertainty’ list appears divided by the total number of words in that set of minutes. We standardize the uncertainty score by multiplying it by 100 and dividing it by the mean score of all the minutes used to construct the uncertainty index as shown in Equation (15).

$$S_s = U_s/N_s, \quad (2)$$

$$F_s = 100 \frac{S_s}{\frac{1}{M} \sum_{m=1}^M S_m}, \quad (3)$$

where, the term U_s is the number of uncertainty words in minute s , and N_s is the total number of words in that set of minutes. Furthermore, S_s and F_s are the uncertainty score and the uncertainty index of minute s , respectively. The denominator of Equation (15) is the mean of all the values of the uncertainty score.

Figure 3 shows the evolution of the minutes uncertainty index. We compare it with the Economic Policy Uncertainty (EPU) index for Brazil created by Baker, Bloom and Davis (2016). There are two main differences in the construction of our minutes uncertainty index and the EPU index that could affect their behavior. First, the minutes uncertainty index is constructed with the corpus of the minutes of the COPOM in which the Brazilian and the international current economic and financial conditions, the updated projections for inflation and the monetary policy decision are discussed in due proportion. Nonetheless, the Brazilian EPU index is constructed starting from newspaper articles of the Brazilian newspaper ‘Folha de Sao Paulo’, which might not always provide information similar to that contained in the minutes. Second, the minutes uncertainty index is constructed using machine learning techniques such as the Skip-Gram model and K-Means. On the contrary, the Brazilian EPU index is based on the counts of the number of articles that contain at least one word in each of three subjectively pre-determined groups of words. The first group contains words related to policy terms such as ‘regulation’ or ‘deficit’, the second comprises the words ‘uncertain’ and ‘uncertainty’, whereas the third group comprises the words ‘economic’ and ‘economy’. The methodology of Baker, Bloom and Davis (2016) might suffer of problems similar to those of the dictionary methods. To compare the two indices, we standardize

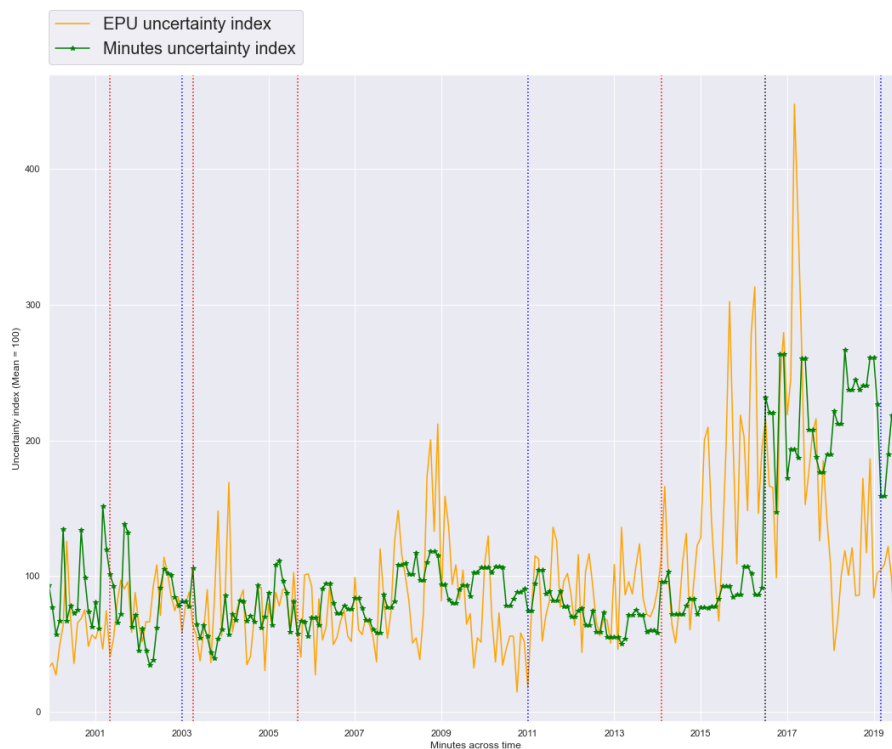


Fig. 3: Minutes uncertainty index - December 1999 to 2019. Notes: The vertical dotted blue lines represent a change in the Governor of the Central Bank of Brazil. The vertical dotted red lines represent a change in the format of the minutes. The vertical dotted black lines indicate a change in the format of the minutes and of the governor of the Central Bank of Brazil

the EPU index following Equation (15) so that the mean of the EPU index is 100 for our sample. Figure 3 shows that the minutes uncertainty index follows a pattern somehow similar to that of the EPU index of Baker, Bloom and Davis (2016), though our index increases significantly in correspondence of the 200th minute in 2016 (represented by the vertical dotted black line), which coincides with the replacement of the governor of the Central Bank of Brazil and with a change in the format of the minutes. It should be noted that the 200th minute introduced a sort of structural break in the behaviour of the index, which does not allow an immediate comparison of the values of the index before and after this event. The results of Fasolo, Graminho and Bastos (2022), which use a dictionary method on the minutes of the COPOM to create an uncertainty index, are similar to ours and show a similar behavior in correspondence of the 200th minute. Differently to our index, the index of Baker, Bloom and Davis (2016) increases rapidly after 2014 capturing one of the worst economic crises in recent decades (lasting from 2014 to 2016) of the Brazilian economy.

We construct two topic-uncertainty indices, creating the first topic-uncertainty index for the paragraphs more likely to include topics related to

‘general economic conditions’. Another topic-uncertainty index is created for the paragraphs more likely to include topics related to ‘inflation’ and the ‘monetary policy decision’. To build the two topic-uncertainty indices, we follow the same procedure as described for the general uncertainty index. With the two topic-uncertainty indices, we can identify the origin of uncertainty either in the ‘general economic situation’ paragraphs or the ‘inflation’ and ‘monetary policy decision’ paragraphs. Figure 4 shows the evolution of the two topic-uncertainty indices and we compare them again to the EPU index of Baker, Bloom, and Davis (2016) for Brazil. From 2000 until 2014, the ‘inflation’ and the ‘monetary policy decision’ topic-uncertainty index is higher for almost all the periods than the ‘general economic conditions’ topic-uncertainty index. In 2014, there was an economic crisis in Brazil, reflected by the fact that the ‘general economic conditions’ uncertainty index outscores the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index. Finally, again there was a considerable increase in both topic-uncertainty indices after the 200th minutes which is represented by the vertical dotted black line, especially in the ‘general economic conditions’ uncertainty index. Nonetheless, the number of paragraphs covering the ‘general economic conditions’ decreases drastically

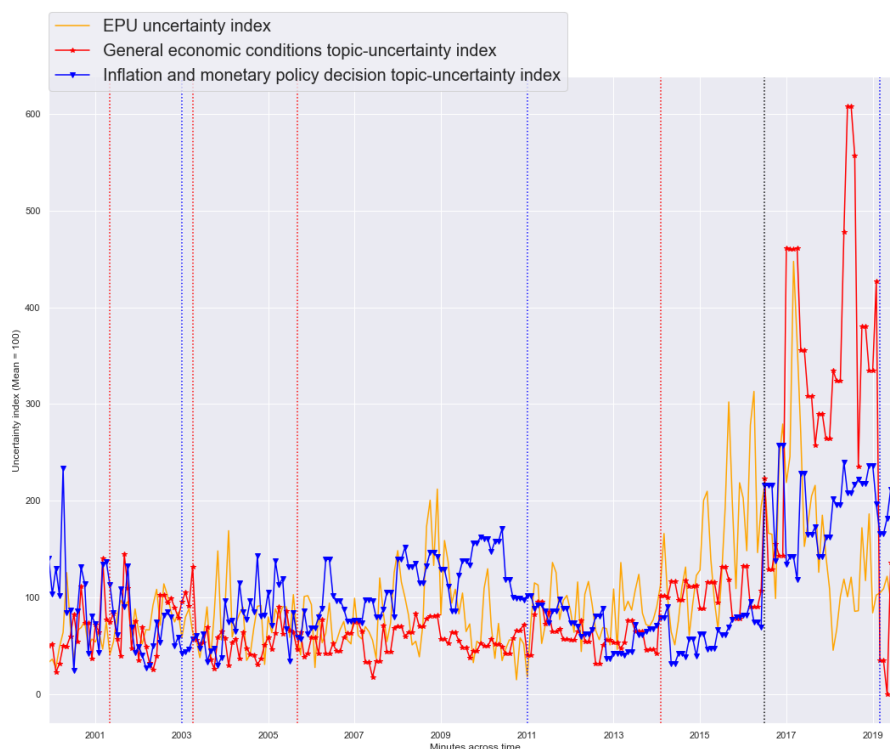


Fig. 4: Topic-uncertainty indices - December 1999 to 2019. Notes: The vertical dotted blue lines represent a change in the Governor of the Central Bank of Brazil. The vertical dotted red lines represent a change in the format of the minutes. The vertical dotted black lines indicate a change in the format of the minutes and of the governor of the Central Bank of Brazil

after the 200th meeting in July 2016, leading to more volatility in this index, including values equal to zero. Therefore, our analysis discards the ‘general economic conditions’ topic-uncertainty index after the 200th minutes.

5 Structural VAR

One of the most similar paper to ours is Hansen and McMahon (2016) who investigate FOMC statements. With LDA and manually they identify the parts of FOMC statements that discuss ‘current economic conditions’ or the ‘monetary policy decision’. For the part related to ‘current economic conditions’ they create a positive-negative index with words associated with expansion and recession in the dictionary list of Apel and Blix Grimaldi (2012). For the ‘monetary policy decision’ parts of FOMC statements, they estimate a topic-uncertainty index by counting the relative frequency of the words in the uncertainty dictionary of Loughran and McDonald (2011). Later, they estimate a Factor-Augmented Vector Autoregression (FAVAR) to investigate the effect of the text measures in the market and real variables. They observe that the effect of communications’ shocks in ‘current economic conditions’ in market and real variables is lower than the effect of communications’ shocks in the ‘monetary policy decision’ part of the FOMC statements.

We investigate the effect of the uncertainty index and the two topic-uncertainty indices in the Brazilian economy. For this purpose, we compute a Structural Vector Autoregression (SVAR) model:

$$B_0 Y_t = \sum_{i=1}^p B_i Y_{t-i} + \omega_t, \quad (4)$$

where, ω_t refers to a structural innovation or structural shock, but also represents the mean zero serially uncorrelated error term. The term Y_t is a K -dimensional time series $t = 1, \dots, T$. The term Y_t is approximated by a vector autoregression of finite order p . The matrix B_0 represents the simultaneous associations of variables in the model (Kilian and Lütkepohl, 2017).

In our paper, the vector $Y_t = [\Delta F_t, \Delta E_t, \Delta \pi_t, \Delta P_t, \Delta C_t]$ where ΔE_t stands for the difference in the Real broad effective exchange rate for Brazil, $\Delta \pi_t$ indi-

cates the difference in the consumer price index in Brazil, ΔP_t is the difference in total industrial output in Brazil, and ΔC_t is the difference in total retail trade. ΔF_t stands for the difference in the range of the uncertainty indices. For clarification, differences indicate first differences of time series, taken over subsequent time instants. For the months with no meetings, we assume the value of the uncertainty index of the previous set of minutes. Moreover, all the macroeconomic variables are extracted with monthly frequency from the Federal Reserve Bank of St. Louis. All variables are differentiated to overcome the non-stationary problem in light of the augmented Dickey-Fuller test indicating I(1).

The optimal number of lags is in line with Akaike Information Criteria (AIC), the Bayesian Information Criterion (SBIC), and the Hannan and Quinn Information Criterion (HQIC). The SVAR model complies with the stability condition since all roots of the characteristic polynomial are outside the unit circle. The identification of structural shock is obtained by appealing to the usually estimated Cholesky decomposition put forward by Sims (1980). The Cholesky decomposition involves the so-called recursiveness assumption, an economic assumption about the timing of the reaction to shocks in the variables. In other words, the recursiveness assumption imposes order between the variables. In our paper, the uncertainty index (ΔF_t) simultaneously affects the other variables, but is not affected by the remainder as in Bloom (2009) and Nodari (2014). Hence, ΔE_t simultaneously affects $\Delta \pi_t$, ΔP_t and ΔC_t . $\Delta \pi_t$ has a simultaneous impact on ΔP_t and ΔC_t . Subsequently, it continues this way for the last two variables. We estimate the Structural VAR model for each of the uncertainty indices. First, we make two estimations with the full sample for the following two uncertainty indices: 1) the minutes uncertainty index; 2) the ‘inflation’ and ‘the monetary policy decision’ topic-uncertainty index. Then, we restrict the sample until the 199th minutes in June 2016 due to a lack of data for the ‘general economic conditions’ topic-uncertainty index. We again estimate Structural VAR with this reduced sample for all the uncertainty indices constructed from the minutes: 3) the general uncertainty index for the minutes; 4) the ‘inflation’ and ‘the monetary policy decision’ topic-uncertainty index; 5) the ‘general economic conditions’ topic-uncertainty index.

6 Results

Figure A.1 and A.2 show the results of the impulse response analysis for the whole sample from 2000 to July 2019. Figure A.1 demonstrates the effects of an increase in a unit shock in the minutes uncertainty index in four Brazilian macroeconomic variables. A rise in one standard shock in the uncertainty index of the minutes depreciates the exchange rate by almost 0.3%. Moreover, an increase in the uncertainty index slightly reduces inflation. However, in two periods after the shock it becomes positive. Lastly, industrial production and the retail trade both decrease by around 0.16% with a unit shock in the general uncertainty index. The results of industrial production and the retail trade are similar to the results of Costa-Filho (2014) after a unit in the uncertainty index. The results of Godeiro and de Oliveira-Lima (2017) also suggest the same negative relationship between macroeconomic uncertainty and industrial production in Brazil. In Figure A.2, the results of the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index are similar to the results of the uncertainty index. The effect on industrial production lasts longer for the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index.

Figures A.3 to A.5 repeat the analysis for all the uncertainty indices constructed from the minutes from 2000 to June 2016. Figure A.3 shows the impulse response functions of the uncertainty index. The results are similar to those computed for the whole sample, as shown in Figure A.1. However, in the reduced sample industrial production decreases drastically in the period following the shock rather than in the same period, as shown in Figure A.1. Figure A.4 shows the results of the impulse response functions for the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index with the reduced sample. Figure A.5 shows the ‘general economic conditions’ topic-uncertainty index with the reduced sample. A unit shock in the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index leads to a larger fall in the exchange rate than in the results of the ‘general economic conditions’ topic-uncertainty index. This might be explained by the large depreciation of the Brazilian Real after the world financial crisis of 2008 during the ‘world currency war’. This depreciation attempted to make Brazilian exports more competitive. In the five years after the financial crisis of 2008, the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index is relatively high. This might be a proxy

of the complex international financial situation facing COPOM board members. The ‘general economic conditions’ topic-uncertainty index has a low value during the five years after the world’s economic crisis of 2008, capturing the growth of the Brazilian economy in that period.

In Figure A.5, we observe that a unit shock in the ‘general economic conditions’ topic-uncertainty index has a positive impact on inflation. However, the impact of a unit shock in the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index has a negative impact on inflation. This might be explained by the fact that the ‘general economic conditions’ topic-uncertainty index is higher than the ‘inflation’ and ‘monetary policy uncertainty’ topic-uncertainty index during periods of higher inflation and tougher economic conditions (beginning of the decade of 2000s and from 2014 to 2016). It might also be related to the fact that COPOM members express more uncertain views in the paragraphs related to ‘inflation’ and ‘monetary policy decision’ during the period after the financial crisis of 2008 characterized by lower inflation.

In addition, the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index has a higher negative effect on industrial production than the ‘general economic conditions’ topic-uncertainty index. This might be explained by the sharp fall in industrial production after the financial crisis of 2008 which may be correlated with an increase in the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index in the same period. Finally, we observe similar results for a unit shock in retail for both topic-uncertainty indices.

We check the validity of our results by estimating the Structural VAR model with an external uncertainty index such as the EPU index for Brazil. Figure A.6 shows the results of the impulse response analysis for the standardized EPU uncertainty index for the whole sample. The results are similar to those of the uncertainty index of the minutes. Nonetheless, an increase in one standard shock of the EPU index leads to a fall in the exchange rate three times higher than is the case for the uncertainty index of the minutes (Figure A.1). Figure A.7 shows results of the impulse response analysis for the standardized EPU uncertainty index for the period 2000 - June 2016. Again, these results are similar to those of the uncertainty index of the minutes, as shown in Figure A.3. In Figure Figure A.7, in the same period, an increase of

one-unit shock in the EPU index has a positive effect on retail and later drop to negative values in the periods after the shock.

7 Conclusion

This paper investigates the relationship between the views expressed in the minutes of the meetings of the Monetary Policy Committee (COPOM) of the Central Bank of Brazil and the real economy. For this purpose, we suggest simple measures of communication to identify the topic and tone of the minutes of the Central Bank of Brazil. First, topic or content analysis enables us to understand what the minutes are talking about. Here, we use Latent Dirichlet Allocation to deduce the content or topics of each paragraph of our sample. We identify two main groups of topics, the ‘current economic conditions’ topics and the ‘inflation’ and ‘monetary policy decision’ topics. By tone analysis, we compute the degree of uncertainty in each paragraph of the minutes. We use the Skip-Gram and the K-means algorithms to create a list of words with similar meaning to ‘uncertain’, ‘uncertainty’, ‘uncertainties’ and ‘fears’ comprising our dictionary of words related to ‘uncertainty’. We then compute the relative frequency of the words from the ‘uncertainty’ dictionary to construct an uncertainty index for the minutes of the Central Bank of Brazil and combine both topic and tone text measures to build two topic-uncertainty indices. The first topic-uncertainty index is constructed from paragraphs that are more likely to include topics related to ‘general economic conditions’. We create a second topic-uncertainty index from the paragraphs that are more likely to include topics related to the ‘inflation situation and expectations’ and the ‘monetary policy decision’.

Finally, with a Structural VAR model we estimate the effect on the real economy corresponding to an increase in the uncertainty index of the minutes and the two topic-uncertainty indices. Our results show that higher uncertainty in the minutes of the COPOM leads to a fall in the exchange rate, industrial production, inflation, and retail sales. We also show the differing impacts on the ‘general economic conditions’ topic-uncertainty and the ‘inflation’ and ‘monetary policy decision’ uncertainty index in relation to macroeconomic variables such as the exchange rate, inflation and industrial production.

Future research could further investigate the communications of the Central Bank of Brazil such as the monetary policy statements or study the effect in the financial markets. Future research could also use alternative unsupervised machine learning methods such as Dynamic Topic Models (Blei and Lafferty, 2006).

References

- Apel, M., and M. Grimaldi (2012). *The information content of central bank minutes*, Riksbank Research Paper Series (92). <http://dx.doi.org/10.2139/ssrn.2092575>
- Ash, E., and S. Hansen (2022). *Text algorithms in economics*. https://elliottash.com/wp-content/uploads/2022/08/annual_reviews_chapter-022-08-16.pdf
- Azqueta-Gavaldón, A., D. Hirschbühl, L. Onorante, et al. (2020). *Economic policy uncertainty in the euro area: An unsupervised machine learning approach*, European Central Bank Working Paper Series (2359). <http://dx.doi.org/10.2139/ssrn.3516756>
- Baker, S. R., N. Bloom, and S. J. Davis (2016). “Measuring economic policy uncertainty”, *The Quarterly Journal of Economics*, 131(4), pp. 1593-1636. <https://doi.org/10.1093/qje/qjw024>
- Bholat, D., S. Hansen, P. Santos, et al. (2015). *Text mining for central banks*, Centre for Central Bank Studies - Bank of England (33). <http://dx.doi.org/10.2139/ssrn.2624811>
- Binder, C. (2017). “Fed speak on main street: Central bank communication and household expectations”, *Journal of Macroeconomics*, 52, pp. 238-251. <https://doi.org/10.1016/j.jmacro.2017.05.003>
- Blei, D. M., and J. D. Lafferty (2006). “Dynamic Topic Models”, In: *ICML '06: Proceedings of the 23rd International Conference on Machine Learning*, pp. 113-120. <https://doi.org/10.1145/1143844.1143859>
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3, pp. 993-1022. <https://dl.acm.org/doi/10.5555/944919.944937>
- Bloom, N. (2009). “The impact of uncertainty shocks”, *Econometrica*, 77(3), pp. 623-685. <https://doi.org/10.3982/ECTA6248>
- Cabral, R., and B. Guimarães (2015). “O comunicado do Banco Central”, *Revista Brasileira de Economia*, 69, pp. 287-301. <https://doi.org/10.5935/0034-7140.20150013>
- Chague, F., R. De-Losso, B. Giovannetti, et al. (2015). “Central bank communication affects the term-structure of interest rates”, *Revista Brasileira de Economia*, 69, pp. 147-162. <https://doi.org/10.5935/0034-7140.20150007>
- Chakraborty, C., and A. Joseph (2017). *Machine learning at central banks*, Bank of England Working Paper (674). <https://www.bankofengland.co.uk/working-paper/2017/machine-learning-at-central-banks>
- Cieslak, A., S. Hansen, M. McMahon, et al. (2021). *Policymakers' uncertainty*, Available at SSRN 3936999 <http://dx.doi.org/10.2139/ssrn.3936999>
- Corrado, D. (2013). *Brasile senza maschere: Politica, economia e società fuori dai luoghi comuni*, EGEA spa.

- Costa Filho, A. E., and F. Rocha (2010). “Como o mercado de juros futuros reage à comunicação do Banco Central?”, *Economia Aplicada*, 14, pp. 265-292. <https://www.scielo.br/j/ecoa/a/xs3tsZMKZVJxhYGGy9BwJ9N/?lang=pt>
- Costa Filho, A. E. (2014). “Incerteza e atividade econômica no Brasil”, *Economia Aplicada*, 18, pp. 421-453. <https://doi.org/10.1590/1413-8050/ea607>
- Fasolo, A. M., F. M. Graminho, and S. B. Bastos (2022). *Seeing the forest for the trees: Using hLDA models to evaluate communication in Banco Central do Brasil*, BIS Working Papers (1021). <https://www.bis.org/publ/work1021.htm>
- García-Herrero, A., E. Girardin, and E. dos Santos (2017). “Follow what I do, and also what I say: Monetary policy impact on Brazil’s financial markets”, *Economia*, 17(2), pp. 65-92. <https://ssrn.com/abstract=3160537>
- Gentzkow, M., B. Kelly, and M. Taddy (2019). “Text as data”, *Journal of Economic Literature*, 57(3), pp. 535-574. <https://doi.org/10.1257/jel.20181020>
- Godeiro, L. L., and L. R. R. de Oliveira Lima (2017). “Measuring macroeconomic uncertainty to Brasil”, *Economia Aplicada*, 21(2), pp. 311-334. <https://doi.org/10.11606/1413-8050/ea156958>
- Griffiths, T. L., and M. Steyvers (2004). “Finding scientific topics”, *Proceedings of the National Academy of Sciences*, 101(suppl. 1), pp. 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Hansen, S., and M. McMahon (2016). “Shocking language: Understanding the macroeconomic effects of central bank communication”, *Journal of International Economics*, 99, pp. S114-S133. <https://doi.org/10.1016/j.jinteco.2015.12.008>
- Hansen, S., M. McMahon, and A. Prat (2017). “Transparency and deliberation within the FOMC: A computational linguistics approach”, *The Quarterly Journal of Economics*, 133(2), pp. 801-870. <https://doi.org/10.1093/qje/qjx045>
- Hansen, S., M. McMahon, and M. Tong (2019). “The long-run information effect of central bank communication”, *Journal of Monetary Economics*, 108, pp. 185-202. <https://doi.org/10.1016/j.jmoneco.2019.09.002>
- Kilian, L., and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press. <https://doi.org/10.1017/9781108164818>
- Larsen, V. H., and L. A. Thorsrud (2019). “The value of news for economic developments”, *Journal of Econometrics*, 210(1), pp. 203-218. <https://doi.org/10.1016/j.jeconom.2018.11.013>
- Loughran, T., and B. McDonald (2011). “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”, *The Journal of Finance*, 66(1), pp. 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Mikolov, T., K. Chen, G. Corrado, et al. (2013). “Efficient estimation of Word representations in vector space”, *arXiv Preprint arXiv*, 1301.3781. <https://arxiv.org/abs/1301.3781>

- Moreno-Pérez, C., and M. Minozzo (2022a). *Monetary policy uncertainty in Mexico: An unsupervised approach*, Documentos de Trabajo, No. 2229, Banco de España. <http://dx.doi.org/10.2139/ssrn.4220416>
- Moreno-Pérez, C., and M. Minozzo (2022b). *Natural language processing and financial markets: Semi-supervised modelling of coronavirus and economic news*, Documentos de Trabajo, No. 2228, Banco de España. <http://dx.doi.org/10.2139/ssrn.4220516>
- Nodari, G. (2014). “Financial regulation policy uncertainty and credit spreads in the US”, *Journal of Macroeconomics*, 41, pp. 122-132. <https://doi.org/10.1016/j.jmacro.2014.05.006>
- Rosa, C., and G. Verga (2007). “On the consistency and effectiveness of central bank communication: Evidence from the ECB”, *European Journal of Political Economy*, 23(1), pp. 146-175. <https://doi.org/10.1016/j.ejpoleco.2006.09.016>
- Sims, C. A. (1980). “Macroeconomics and reality”, *Econometrica: Journal of the Econometric Society*, pp. 1-48. <https://doi.org/10.2307/1912017>
- Soto, P. E. (2021). “Breaking the word bank: Measurement and effects of bank level uncertainty”, *Journal of Financial Services Research*, 59(1), pp. 1-45. <https://doi.org/10.1007/s10693-020-00338-5>

Appendix

A Appendix

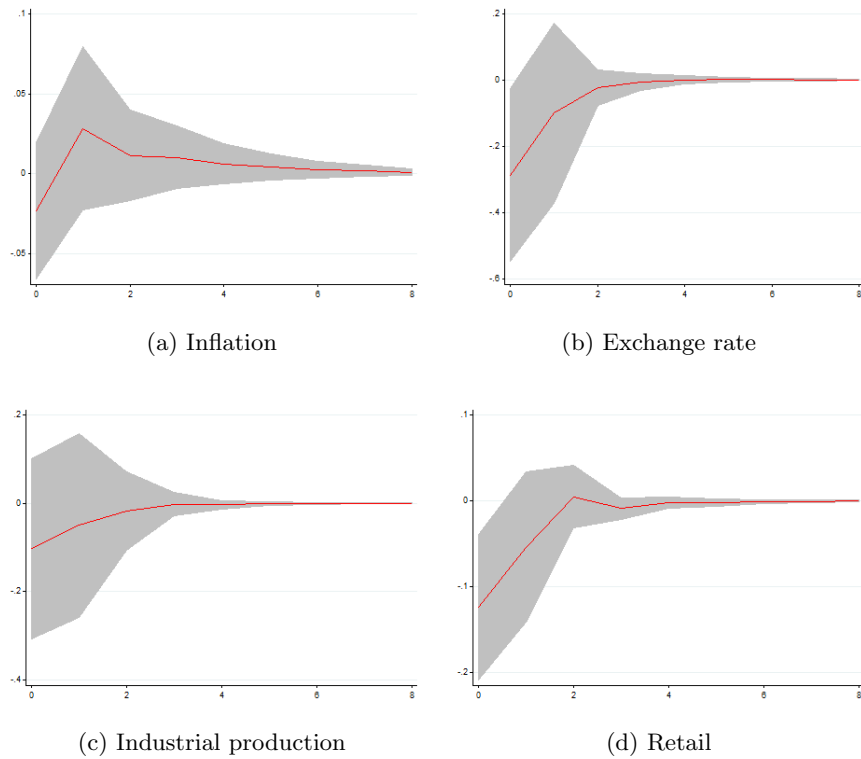


Fig. A.1: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the uncertainty index of the minutes of the COPOM from 2000 to July 2019. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points of each of the four macroeconomic variable and the X-axis represents time in months (8 months).

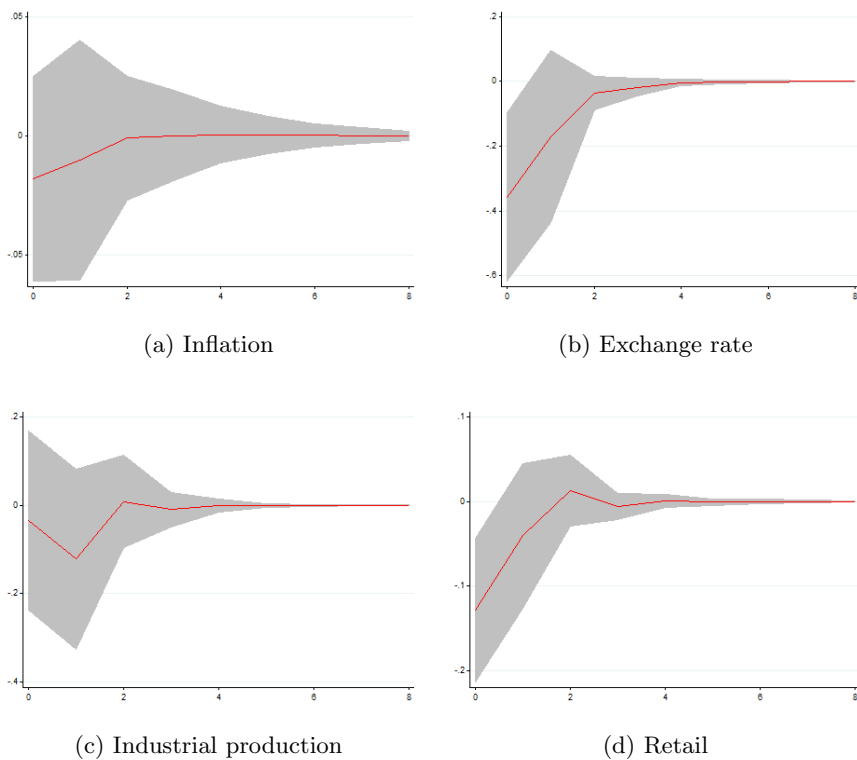


Fig. A.2: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index of the minutes of the COPOM from 2000 to July 2019. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points of each of the four macroeconomic variable and the X-axis represents time in months (8 months).

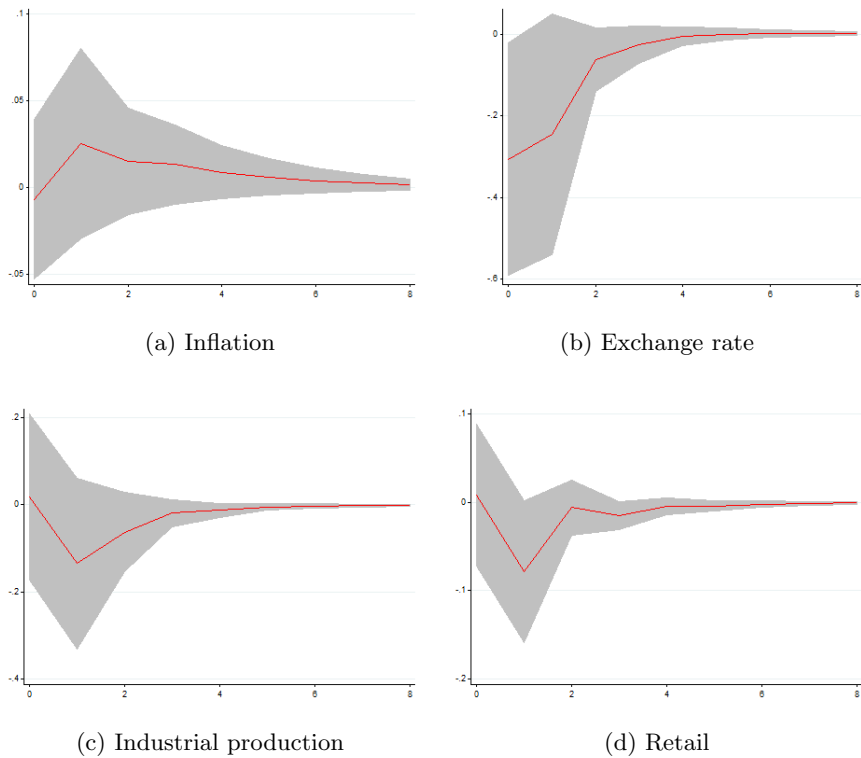
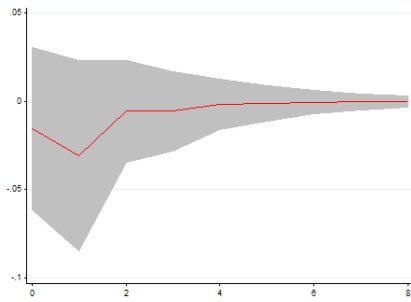
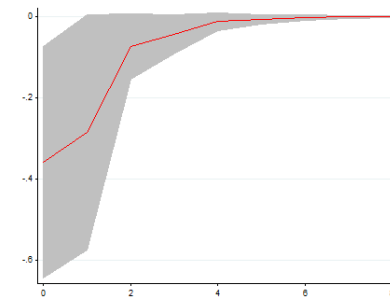


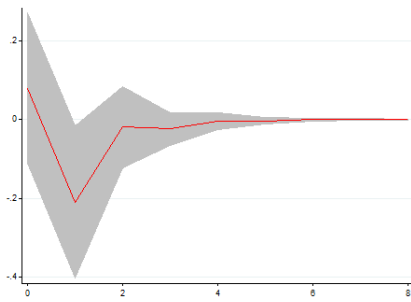
Fig. A.3: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the uncertainty index of the minutes of the COPOM from 2000 to June 2016. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points change for each one of the four macroeconomic variables and the X-axis represents time in months (8 months).



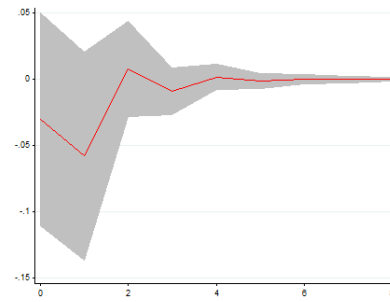
(a) Inflation



(b) Exchange rate



(c) Industrial production



(d) Retail

Fig. A.4: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the ‘inflation’ and ‘monetary policy decision’ topic-uncertainty index of the minutes of the COPOM from 2000 to June 2016. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points change for each one of the four macroeconomic variables and the X-axis represents time in months (8 months).

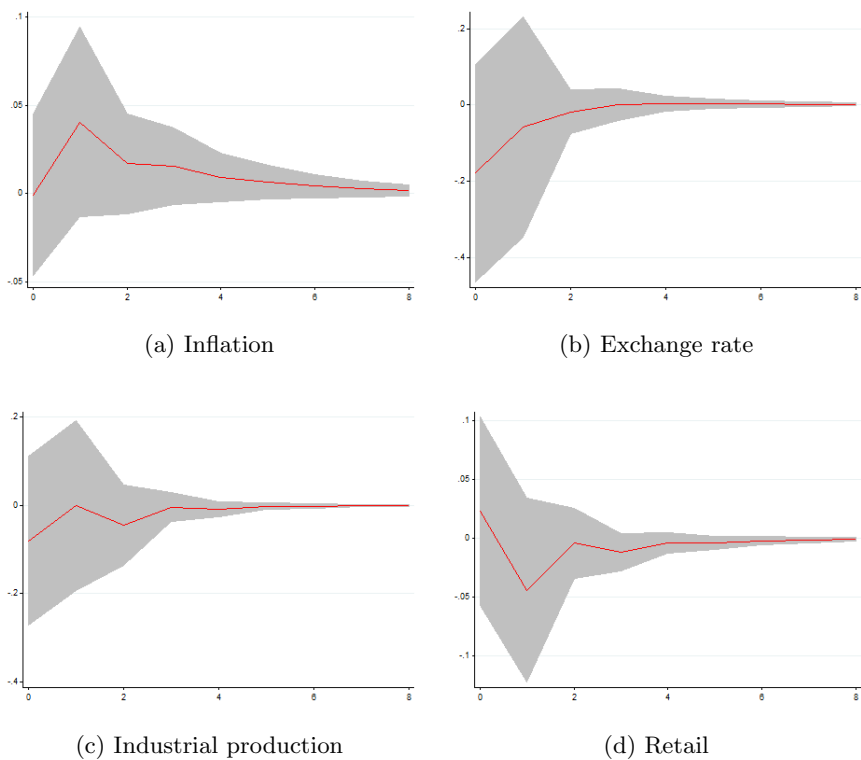


Fig. A.5: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the ‘general economic conditions’ topic-uncertainty index of the minutes of the COPOM from 2000 to June 2016. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y -axis is in % points change for each one of the four macroeconomic variables and the X -axis represents time in months (8 months).

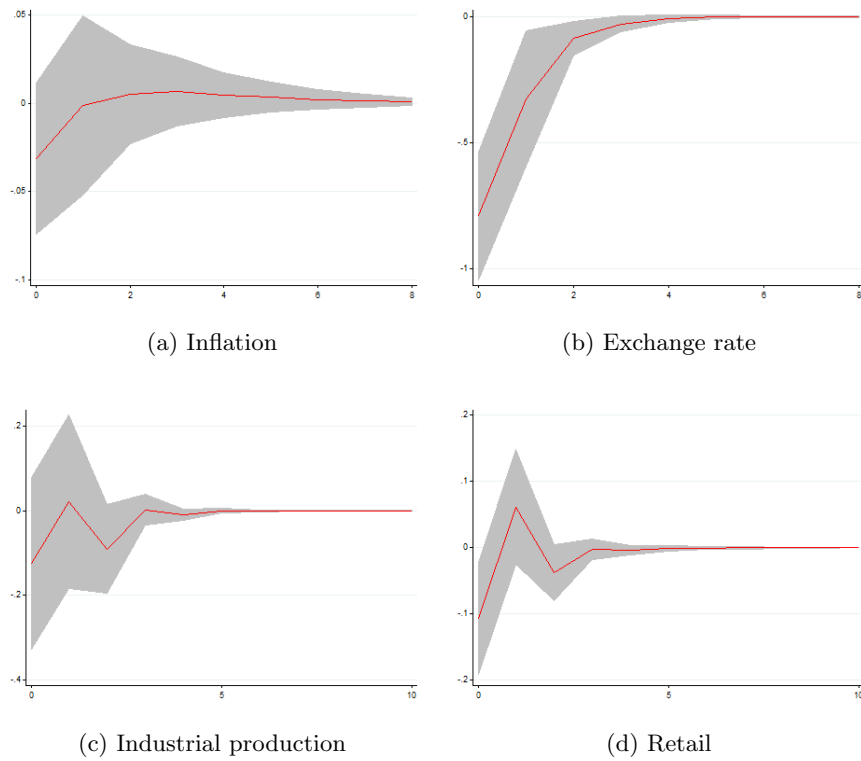


Fig. A.6: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the Economic Policy Uncertainty (EPU) index for Brazil created by Baker, Bloom, and Davis (2016) from 2000 to July 2019. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points change for each one of the four macroeconomic variables and the X-axis represents time in months (8 months).

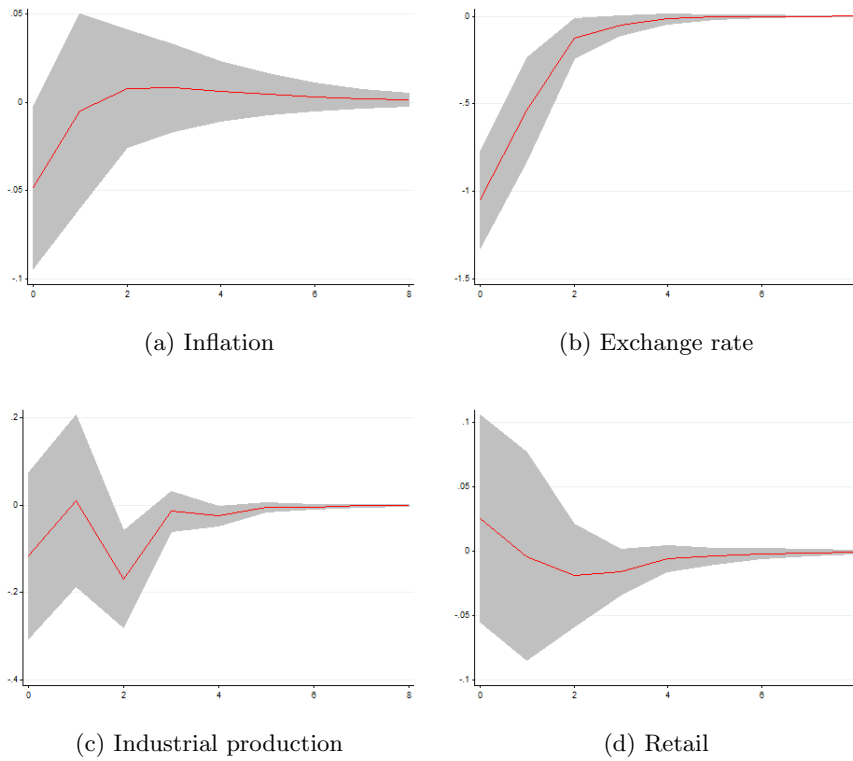


Fig. A.7: Impulse response functions from the Structural VAR model corresponding to one standard-deviation in the Economic Policy Uncertainty (EPU) index for Brazil created by Baker, Bloom, and Davis (2016) from 2000 to June 2016. The gray area displays the 90% confidence intervals computed using bootstrapped standard errors (200 replications). The Y-axis is in % points change for each one of the four macroeconomic variables and the X-axis represents time in months (8 months).

Table A.1: This table shows the first twelve words with the highest probability for each of the nine topics of the LDA results. A tag is included for each topic to provide a better understanding of the topic. These tags do not influence the results.

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11	Word 12
0. Inflation	price	twelv	chang	index	ipca	food	agricultur	accumul	di	compar	regul	reflect
	0.164	0.054	0.039	0.033	0.021	0.021	0.02	0.019	0.016	0.015	0.015	0.014
1. Inflation / COPOM	inflat	expect	core	measur	copom	last	futur	pressur	short	monetari	smooth	mean
	0.161	0.057	0.031	0.029	0.019	0.017	0.015	0.015	0.014	0.014	0.013	0.012
2. Economic activity	economi	econom	market	intern	activ	remain	recoveri	global	growth	despit	financi	continu
	0.042	0.024	0.024	0.022	0.018	0.017	0.017	0.016	0.015	0.015	0.014	0.013
3. COPOM meeting	rate	project	meet	scenario	consid	copom	interest	target	exchang	market	selic	inflat
	0.104	0.044	0.038	0.036	0.036	0.032	0.031	0.025	0.024	0.019	0.019	0.019
4. Trade / credit operations	billion	total	credit	oper	reach	averag	period	export	trade	matur	day	respect
	0.081	0.042	0.041	0.039	0.035	0.032	0.025	0.025	0.024	0.02	0.019	0.018
5. COPOM meetings	monetari	polic	committe	will	risk	demand	copom	effect	factor	econom	process	time
	0.034	0.029	0.025	0.022	0.018	0.017	0.017	0.015	0.014	0.013	0.011	0.01
6. Sales retails	quarter	sale	decreas	retail	accord	adjust	end	survey	index	data	growth	confid
	0.056	0.048	0.045	0.026	0.025	0.024	0.023	0.021	0.021	0.02	0.019	0.018
7. Employment	rate	employ	compar	indic	sector	real	accord	record	labor	reach	thousand	result
	0.029	0.027	0.027	0.026	0.025	0.025	0.023	0.021	0.018	0.017	0.017	0.017
8. Industrial production	product	industri	good	capit	adjust	consum	season	accord	durabl	manufactur	expans	decreas
	0.081	0.073	0.07	0.03	0.03	0.03	0.026	0.02	0.019	0.017	0.016	0.015

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

- 2120 ALEJANDRO FERNÁNDEZ-CEREZO, BEATRIZ GONZÁLEZ, MARIO IZQUIERDO and ENRIQUE MORAL-BENITO: Firm-level heterogeneity in the impact of the COVID-19 pandemic.
- 2121 EDUARDO GUTIÉRREZ and CÉSAR MARTÍN MACHUCA: The effect of tariffs on Spanish goods exports.
- 2122 JACOPO TIMINI: Revisiting the 'Cobden-Chevalier network' trade and welfare effects.
- 2123 ALEJANDRO BUESA, JAVIER J. PÉREZ and DANIEL SANTABÁRBARA: Awareness of pandemics and the impact of COVID-19.
- 2124 ALICIA AGUILAR and DIEGO TORRES: The impact of COVID-19 on analysts' sentiment about the banking sector.
- 2125 SILVIA ALBRIZIO, IVÁN KATARYNIUK, LUIS MOLINA and JAN SCHÄFER: ECB euro liquidity lines.
- 2126 ANTHONY BALD, ERIC CHYN, JUSTINE HASTINGS and MARGARITA MACHELETT: The causal impact of removing children from abusive and neglectful homes.
- 2127 IRMA ALONSO, PEDRO SERRANO and ANTONI VAELO-SEBASTIÀ: The impact of heterogeneous unconventional monetary policies on the expectations of market crashes.
- 2128 MARÍA T. GONZÁLEZ-PÉREZ: Lessons from estimating the average option-implied volatility term structure for the Spanish banking sector.
- 2129 SIMÓN A. RELLA, YULIYA A. KULIKOVA, EMMANOUIL T. DERMITZAKIS and FYODOR A. KONDRASHOV: Rates of SARS-COV-2 transmission and vaccination impact the fate of vaccine-resistant strains.
- 2130 MATÍAS LAMAS and DAVID MARTÍNEZ-MIERA: Sectorial holdings and stock prices: the household-bank nexus.
- 2131 ALBERT BANAL-ESTAÑOL, ENRIQUE BENITO, DMITRY KHAMETSHIN and JIANXING WEI: Asset encumbrance and bank risk: theory and first evidence from public disclosures in Europe.
- 2132 ISABEL ARGIMÓN and MARÍA RODRÍGUEZ-MORENO: Business complexity and geographic expansion in banking.
- 2133 LUIS GUIROLA: Does political polarization affect economic expectations?: Evidence from three decades of cabinet shifts in Europe.
- 2134 CHRISTIANE BAUMEISTER, DANILO LEIVA-LEÓN and ERIC SIMS: Tracking weekly state-level economic conditions.
- 2135 SERGI BASCO, DAVID LÓPEZ-RODRÍGUEZ and ENRIQUE MORAL-BENITO: House prices and misallocation: The impact of the collateral channel on productivity.
- 2136 MANUEL ARELLANO, STÉPHANE BONHOMME, MICOLE DE VERA, LAURA HOSPIDO and SIQI WEI: Income risk inequality: Evidence from Spanish administrative records.
- 2137 ANGELA ABBATE and DOMINIK THALER: Optimal monetary policy with the risk-taking channel.
- 2138 MARTA BAÑBURA, DANILO LEIVA-LEÓN and JAN-OLIVER MENZ: Do inflation expectations improve model-based inflation forecasts?
- 2139 MÁXIMO CAMACHO, MARÍA DOLORES GADEA and ANA GÓMEZ LOSCOS: An automatic algorithm to date the reference cycle of the Spanish economy.
- 2140 EDUARDO GUTIÉRREZ, AITOR LACUESTA and CÉSAR MARTÍN MACHUCA: Brexit: Trade diversion due to trade policy uncertainty.
- 2141 JULIO A. CREGO and JULIO GÁLVEZ: Cyclical dependence in market neutral hedge funds.
- 2142 HERVÉ LE BIHAN, MAGALI MARX and JULIEN MATHERON: Inflation tolerance ranges in the new keynesian model.
- 2143 DIEGO COMIN, JAVIER QUINTANA, TOM SCHMITZ and ANTONELLA TRIGARI: Measuring TFP: the role of profits, adjustment costs, and capacity utilization.
- 2144 ROBERTO PASCUAL: Do analysts forecast differently in periods of uncertainty? An empirical analysis of target prices for Spanish banks.
- 2145 BEATRIZ GONZÁLEZ, GALO NUÑO, DOMINIK THALER and SILVIA ABRIZIO: Firm heterogeneity, capital misallocation and optimal monetary policy.
- 2201 RYAN BANERJEE and JOSÉ-MARÍA SERENA: Dampening the financial accelerator? Direct lenders and monetary policy.
- 2202 JUAN S. MORA-SANGUINETTI and ISABEL SOLER: La regulación sectorial en España. Resultados cuantitativos.
- 2203 JORGE E. GALÁN, MATÍAS LAMAS and RAQUEL VEGAS: Roots and recourse mortgages: handing back the keys.
- 2204 MONICA MARTINEZ-BRAVO and CARLOS SANZ: Inequality and psychological well-being in times of COVID-19: evidence from Spain.
- 2205 CORINNA GHIRELLI, DANILO LEIVA-LEÓN and ALBERTO URTASUN: Housing prices in Spain: convergence or decoupling?
- 2206 MARÍA BRU MUÑOZ: Financial exclusion and sovereign default: The role of official lenders.
- 2207 RICARDO GIMENO and CLARA I. GONZÁLEZ: The role of a green factor in stock prices. When Fama & French go green.

- 2208 CARLOS MONTES-GALDÓN and EVA ORTEGA: Skewed SVARs: tracking the structural sources of macroeconomic tail risks.
- 2209 RODOLFO G. CAMPOS, ILIANA REGGIO and JACOPO TIMINI: Thick borders in Franco's Spain: the costs of a closed economy.
- 2210 MARCO CELENTANI, MIGUEL GARCÍA-POSADA and FERNANDO GÓMEZ POMAR: Fresh start policies and small business activity: evidence from a natural experiment.
- 2211 JOSE GARCIA-LOUZAO, LAURA HOSPIDO and ALESSANDRO RUGGIERI: Dual returns to experience.
- 2212 ADRIÁN CARRO and PATRICIA STUPARIU: Uncertainty, non-linear contagion and the credit quality channel: an application to the Spanish interbank market.
- 2113 MARIO IZQUIERDO, ENRIQUE MORAL-BENITO, ELVIRA PRADES and JAVIER QUINTANA: The propagation of worldwide sector-specific shocks.
- 2214 HENRIQUE S. BASSO: Asset holdings, information aggregation in secondary markets and credit cycles.
- 2215 JOSÉ MANUEL CARBÓ and SERGIO GORJÓN: Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin.
- 2216 LUIS GUIROLA and MARÍA SÁNCHEZ-DOMÍNGUEZ: Childcare constraints on immigrant integration.
- 2217 ADRIÁN CARRO, MARC HINTERSCHWEIGER, ARZU ULUC and J. DOYNE FARMER: Heterogeneous effects and spillovers of macroprudential policy in an agent-based model of the UK housing market.
- 2218 STÉPHANE DUPRAZ, HERVÉ LE BIHAN and JULIEN MATHERON: Make-up strategies with finite planning horizons but forward-looking asset prices.
- 2219 LAURA ÁLVAREZ, MIGUEL GARCÍA-POSADA and SERGIO MAYORDOMO: Distressed firms, zombie firms and zombie lending: a taxonomy.
- 2220 BLANCA JIMÉNEZ-GARCÍA and JULIO RODRÍGUEZ: A quantification of the evolution of bilateral trade flows once bilateral RTAs are implemented.
- 2221 SALOMÓN GARCÍA: Mortgage securitization and information frictions in general equilibrium.
- 2222 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Accuracy of explanations of machine learning models for credit decisions.
- 2223 JAMES COSTAIN, GALO NUÑO and CARLOS THOMAS: The term structure of interest rates in a heterogeneous monetary union.
- 2224 ANTOINE BERTHEAU, EDOARDO MARIA ACABBI, CRISTINA BARCELÓ, ANDREAS GULYAS, STEFANO LOMBARDI and RAFFAELE SAGGIO: The Unequal Consequences of Job Loss across Countries.
- 2225 ERWAN GAUTIER, CRISTINA CONFLITTI, RIEMER P. FABER, BRIAN FABO, LUDMILA FADEJEVA, VALENTIN JOUVANCEAU, JAN-OLIVER MENZ, TERESA MESSNER, PAVLOS PETROULAS, PAU ROLDAN-BLANCO, FABIO RUMLER, SERGIO SANTORO, ELISABETH WIELAND and HÉLÈNE ZIMMER: New facts on consumer price rigidity in the euro area.
- 2226 MARIO BAJO and EMILIO RODRÍGUEZ: Integrating the carbon footprint into the construction of corporate bond portfolios.
- 2227 FEDERICO CARRIL-CACCIA, JORDI PANIAGUA and MARTA SUÁREZ-VARELA: Forced migration and food crises.
- 2228 CARLOS MORENO PÉREZ and MARCO MINOZZO: Natural Language Processing and Financial Markets: Semi-supervised Modelling of Coronavirus and Economic News.
- 2229 CARLOS MORENO PÉREZ and MARCO MINOZZO: Monetary Policy Uncertainty in Mexico: An Unsupervised Approach.
- 2230 ADRIAN CARRO: Could Spain be less different? Exploring the effects of macroprudential policy on the house price cycle.
- 2231 DANIEL SANTABÁRBARA and MARTA SUÁREZ-VARELA: Carbon pricing and inflation volatility.
- 2232 MARINA DIAKONOVA, LUIS MOLINA, HANNES MUELLER, JAVIER J. PÉREZ and CRISTOPHER RAUH: The information content of conflict, social unrest and policy uncertainty measures for macroeconomic forecasting.
- 2233 JULIAN DI GIOVANNI, MANUEL GARCÍA-SANTANA, PRIIT JEENAS, ENRIQUE MORAL-BENITO and JOSEP PIJOAN-MAS: Government Procurement and Access to Credit: Firm Dynamics and Aggregate Implications.
- 2234 PETER PAZ: Bank capitalization heterogeneity and monetary policy.
- 2235 ERIK ANDRES-ESCAIOLA, CORINNA GHIRELLI, LUIS MOLINA, JAVIER J. PÉREZ and ELENA VIDAL: Using newspapers for textual indicators: which and how many?
- 2236 MARÍA ALEJANDRA AMADO: Macroprudential FX regulations: sacrificing small firms for stability?
- 2237 LUIS GUIROLA and GONZALO RIVERO: Polarization contaminates the link with partisan and independent institutions: evidence from 138 cabinet shifts.
- 2238 MIGUEL DURO, GERMÁN LÓPEZ-ESPINOSA, SERGIO MAYORDOMO, GAIZKA ORMAZABAL and MARÍA RODRÍGUEZ-MORENO: Enforcing mandatory reporting on private firms: the role of banks.
- 2239 LUIS J. ÁLVAREZ and FLORENS ODENDAHL: Data outliers and Bayesian VARs in the Euro Area.
- 2240 CARLOS MORENO PÉREZ and MARCO MINOZZO: "Making text talk": The minutes of the Central Bank of Brazil and the real economy.