

# Mensurando o Sentimento de Incerteza da Política Econômica: Uma Análise a Partir da Comunicação do Banco Central do Brasil

Pierre Hítalo Nascimento Silva\*      Dr. Cássio da Nóbrega Besarria†  
Dr<sup>a</sup>. Maria Daniella de Oliveira Pereira da Silva‡

Julho, 2019

## RESUMO

Este artigo tem como propósito criar um índice capaz de mensurar o grau de incerteza da economia no Brasil. Esse índice será construído a partir da estimação do sentimento textual contido nas atas de reuniões do Comitê de Política Monetária (Copom) considerando o período de janeiro de 2000 a dezembro de 2018. Posteriormente é analisado como um choque de incerteza afeta a dinâmica de um conjunto de variáveis macroeconômicas por meio de um Modelo de Vetores Autorregressivos (VAR) com restrição de sinais. Conclui-se que a incerteza afeta de forma negativa a atividade econômica e que é possível estabelecer uma ligação entre os resultados advindos da análise do choque de incerteza e o desempenho da economia brasileiros nos últimos anos.

**Palavras-chaves:** Incerteza. Copom. Análise de sentimento. VAR.

**Classificação JEL:** C83, D8, E32

**Área 4-** Macroeconomia, Economia Monetária e Finanças

## ABSTRACT

This article aims to create an index capable of measuring the degree of uncertainty of the economy in Brazil. This index will be constructed based on the estimation of the textual sentiment contained in the minutes of meetings of the Monetary Policy Committee (Copom) considering the period from January 2000 to December 2018. It is later analyzed how a shock of uncertainty affects the dynamics of a set of macroeconomic variables by means of a Signal Restricted Autoregressive Vector Model (VAR). It is concluded that uncertainty negatively affects economic activity and that it is possible to establish a link between the results of the analysis of the shock of uncertainty and the performance of the Brazilian economy in recent years.

**Key-words:** Uncertainty. Copom. Sentiment analysis. VAR.

**JEL Codes:** C83, D8, E32,

---

\*Mestrando em Economia Universidade Federal da Paraíba (PPGE/UFPB) e estudante do Laboratório de Inteligência Artificial e Macroeconomia Computacional (LABIMEC). <[pierrehitalo@gmail.com](mailto:pierrehitalo@gmail.com)>

†Professor do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba (UFPB) e coordenador do Laboratório de Inteligência Artificial e Macroeconomia Computacional (LABIMEC). <[cassiodanobrega@yahoo.com.br](mailto:cassiodanobrega@yahoo.com.br)>

‡Professora Assistente do Departamento de Gestão Pública da Universidade Federal da Paraíba (UFPB) e coordenadora do Laboratório de Inteligência Artificial e Macroeconomia Computacional (LABIMEC). <[mariadaniella75@hotmail.com](mailto:mariadaniella75@hotmail.com)>

# 1 INTRODUÇÃO

Risco e incerteza por muito tempo foram tratados como sinônimos pela teoria econômica, [Knight \(1921\)](#) foi o primeiro autor a propor a separação desses conceitos ao apresentar uma definição desagregada. O risco é caracterizado pela capacidade de se atribuir probabilidade *a priori* a um determinado evento mesmo que não seja possível visualizar o resultado futuro<sup>1</sup>, que é passível de modelagem. A incerteza, então, diferencia-se do risco pelo fato de não ser possível atribuir probabilidade aos eventos futuros, as decisões são tomadas a partir da influência do ambiente para com o agente econômico, sendo assim, é impossível modelar o conjunto de resultados possíveis/esperados como acontece sob o cenário de risco. As economias vivendo sobre esse cenário apresentam eventos que não conseguem ser explicados, por exemplo, a inabilidade de recuperação apresentadas por alguns países mesmo após as medidas de ajustes necessárias terem sido tomadas.

A discussão sobre os fatores geradores da incerteza e suas consequências não é recente, mas novas formas de lidar com esse componente sempre estão em evidência. Algumas pesquisas empíricas têm direcionado o foco para o estudo das consequências trazidas pela incerteza, principalmente em relação ao efeito psicológico dos agentes econômicos. [Alchian \(1950\)](#) argumenta que sob sua presença, julgamentos e opiniões serão divergentes, mesmo embasados nas melhores evidências disponíveis. [Filho \(2014\)](#) encontra evidências que a incerteza apresenta um comportamento anticíclico no Brasil e que os períodos prolongados de incerteza são caracterizados por uma atividade econômica deprimida e, conseqüentemente, uma fase de recessão mais duradoura. A junção desses fatores dificulta a capacidade preditiva dos agentes econômicos e autoridades, fazendo com que esses retardem tomadas de decisões importantes ou as façam de forma incorreta.

A presença de incerteza em uma determinada economia traz como consequência o adiamento das decisões econômicas em função do seu caráter irreversível e do alto grau de volatilidade, expectativas negativas acabam sendo formadas e empilhadas dentro desse contexto, além de transcenderem de um período para o outro<sup>2</sup>, a longo prazo é criada uma memória pessimista que afeta suas escolhas. A expansão duradoura desses choques modifica a estrutura econômica doméstica que fica condicionada a essa nova conjuntura. No que se refere as diretrizes da política econômica, a incapacidade de projeção do cenário econômico de forma clara acarreta na postergação de decisões estratégicas sobre diversos setores.

A incerteza trata-se de uma informação importante na qual ajuda a projetar a tomada de decisões a respeito do futuro, choques de incerteza acarretam em impactos negativos perante a política monetária, tornando-a mais sensível e/ou comprometendo sua eficiência devido ao seu efeito sobre os agentes econômicos [Schymura \(2017\)](#). No que se refere as diretrizes da política econômica, a incapacidade de projeção do cenário econômico de forma clara acarreta na postergação de decisões estratégicas sobre diversos setores.

O principal problema consiste no fato da incerteza não se tratar de uma variável que possa ser quantificada de maneira usual, como visto por [Knight \(1921\)](#) ela não apresenta nenhuma distribuição de probabilidade que permita observar e modelar o seu comportamento. Então, como mensurá-la? Uma das soluções é a estimação do sentimento<sup>3</sup>, um método alternativo que viabiliza a extração dessa informação ao tentar captá-la mediante a ferramentas de estimação do sentimento textual. Esse sentimento é obtido a partir da análise textual com o auxílio de um dicionário específico, é possível polarizar os termos. Essa polarização exprime o grau de positividade, negatividade, incerteza ou neutralidade de uma palavra ou frase, apoiado nessas informações e aplicando métricas que calculam o peso de cada palavra é possível determinar o sentimento textual [You, DesArmo e Joo \(2013\)](#), [Hutto e Gilbert \(2014\)](#), [Graminho et al. \(2015\)](#), [Nopp e Hanbury \(2015\)](#), [Ormerod, Nyman e Tuckett \(2015\)](#), [Silva et al. \(2017\)](#), [Machado, Silva et al. \(2017\)](#), [Aguilar et al. \(2018\)](#), [Li et al. \(2019\)](#).

Os métodos de *text regression* permitem quantificar palavras e frases de modo a contribuir com

<sup>1</sup> Tome como exemplo o lançamento de um dado, mesmo sem saber qual será o resultado é possível atribuir a probabilidade de ocorrência para cada resultado provável.

<sup>2</sup> Considerando a natureza adaptativa das expectativas.

<sup>3</sup> Em alguns estudos a incerteza é medida pela volatilidade [Jurado, Ludvigson e Ng \(2015\)](#), em outros utilizam volatilidade e estimação do sentimento como em [Barboza e Zilberman \(2018\)](#).

previsão, por exemplo, de variáveis financeiras. Neste ensaio, aplicá-se este método com o intuito de estimar o sentimento de incerteza contido nas palavras a partir de bases textuais. [Kearney e Liu \(2014\)](#) enfatizam a existência de muitas fontes de informações prontas para serem “coletadas”: documentos, reportagens, *blogs*, discursos públicos etc. Destaca-se o emprego do dicionário que permite averiguar o tom contido na escrita concomitantemente com o aprendizado de máquina que promove a conversão de texto em *vector space model* através do uso de matrizes.

A Linguagem de Processamento Natural viabiliza a transformação dos dados qualitativos em valores numéricos por meio de informações estruturadas sob a forma de matriz, facilitando sua incorporação em estudos econômicos através de modelos econométricos. A base de dados é organizada em uma matriz (A), cada linha dessa matriz representa uma palavra (i) e cada coluna representa um documento (j), assim, cada célula reflete a frequência ponderada ( $a_{ij}$ ) do termo “i” no documento “j” [Hendry e Madeley \(2010\)](#). O processo de extração de informações como sentimento é possível por conta de aprendizagem de máquina baseadas em dicionários.

O sentimento está diretamente atrelado à escrita, ao selecionar um conjunto de palavras para analisar, consegue-se determinar o seu tom<sup>4</sup>. A partir da tonalidade gerada e mediante o cálculo do peso de cada palavra é possível obter o parâmetro que expressa o sentimento requerido. Com base nos resultados alcançados o pesquisador é capaz de fazer uso dessa informação para derivar diferentes indicadores de sentimento: otimismo, pessimismo, felicidade, incerteza etc., conforme a necessidade do seu problema de pesquisa. Desenvolveu-se assim, um instrumento de estudo dos padrões de conduta dos agentes econômicos, no qual as respostas obtidas são utilizadas com o intuito de acalmar os ânimos individuais ao fornecer explicações sobre o comportamento da variável analisada, diagnósticos, previsões etc.

A eficiência dessa técnica foi demonstrada pelo [Correa et al. \(2017\)](#), que construiu um índice capaz de verificar o sentimento de estabilidade financeira a partir dos relatórios de diversos Bancos Centrais. [Machado, Silva et al. \(2017\)](#) encontram evidências que o tom de incerteza contido nos relatórios de desempenho trimestrais das indústrias influencia a dinâmica do mercado financeiro. [Aarle e Moons \(2017\)](#) ao analisar o ajuste econômico ocorrido na Europa após a Crise Financeira Global e a Crise da Zona do Euro, concluem que os fatores psicológicos possuem efeitos não negligenciáveis na explicação dos ciclos de negócios.

Contudo, estimar o sentimento “transcrito” ainda se trata de uma metodologia recente e em pleno desenvolvimento, novas ferramentas que buscam otimizar cada vez mais o processo continuam sendo incorporadas. No Brasil o status de avanço da temática ainda é embrionário, o leque de trabalhos que fazem uso desse procedimento ainda é pequeno e se torna ainda mais restrito quando se direciona para a incerteza, podemos destacar alguns trabalhos como: [Filho \(2014\)](#), [Ferreira et al. \(2017\)](#), [Barboza e Zilberman \(2018\)](#). [Godeiro e Lima \(2017\)](#) criaram um Índice de incerteza Macroeconômica agregada para o Brasil a partir da volatilidade, seguindo a metodologia de [Jurado, Ludvigson e Ng \(2015\)](#).

Baseado no arcabouço literário internacional acerca da mensuração da incerteza e dos seus efeitos sobre a economia, [Ferreira et al. \(2017\)](#) desenvolveu o Indicador de Incerteza Econômica – Brasil (IIE-Br), divulgado mensalmente pela Fundação Getúlio Vargas. Esse indicador apresenta forte relação com os grandes momentos de incerteza vivido pelo país nos últimos anos, após um estudo econométrico o autor chega a conclusão de que choques de incerteza produzem efeitos negativos sobre a atividade econômica e produção industrial<sup>5</sup>.

A incerteza da política econômica foi observada pela primeira vez em [Baker, Bloom e Davis \(2016\)](#)<sup>6</sup>, gerou-se um indicador capaz de captar esse sentimento a partir de artigos de jornais baseado na frequência de palavras associadas a três categorias de palavras: incerteza, política e econômica. Os autores identificaram uma relação positiva entre incerteza e volatilidade, além de uma relação negativa entre ela e o investimento, ou seja, fortes indícios indicam a sua capacidade de afetar variáveis macroeconômicas.

<sup>4</sup> Neste trabalho, sentimento e tom são tratados como sinônimos

<sup>5</sup> [Ferreira et al. \(2017\)](#) utilizaram as variáveis: IBC-Br e PIM-PF para captar esses efeitos através de um modelo VAR.

<sup>6</sup> Os autores são referências nessa temática e atualmente possuem índices de incerteza para diversos países.

A ata do Copom, nacionalmente e internacionalmente, é um dos principais instrumentos de comunicação utilizado pelo BACEN ao apresentar dados macroeconômicos relevantes, projeções econômicas a respeito da conjuntura doméstica e externa, além de detalhar os procedimentos da tomada de decisão em relação a política monetária. Optou-se por trabalhar com elas com o objetivo de desenvolver um índice capaz de mensurar o sentimento de incerteza no Brasil.

As comunicações dos Bancos Centrais surgem como uma fonte promissora de análise do sentimento textual sua transparência é vista como fundamental para controlar as expectativas dos agentes econômicos, [Montes e Nicolay \(2015\)](#) realizam um estudo empírico sobre a influência da comunicação do Banco Central do Brasil sobre as expectativas de inflação por meio das atas das reuniões do COPOM obtém-se indícios de que os especialistas reagem ao “tom” empregado nelas<sup>7</sup> comprovando assim, a eficiência desse instrumento de comunicação utilizado pelo Bacen. Destacam-se alguns autores que realizam estudos empíricos sobre a comunicação dos Bancos Centrais: [Apel e Grimaldi \(2012\)](#), [Bholat et al. \(2015\)](#), [Bruno \(2016\)](#), [Bruno \(2017\)](#), [Grignani e Fontana \(2018\)](#), [Rybinski \(2018\)](#). Na esfera da análise baseada em documentos envolvendo Autoridades Monetárias, [Hansen e McMahon \(2016\)](#) exploram como as informações divulgadas pelo *Federal Open Market Committee* têm efeitos sobre as variáveis reais e de mercado.

Este trabalho tem como objetivo contribuir com a literatura em desenvolvimento no país no tocante referente a análise de sentimento textual. Foi criado o Índice de Incerteza da Política Econômica (IIPE) extraído a partir das atas do Copom, que apresentou resultados consistentes com os momentos de incerteza vivido pelo país nas duas últimas décadas. Posteriormente, como primeiro exercício observou o impacto de choques de incerteza sobre um conjunto de variáveis macroeconômicas através de um Modelo de Vetores Autorregressivos (VAR) com restrições de sinais proposto por [Uhlig \(2005\)](#). Ademais, faz-se necessário fornecer aos tomadores de decisões e formuladores de política econômica as melhores informações visando diminuir o grau de assimetria com relação a incerteza, possibilitando escolhas eficiente por ambos.

## 2 DADOS

Os dados utilizados foram coletados através do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil, Confederação Nacional da Indústria (CNI), Fundação Getulio Vargas (FGV), Instituto Brasileiro de Geografia e Estatística (IBGE) e Instituto de Pesquisa Econômica Aplicada (IPEA). Todas as informações relacionadas às atas do Copom foram reunidas diretamente da base de dados compreendida no portal do BACEN, o critério de seleção levou em conta a disponibilidade das atas em inglês, por isso optou-se por trabalhar com os documentos alusivos ao período janeiro de 2000 a dezembro de 2018.

Deve se levar em conta uma mudança referente a periodicidade de publicação, durante o período 2000 a 2005 as reuniões ocorriam de forma mensal assim como as publicações, a partir do ano de 2006 as reuniões passaram a acontecer a cada quarenta e cinco dias, sendo divulgadas oito atas ao ano. Outro ponto que merece ser destacado é o fato de que em 2002 foram publicadas treze atas ao invés de doze. O índice após ser estimado precisou ser convertido em dados trimestrais com o objetivo de que toda a amostra passasse a ter a mesma frequência.

A ata do Copom é um dos principais instrumentos de comunicação do Banco Central do Brasil, apresenta projeções econômicas para o cenário nacional e internacional, controle da inflação, decisões a respeito dos juros etc. É através dela que a autoridade monetária explica os procedimentos utilizados para a tomada de decisão de política monetária com o objetivo de tornar a comunicação mais transparente e manter sob controle as expectativas [Filho e Rocha \(2010\)](#).

Desse modo, este trabalho se propôs a criar o Índice de Incerteza da Política Econômica (IIPE) derivado a partir das atas do Copom. Em posse desse índice, optou por aplicá-lo em um modelo VAR com

<sup>7</sup> [Montes e Nicolay \(2015\)](#) criaram um indicador de percepção da inflação, no qual mostra que os especialistas reagem de forma negativa ao tom empregado do indicador. Se o indicador demonstra que a inflação está sob controle, as expectativas de inflação são reduzidas e *vice-versa*.

restrições de sinais com o objetivo de observar a resposta de um conjunto de variáveis macroeconômicas a um choque de incerteza. A série temporal criada nesse estudo considerou o período de 01 de janeiro de 2000 a 31 de dezembro de 2018, o que totalizou uma amostra de 177 atas, que após transformação de periodicidade em valores trimestrais resultou em 76 observações.

## 2.1 Base de dados

A base de dados utilizada nesta pesquisa é composta por observações trimestrais, no período que compreende janeiro de 2000 a dezembro de 2018. A taxa Selic Meta é a principal taxa de juros da economia brasileira é regulada pelo Copom, o Índice de Preços ao Consumidor é utilizado como medida de inflação, mede a variação de preços de um conjunto fixo de bens cobrindo a pesquisa cobre as principais capitais do país: São Paulo, Rio de Janeiro, Belo Horizonte, Salvador, Recife, Porto Alegre e Brasília. A variável Horas trabalhadas na indústria é um indicador de desempenho que visa medir a evolução da produção industrial, a Formação Bruta de Capital Fixo mostra o quanto as empresas aumentaram os seus bens de capital, atuando como indicador que mensura o investimento. A variável consumo trata-se da soma das despesas de consumo privada e pública, o Produto Interno Bruto considerado foi a preços correntes. O spread devido à falta de dados que abrangesse toda a amostra, utilizou como medida a diferença entre taxa Selic e o CDI, o saldo desse indicador será a medida de crédito incorporada neste trabalho. A Tabela 1 mostra as variáveis por código, descrição e fonte.

Tabela 1 – Descrição dos dados

Variável	Descrição	Fonte
Selic	Taxa de juros - Meta Selic definida pelo Copom	Bacen
IPC	Índice de preços ao consumidor-Brasil	FGV
Horas	Horas trabalhadas na produção industrial	CNI
FBCF	Formação bruta de capital fixo	Ipea
Consumo	Consumo privado + consumo do governo	IBGE
PIB	Produto interno Bruto a preços correntes em milhões	IBGE
Spread	Obtido mediante a diferença entre a Selic e o CDI	Bacen

Fonte: Elaboração própria

Destaca-se que neste artigo a seleção das variáveis, assim como a definição da ordem de inclusão das mesmas no modelo VAR seguiu a proposta do trabalho do Redl (2017).

## 3 ESTIMAÇÃO DO SENTIMENTO TEXTUAL

A Linguagem de Processamento Natural permite uma interação entre humano e máquina, através dela é possível determinar os parâmetros que iriam guiar a máquina e permitir a conversão do texto em um *vector space model*<sup>8</sup> que recebeu as palavras (*strings*). A análise das atas do Comitê de Política Monetária demandou a utilização de algumas técnicas específicas de manipulação e mineração textual, de modo a proporcionar que toda a informação qualitativa contida nos documentos seja extraída e posteriormente transformada em números.

É possível encontrar na literatura algumas metodologias de como estimar o peso das palavras, por exemplo, Salton e Buckley (1988) utilizam o *vector space model* a partir de uma técnica de normalização dos termos para documentos de diferentes tamanhos visando obter melhores resultados referente ao processo de minimização de *outliers* dentro do vetor corpus<sup>9</sup>.

Neste trabalho foi escolhida a abordagem proposta por Chisholm e Kolda (1999), na qual é aplicada a ponderação nas frequências de palavras com o objetivo de maximizar a eficiência dos

<sup>8</sup> Representa documentos de texto na forma de vetores de termos, além de permitir modelagem.

<sup>9</sup> O Corpus se trata de uma coletânea restrita de documentos linguísticos manuseados como objeto de estudo que permite modelagem a partir de ferramentas computacionais, para maiores esclarecimentos consultar Aluísio e Almeida (2006)

resultados alcançados pelo *vector space model*. O desempenho deste modelo está associado ao esquema de ponderação, fez-se o uso dos logaritmos com o intuito de ajustar melhor a frequência dos termos em cada ata, impedindo que termos com maiores frequências (*outliers*) venham a ter um peso maior dentro do documento Chisholm e Kolda (1999). A estimação do peso das palavras é representada pela equação (1):

$$P_{i,j} \begin{cases} \frac{(1+\log(Tf_{i,j}))}{(1+\log(a_j))} x \log \frac{N}{df_i}, & \text{se } Tf_{i,j} \geq 1 \\ 0, & \text{se } Tf_{i,j} = 0 \end{cases} \quad (1)$$

em que  $P_{i,j}$  representa o peso da palavra  $i$  no documento  $j$ ;  $Tf_i$  é o total de ocorrências de uma palavra  $i$  em um documento  $j$ ;  $a_j$  é a média<sup>10</sup> de frequência das palavras de incerteza presente em um documento  $j$ ;  $N$  é o total de documentos na amostra e;  $df_i$  representa o total de documentos com ao menos uma ocorrência da palavra  $i$ .

O peso de cada termo pode ser calculado mediante ponderação aplicada a frequência das palavras em cada ata, pode-se também dividir em duas partes a equação (1). A primeira metade expressa por:  $\frac{1+\log Tf_{i,j}}{1+\log a_j}$ , representa a parte local da expressão, na qual retrata o peso que a palavra  $i$  tem sobre o documento  $j$ , o termo  $1 + \log(a_j)$  está normalizando<sup>11</sup> a expressão evitando discrepância entre palavras com frequências muito distintas dentro do mesmo corpo textual, essa solução de normalização é específica para o *vector space model*. A segunda metade da equação é representada por  $\log \frac{N}{df_i}$  e esse termo denota o peso global que a palavra  $i$  tem sobre todo o conjunto de palavras de incerteza.

O sentimento textual foi então estimado com base no peso de cada palavras de incerteza contida nas atas de reuniões do Copom, utilizou-se da abordagem de dicionário através de uma lista de palavras de incerteza previamente determinada. O cálculo do sentimento individual de incerteza da ata é dado por:

$$SI_j = \sum \text{Peso das palavras de incerteza} \quad (2)$$

em que o termo  $SI_j$  é o sentimento textual de incerteza da ata  $j$ .

A estimação do Índice de Incerteza da Política Econômica se deu mediante o agrupamento desse sentimento individual gerado por cada ata a partir de (2), ao associá-los obteve-se a série temporal que permite observar o comportamento da incerteza no Brasil.

### 3.1 PROCEDIMENTO DE ESTIMAÇÃO TEXTUAL

A Figura (1) representa um fluxograma no qual é descrito todo o processo automatizado<sup>12</sup> desta pesquisa. O processo inclui desde a coleta dos dados até o cálculo do sentimento e, posteriormente, a estimação do Índice de Incerteza da Política Econômica, que foi derivado das atas do COPOM. Optou-se por captar manualmente todos os arquivos necessários em PDF diretamente do site do BACEN e então criou-se o banco de dados para trabalho. O objetivo é organizar a base de dados de modo que o acesso aos documentos seja facilitado, além de ficar disponível *off-line* para aplicações. Todas as etapas de manuseio e tratamento das atas do Copom foram realizadas a partir do *software* R.

No fluxograma os *templates* atuam como uma etapa de pré-ajuste, também conhecida como conjunto de regras (remoção de espaços duplos, pontuação, números, quebra de linha, quebra de página,

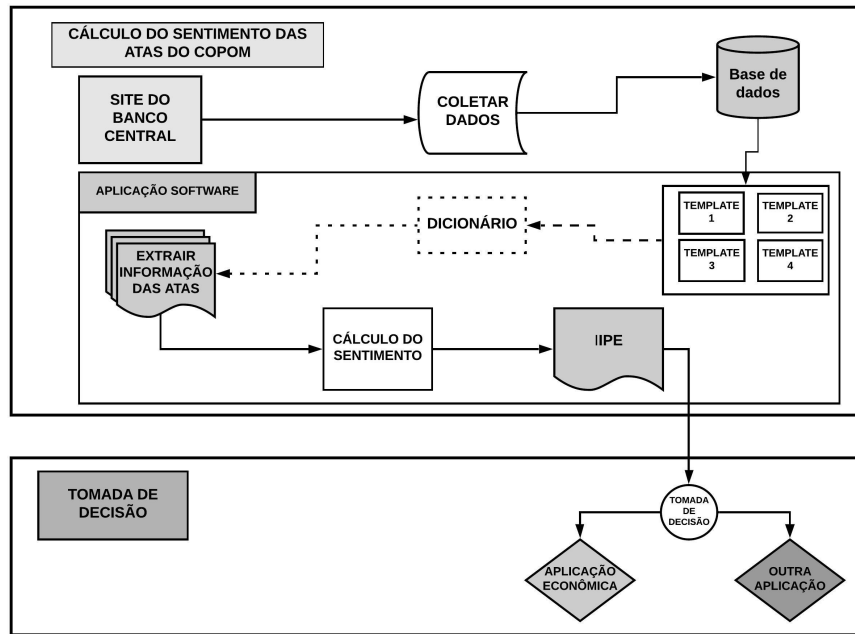
<sup>10</sup> A média de frequência das palavras de incerteza é obtida mediante a equação:  $a_j = \frac{1}{n} \sum_{i=1}^n FPI_j$ , em que  $n$  representa o número de termos de incerteza propostos por este estudo, que estão presentes no texto;  $FPI$  é a frequência de palavras de incerteza que aparecem no documento  $j$ .

<sup>11</sup> Para maiores esclarecimentos consultar Chisholm e Kolda (1999) os autores apresentam detalhadamente técnicas ponderação para o *vector space model*.

<sup>12</sup> Transformação de informação qualitativa em quantitativa.



Figura 1 – Processo de coleta e estimação do sentimento textual



Fonte: Elaboração própria

marca de parágrafo, caracteres maiúsculos), e é aplicada *a priori* sobre a base de dados de modo a prepará-la a fim de que o algoritmo possa ser executado sem problemas. Uma das maiores dificuldades encontradas no processo de coleta dos dados foi a falta de padronização entre as atas, documentos apresentaram diferença de escrita, alteração na aparência textual que se torna um problema sério devido ao fato do *software* diferenciar letras maiúsculas e minúsculas, organização dos tópicos e exposição da decisão de política monetária. Sendo possível, identificar ao longo de toda a base de dados vários formatos distintos para os documentos que serão analisados.

O uso da técnica de *web scraping*<sup>13</sup> é impossibilitado. Como pode ser visto na Figura (2), existe uma quebra textual que impede o algoritmo obter uma enumeração eficiente das palavras de incerteza presente ao longo dos textos. Ao estar trabalhando com leitura e contagem automatizada de palavras é preciso que o sistema consiga identificá-las e agrupá-las sem maiores problemas, devido a isso, foi preciso tratar toda amostra de modo a criar um padrão entre os documentos.

Terminada esta fase poucas atas apresentaram problemas ao serem convertidas para arquivos *.txt* (títulos de seções com letras trocadas e o surgimento de símbolos). O algoritmo conseguiu contorná-los pelo fato desses problemas estarem nos títulos e não no corpo textual, consequentemente não interferiu na contagem de palavras e mesmo que acontecesse alguma interferência o impacto seria irrisório sobre o resultado, Pao (1978) argumenta que os termos com maior valor semântico se encontram no meio da amostra e não nas extremidades. Todavia, as atas de número 44 e 49 precisaram ser digitadas manualmente devido ao fato de o texto ter ficado codificado<sup>14</sup>, não sendo possível obter nenhuma contagem de frequência.

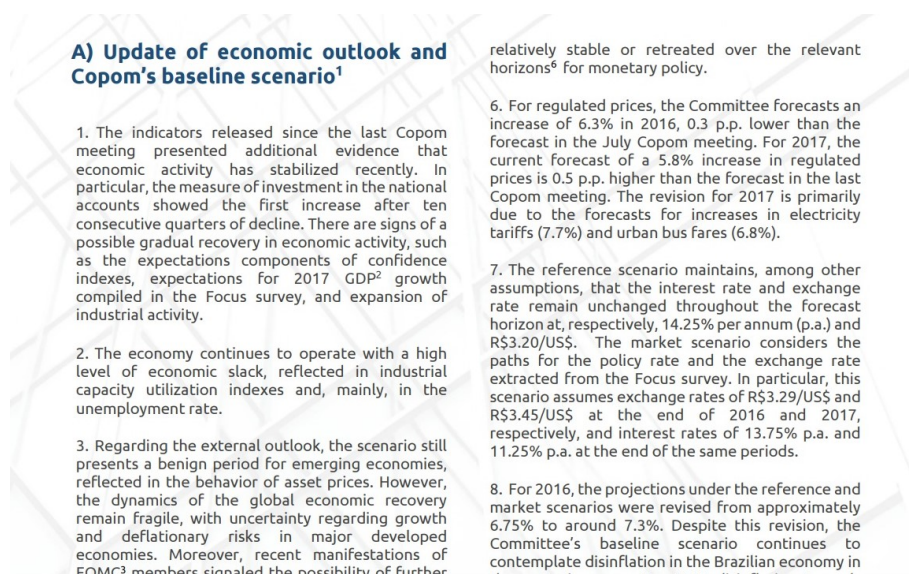
O dicionário<sup>15</sup> nesse caso não é um termo rígido do algoritmo, podendo ter a sua lista de palavras adaptada ou até mesma substituída conforme necessidade do pesquisador. Finalmente, produz-se como resposta valores numéricos por meio de variáveis definidas e obtidas a partir da extração de informações qualitativas contidas no corpo textual das atas. Decorrida esta etapa inicia-se o processo de cálculo do sentimento de incerteza de todas as palavras contidas em cada registro coletado, esse procedimento se

<sup>13</sup> Consiste em importar todos os arquivos diretamente do software, dispensando o *download*.

<sup>14</sup> Surgimento de símbolos ao invés do texto esperado o que impedia o *software* de conseguir lê-las.

<sup>15</sup> Neste trabalho, dicionário e lista de palavras são tratados como sinônimos

Figura 2 – Ata Nº 201: Reunião do Copom - 06 Setembro de 2016



Fonte: Elaboração própria

repete em *loop*<sup>16</sup> até que o último registro tenha o seu sentimento estimado, ao final do procedimento é criado um arquivo único com as informações necessárias. O IIPE é então gerado a partir do agrupamento dos sentimentos individuais de cada ata, ao fim de todo o processo é fornecido um arquivo estruturado no formato *.CSV* com os dados do índice.

## 3.2 DICIONÁRIO

### 3.2.1 Procedimento de Organização da Base de Dados e Criação da Lista de Palavras de Incerteza

Dentro das escolhas realizadas em *text regression*, o dicionário é necessário para extrair o sentimento de incerteza dos documentos, optou-se por trabalhar com as atas em inglês devido ao fato de que o dicionário mais utilizado nas investigações do sentimento encontra-se nesse idioma<sup>17</sup>, ao fazer aplicação desse instrumento permitiu-se utilizar a técnica de *clustering*<sup>18</sup> que consiste em agrupar um conjunto de objeto, nesse caso, palavras. Atendendo melhor a proposta deste estudo.

O dicionário Harvard-IV<sup>19</sup> (*Harvard Psychosocial Dictionary*) empregado anteriormente como referência para esse tipo de análise classificava erroneamente o tom de alguns termos financeiros. Devido a este problema Loughran e McDonald (2011) desenvolveram um dicionário específico capaz de classificar melhor o tom das palavras contidas em textos financeiros, os autores foram pioneiros ao aplicar uma nova metodologia de análise textual.

A hipótese proposta por esta pesquisa demandou a criação de uma lista de palavras próprias que englobasse o contexto de incerteza da política econômica no Brasil, a partir desta lista e concomitantemente com a aplicação de um algoritmo sobre as atas do Comitê de Política de Política Monetária pode-se construir o Índice de Incerteza da Política Econômica para o Brasil.

Desenvolveu-se especificamente para o Brasil um dicionário derivado de uma fusão entre termos retirados da lista de palavras de incerteza definidas pelo Loughran e McDonald (2011) combinada com algumas sentenças escolhidas pelo Baker, Bloom e Davis (2016) que as subdividiu em três categorias em seu estudo para criação do seu Índice de Incerteza da Política Econômica, além de acrescentar alguns termos próprios relevantes para o contexto brasileiro.

<sup>16</sup> O algoritmo ocorre de forma única para cada documento, criando um laço até que todos passem por ele.

<sup>17</sup> Devido à falta de tempo hábil não foi possível desenvolver um dicionário em português.

<sup>18</sup> Técnica que permite agrupar e contar palavras.

<sup>19</sup> Dicionário de Psicologia.



A escolha desses termos se deu mediante a leitura de diversas atas, percebeu-se a existência de um grupo de palavras significativos, quer seja por sua presença sazonal ligada a acontecimentos políticos expressivos para o país, ou por simplesmente atuarem como bigramas fazendo referência a incerteza. O cenário brasileiro se fez necessário a inclusão dessas palavras “exclusivas”, devido a toda conjuntura peculiar que tem pairado sobre o país nos últimos anos, a adição desses termos é unicamente com o propósito de garantir uma melhor discussão acerca do papel que a incerteza desempenha sobre o comportamento das variáveis macroeconômicas. As palavras próprias que foram incorporadas na lista de incerteza para o Brasil foram: *readjustment*, *readjustments*, *reform*, *reforms*, *shock*, *shocks*.

A estruturação da lista de palavras por este trabalho é derivada dos dicionários citados anteriormente além de seguir a divisão de categorias adotada por Baker, Bloom e Davis (2016). A decisão de filtrar as sentenças adotadas neste estudo se deu após a leitura das respectivas listas de palavras, das atas do Copom e aplicação de alguns testes que serão discutidos na próxima sessão. Chegou-se à conclusão que muitas dessas sentenças eram irrelevantes e por isso optou-se por excluí-las. O dicionário escolhido para a construção do Índice Incerteza da Política Econômica é composto por 45 termos subdivididos em três categorias: incerteza que corresponde a 59% do total de palavras, política que representa 21% e econômica com 20%.

Dos termos presentes no IPE 68,89% corresponde a sentenças retiradas do dicionário do Loughran e McDonald (2011) que foram totalmente alocadas na categoria incerteza devido a sua lista ser totalmente composta por palavras de incerteza, 13,33%<sup>20</sup> e 4,44% correspondem as palavras utilizadas pelo Baker, Bloom e Davis (2016) para criar os índices de incerteza da política econômica para o Brasil e Estados Unidos respectivamente que foram destinadas as categorias política e econômica. Por fim; 13,33% das palavras são termos próprios, após uma análise da conjuntura econômica vivida pelo país, da frequência delas ao longo da série temporal e o teor textual de algumas atas que remetiam a quem lê-se o sentido de incerteza, esses termos foram encaixados nas categorias política e econômica.

### 3.2.2 Procedimento de Validação da Lista de Palavras

O processo de validação dos termos próprios aconteceu em três etapas: a) criou-se um mapa de calor apenas com essas sentenças a fim de acompanhar sua evolução ano a ano, b) a partir dos dados obtidos em “a” utilizou-se um algoritmo que capta a frequência de buscar realizadas no Google no país durante períodos específicos, e c) consistiu em gerar o IPE utilizando as listas de palavras de incerteza utilizadas pelo Loughran e McDonald (2011) e Baker, Bloom e Davis (2016). Os resultados obtidos ao longo do estágio de validação permitiram obter informação que comprovam que a inclusão desses termos não foi uma ação aleatória, mas sim um processo pensado e previamente definido com objetivo de aumentar o poder do indicador em captar com precisão o sentimento de incerteza.

A Figura 3 mostra o mapa (*heatmap*) de calor dessas palavras nas atas das reuniões ao longo do período observado, optou-se por usar essa técnica devido a sua clareza. Wilkinson e Friendly (2009), argumentam que ao utilizar essa ferramenta é possível compactar matrizes com uma grande quantidade de dados<sup>21</sup> podem ser exibidas de forma compacta facilitando sua inspeção.

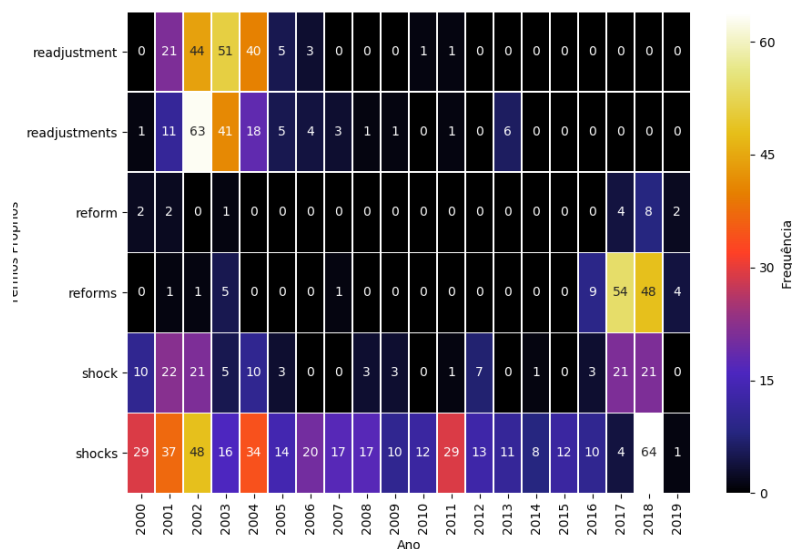
Ao analisar o *heatmap* percebemos a presença de palavras sazonais, nos primeiros anos da amostra os termos *readjustment* e *readjustments* apareceram de forma mais incisiva nos textos das atas, muito devido a transição de governo e a forma diferente que cada um conduzia a política econômica. Posteriormente, ao observar a parte final, notou-se que as sentenças *reform* e *reforms* apresentam-se de modo significativo, após o *impeachment*. O atual contexto do Brasil favoreceu a sua inclusão ao considerá-las vocábulos que se aplicam ao grupo política do IPE, tendo em vista que o Governo defende a realização de algumas reformas a respeito de questões<sup>22</sup> que trazem impactos significativos para a economia, a indefinição a respeito de sua ocorrência, por sua vez, sinaliza uma mensagem de incerteza para os especialistas. Por fim, as palavras *shock* e *shocks* presentes ao longo de toda série temporal devem

<sup>20</sup> A respeito de palavras repetidas considerou apenas a primeira contagem na seguinte ordem de aparecimento: Loughran e McDonald, Barker lista Brasil e Barker lista USA, evitando contar duas vezes a mesma palavra.

<sup>21</sup> Número muito grande de colunas e linhas.

<sup>22</sup> Reforma da previdência, reforma trabalhista etc.

Figura 3 – Mapa de Calor Frequência dos Termos Próprios



Fonte: Elaboração própria

ser consideradas como bigramas<sup>23</sup>, sozinhas elas possuem um efeito neutro, mas quando precedem outro termo como por exemplo, “choque de juros”, traz alusão de incerteza. É preciso levar em conta o fato de que para estudar o ambiente brasileiro é necessário utilizar alguns vocábulos próprios que permitem examinar de forma eficiente o esse sentimento.

A segunda etapa de validação consistiu em aplicar as informações obtidas através do mapa de calor, mediante uso de um algoritmo foi possível verificar a relevância de alguns termos relevantes para o nosso estudo pesquisados no Google. No anexo 2 é apresentado alguns gráficos que reflete o interesse ao longo do tempo no Brasil de alguns bigramas derivados a partir de uma combinação entre as informações das atas e do 3. Mesmo sendo considerado o fato de que os leitores podem ter visto essa sentença de outras fontes e não apenas das atas é possível notar claramente que o comportamento dos dados reflete o que foi apresentado no mapa de calor<sup>24</sup>.

Como pôde ser visto no fluxograma o dicionário não é um termo rígido podendo ser facilmente substituído para a estimação do índice, com a finalidade de validar a lista de palavras utilizadas por esse estudo, estimou-se novamente o IIPE três vezes com o dicionário [Loughran e McDonald \(2011\)](#), e a lista de palavras utilizadas pelo [Baker, Bloom e Davis \(2016\)](#) para o Brasil e Estados Unidos, após observar os resultados percebeu que o comportamento do índice praticamente não sofre alterações, comprovando a robustez da lista de palavras utilizada por esse estudo<sup>25</sup>.

## 4 MODELO EMPÍRICO

### 4.1 VAR

Modelo de Vetores Autorregressivo em sua forma estrutural pode ser representado da seguinte forma:

<sup>23</sup> Pares ordenados de palavras.

<sup>24</sup> Para maiores esclarecimentos à respeito do processo de justificação e validação dos termos próprios contatar um dos autores.

<sup>25</sup> Os resultados desse teste de robustez não estão sendo representados nesse trabalho, para qualquer esclarecimento a respeito deles contatar um dos autores.

$$A_0 y_t = \sum_{\ell=1}^p A_\ell y_{t-\ell} + \epsilon_t \quad \text{para } t = 1, \dots, T \quad (3)$$

em que:  $y_t$  é um vetor coluna  $n$ -dimensional de variáveis endógenas,  $A_0$  é uma matriz  $n \times n$  de impactos contemporâneos,  $A_\ell$  são  $n \times n$  matrizes de parâmetro de variáveis defasadas,  $\epsilon_t$  é um processo de ruído branco,  $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$ ,  $p$  é o número de atrasos e  $T$  o tamanho da amostra.

Ao multiplicar (3) por  $A_0^{-1}$  obtém-se a forma reduzida usual do VAR:

$$y_t = \sum_{\ell=1}^p B_\ell y_{t-\ell} + u_t \quad \text{for } t = 1, \dots, T \quad (4)$$

onde:  $B_\ell = A_0^{-1} A_\ell$  para  $\ell = 1, 2, \dots, p$ ;  $u_t = A_0^{-1} \epsilon_t$  e  $\Omega = E[u_t u_t']$  é a matriz de variância-covariância dos resíduos.

O sistema obtido em (4) é o que deve ser estimado, a partir dos erros de forma reduzida,  $u_t$  é possível conseguir as funções de respostas aos impulsos ortogonais.

#### 4.1.1 VAR com restrição de sinais

Um dos benefícios de fazer uso do método de restrição de sinais é o fato de que não é necessário impor restrições sobre toda a matriz  $A_0$ , o propósito de aplicar a abordagem da restrição de sinais é o fato de poder identificar os choques estruturais, ao impor sobre algumas variáveis uma restrição ao sinal da resposta impulsiva, ao passo que para as variáveis de interesse é possível manter o sinal irrestrito com o objetivo de observar a sua resposta ao impulso. Os choques estruturais identificados: choque de produtividade, choque na política monetária e choque de incerteza.

Entendendo melhor o funcionamento dessa estratégia de identificação, denote  $a_k$  o vetor de coluna  $k$  da matriz  $A$ , de modo que  $AA' = \Sigma$  e  $A_k$  o vetor de impulso para a  $k$ -ésima variável. O objetivo é identificar apenas  $m$  choques em que  $m \leq n$ . Essa matriz de impulso  $[a^{(1)}, \dots, a^{(m)}]$ , pode ser representada como sendo o produto:

$$[a^{(1)}, \dots, a^{(m)}]_{(m \times n)} = \tilde{A} \Lambda$$

em que:  $\Lambda = [\lambda^{(1)}, \dots, \lambda^{(m)}]$  é uma matriz ortonormal  $m \times n$  de modo que:  $\Lambda \Lambda' = I_m$  e  $\tilde{A}$  é um fator Cholesky triangular inferior de  $\Sigma$ .

Seja  $a = a^{(s)}$ ,  $s = 1, 2, \dots, m$  uma das colunas de impulso da matriz  $\tilde{A} \Lambda$  e  $\lambda = \lambda^{(s)} = \tilde{A}^{-1}$  seja a coluna correspondente de  $\Lambda$ . Uhlig (2005) demonstra que qualquer resposta de impulso  $a$ , pode ser obtida mediante uma combinação linear das respostas de impulso perante uma decomposição de Cholesky  $\Sigma$  que satisfaz  $\Sigma = E[u_t u_t'] = A E[v_t v_t'] A' = A A'$ . Admitindo  $r_i(k) \in R^m$  como sendo a resposta vetorial no horizonte  $k$  para o  $i$ -ésimo choque em uma decomposição de Cholesky de  $\Sigma$ . A função impulso-resposta  $r_a(k)$  para  $a^{(s)}$  é dada por:

$$r_a^{(s)}(k) = \sum_{i=1}^m \lambda_i r_i(k) \quad (5)$$

A equação (5), permite identificar o vetor de resposta ao impulso referente as inovações estruturais, porém, essas restrições não são condições suficientes para identificar corretamente os choques. Considerando a literatura empírica e tendo como base o que foi apresentado em Baker, Bloom e Davis (2016), os efeitos precaucionistas de um choque de incerteza se assemelham a um choque de demanda negativo, por isso, ao determinar as restrições dos sinais adota-se o que foi definido em Redl (2017).

#### 4.1.2 Restrição de Sinal

Ao adotar o modelo com restrições de sinais, o VAR permite identificação com maior grau de flexibilidade em relação a suposição em torno do tempo de respostas das variáveis aos choques, que resulta no conjunto de respostas identificadas para esse choque.

As restrições impostas no exercício empírico são de certa forma intuitivas, um choque de produtividade implica em um aumento do PIB enquanto a inflação se reduz, o Banco Central em resposta a elevação do PIB e a um cenário de baixa inflação reduz a taxa de juros nominal. Por outro lado, optamos por um choque de política monetária, a taxa de juros eleva-se, causando uma redução no PIB e na inflação. Nenhuma outra restrição adicional é imposta. Por fim, o choque de incerteza reduz a taxa de juros básica, o que consequentemente traz um impacto negativo no crédito e devido a relação horas/produtividade então, indiretamente com a redução do crédito a produtividade indústria cai, assim como as horas trabalhadas. Devido ao seu efeito cumulativo e considerando as expectativas formadas, a incerteza aumenta. As variáveis restantes são deixadas irrestritas tendo em vista que estamos interessados em observar o seu comportamento, principalmente após o choque de incerteza, as restrições descritas podem ser observadas na Tabela 2.

Tabela 2 – Restrição de sinal para 3 trimestres

	Selic	IPC	Horas	FBCF	Consumo	PIB	Spread	IIFE
<b>Choque de Produtividade</b>	-	-	?	?	?	+	?	?
<b>Choque de Política Monetária</b>	+	-	?	?	?	-	?	?
<b>Choque de Incerteza</b>	-	?	-	?	?	?	-	+

Fonte: Elaboração própria

## 5 RESULTADOS DA ESTIMAÇÃO TEXTUAL

Os primeiros resultados demonstram que o Índice de Incerteza da Política Econômica desenvolvido nesse trabalho é capaz de capturar o sentimento de incerteza a partir das atas do Copom, os picos apresentados pelo IIFE condizem com a realidade vivida pelo Brasil durante as duas últimas décadas, principalmente no âmbito político.

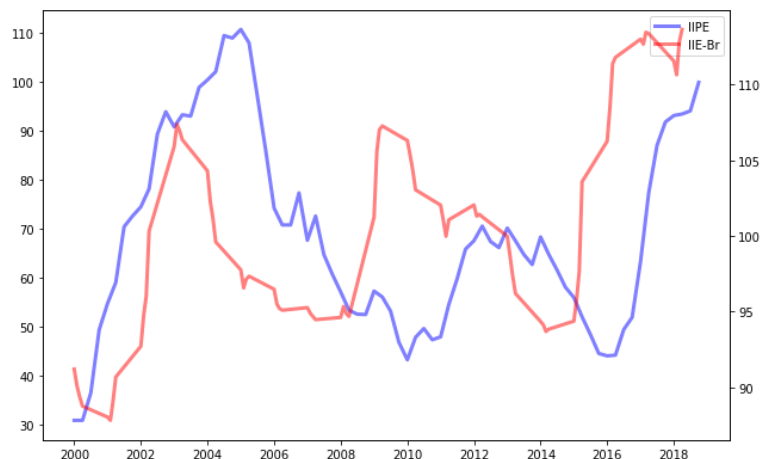
Apesar de que na teoria o IIE-Br desenvolvido por [Ferreira et al. \(2017\)](#) e o IIFE construído neste trabalho são indicadores de medidas diferentes que mensuram o grau de incerteza econômica e a incerteza da política econômica respectivamente, optou-se por compará-los devido ao fato de ser o único indicador construído nacionalmente. Ao contrapor esses dois indicadores observamos trajetórias divergentes e convergentes, no entanto, os períodos de ciclo e de vale apresentam o mesmo comportamento, a diferença mais significativa está no intervalo de duração e em alguns pontos na sensibilidade, conforme pode ser observado na Figura 4.

Enquanto o IIE-Br reduz drasticamente a incerteza após 2003 com a mudança Presidencial e a possível “euforia” dos jornais nesse ano, por sua vez o IIFE considera a incerteza alta até 2005, este resultado influenciado pela elevação dos juros dos Estados Unidos em 2004 e a Primeira Crise Política vivida pelo Partido dos Trabalhadores em 2005, o que justifica uma postura um pouco mais conservadora do Banco Central até 2005. Destaca-se também o fato de o IIE-Br ser mais sensível a Crise Internacional de 2008, assim como o IIFE absorveu melhor a Crise na Europa. Por fim, observa-se ambos os índices adotando o mesmo curso ascendente nos três anos. Sendo assim, os primeiros resultados obtidos com o indicador construído nesse trabalho são satisfatórios, conseguiu apresentar um indicador derivado de uma metodologia mais simples e que pode auxiliar na tomada de decisão.

Em um segundo exercício comparou o IIFE com o *Economic Policy Uncertainty Index Brazil*<sup>26</sup> (EPU) criado por [Baker, Bloom e Davis \(2016\)](#), ambos os índices estão concentrados em captar justamente esse sentimento de incerteza da política econômica mediante análise textual, a diferença se deu mediante

<sup>26</sup> Disponível em: <http://www.policyuncertainty.com/brazilmonthly.html>.

Figura 4 – Comparativo IIPExIIE-Br



Fonte: Elaboração própria

Nota: Índices suavizados exponencialmente

a fonte de coleta e o dicionário utilizado. Enquanto o EPU Brasil utiliza jornal e um dicionário com 17 palavras o IIPE tem como foco as atas das reuniões do Copom e um dicionário com 45 palavras.

Apesar de existir grande diferença entre a quantidade de documentos utilizados por ambos indicadores<sup>27</sup>, as conclusões iniciais se mostraram satisfatórias, as trajetórias apesar de mostrarem alguns pontos de divergências em alguns momentos, os períodos relevantes demonstraram convergência. Como pode ser observado no período de mudança de partido presidencial (2002-2003), a Crise na Europa (2011-2013<sup>28</sup>), e finalmente o período que antecedeu e sucedeu o *impeachment* da Presidente Dilma Rousseff (2016-2018), como pode ser visto na Figura 5.

Figura 5 – Comparativo IIPExEPU



Fonte: Elaboração própria

Nota: Índices suavizados exponencialmente

Como parte do processo de validação do indicador construído neste trabalho, foi realizada uma associação dos índices a um conjunto de variáveis macroeconômicas através do coeficiente de correlação de Pearson com objetivo de obter o grau e a direção dessa correlação. O foco principal é analisar o comportamento dos sinais em resposta a um aumento da incerteza, como pode ser constatado na Tabela

<sup>27</sup> O jornal possui matérias diárias, por sua vez, a ata que até 2005 era mensal, atualmente é divulgada a cada 45 dias.

<sup>28</sup> A Crise na Europa iniciou em maio de 2010, porém, os dois índices começaram a captar o aumento a incerteza doméstica em 2011.



3, apenas o IIPE mostra os resultados esperados levando em consideração a literatura. Um aumento de incerteza resulta no aumento da taxa de juros e do *spread*, assim como uma redução da inflação, horas trabalhadas na indústria, investimento, consumo e do PIB.

Tabela 3 – Coeficiente de correlação

	Selic	IPC	Horas	FBCF	Consumo	PIB	<i>Spread</i>
<b>IIPE</b>	0.18	-0.03	-0.40	-0.16	-0.03	-0.05	0.17
<b>IIE-Br</b>	-0.05	0.21	-0.38	0.28	0.43	0.41	-0.05
<b>EPU</b>	-0.18	-0.002	-0.44	0.46	0.63	0.60	-0.18

Fonte: Elaboração própria

A partir dos dados contidos na Tabela 3 e apoiado na literatura de Filho (2014), Baker, Bloom e Davis (2016), Barboza e Zilberman (2018), identificou-se que a influência da incerteza sobre a atividade econômica se dá por meio de canais de transmissões: *i*) consumo, *ii*) investimento e *iii*) risco.

Ao observar os efeitos da incerteza sobre o canal consumo nota-se que ela irá afetá-lo de forma negativa, as famílias reduzirão consumo ao mesmo tempo que deslocam recursos para a poupança precaucionária; o investimento, por sua vez, é afetado de forma mais significativa, as firmas e os agentes econômicos postergam decisões à medida que o cenário de indefinição se eleva, o que ocasiona na redução da produção; o risco está ligado a possibilidade de um *default* financeiro, ou seja, esse canal sinaliza mensagens para o mercado sobre um possível colapso financeiro, espera-se a elevação das taxas de juros como medida de “fidelização” dos agentes econômicos. Por fim, o impacto acumulado sobre a atividade econômica é negativo.

### 5.1 Choque de incerteza e seu impacto macroeconômico

Nesta subseção apresenta-se os resultados do modelo empírico, aplicou-se a restrição de sinal sobre os três primeiros trimestres nas variáveis Selic, Horas, *Spread* e IIPE. Com objetivo de observar o impacto que um choque de incerteza causaria sobre as demais variáveis do modelo. A Figura 6 exibe as funções impulso-resposta para a restrição de sinal temporária  $k = 3$  assim, as variáveis seguem a restrição imposta até um período de 3 trimestres após o choque em que: a incerteza não pode apresentar uma resposta negativa, a taxa de juros e o crédito apresentam retorno negativo e por fim, a variável hora reduz-se durante o choque de incerteza.

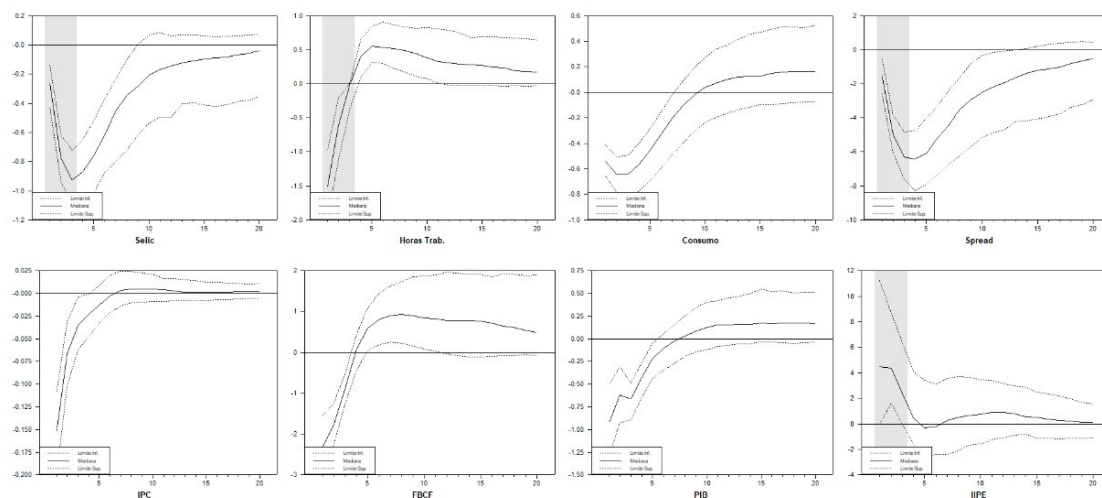
A taxa Selic continua apresentar um efeito negativo significativo por seis trimestres após a retirada da restrição imposta, com relação ao *spread* do crédito, o impacto negativo do choque manteve-se significativo até o nono trimestre. A variável horas teve o efeito imposto perdurando até o primeiro trimestre, posteriormente apresentou um efeito positivo significativo.

Como pode ser visto na Figura 6, as funções de impulso-resposta das variáveis irrestritas apresentaram levando em consideração a literatura resultados esperados. Há impactos negativos e estatisticamente significativos, a inflação demonstrou uma queda inicial de 0.15 no primeiro trimestre e após o quinto a trajetória foi de encontro até que o efeito seja nulo. Com relação ao consumo, a redução ocorre de forma mais negativa nos dois primeiros trimestres, em seguida estabiliza-se durante um período e logo em seguida segue uma trajetória de redução dos valores negativos. O investimento evidenciou uma queda brusca no primeiro trimestre, porém, logo em seguida apresentou um curso de redução desse efeito negativo, até que a partir do quarto trimestre os valores passam a ser positivos sendo considerados significativos até pelo menos doze trimestres. O PIB em um primeiro momento apresenta uma redução que seguida de um leve trajeto de recuperação por dois trimestres e novamente exibe um declínio por mais um trimestre, até que essa trajetória siga o caminho de redução dos efeitos negativos que se estende até o sétimo trimestre.

Os resultados sugerem então que os setores analisados são afetados de forma significativa pela incerteza da política econômica, trazendo esse resultado à luz do desempenho apresentado pelo Brasil ao longo dos últimos quatro anos percebe a possibilidade de estabelecer uma conexão entre eles e o

desempenho das variáveis macroeconômicas. Se considerarmos o fato de a incerteza apresentar um comportamento duradouro ao invés de um choque, pode-se deduzir que os efeitos sobre a atividade econômica no Brasil são consideráveis.

Figura 6 – Choque de Incerteza

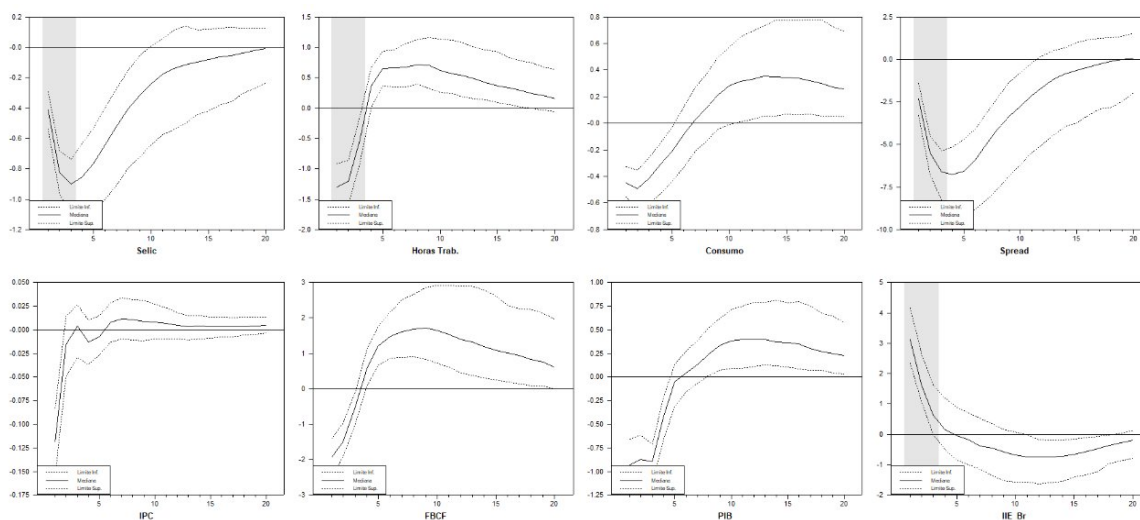


Fonte: Elaboração própria

### 5.1.1 Robustez

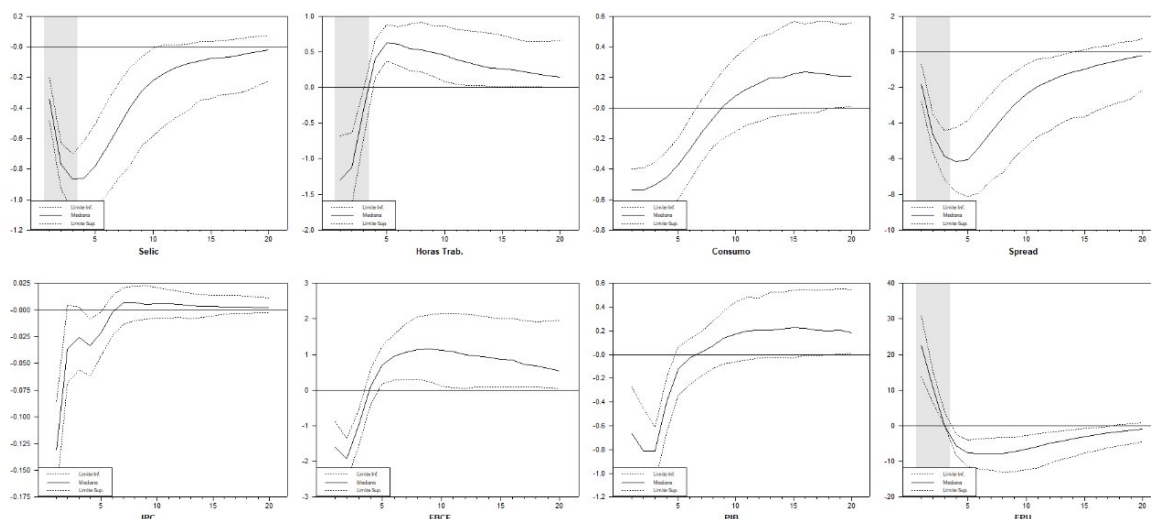
Tendo como objetivo testar a robustez do índice, substituímos o IIPE pelo IIE-Br e o EPU Brasil e repetimos o exercício anterior, o que permitiu analisar o impacto de um choque de incerteza advindo de outros indicadores sobre o mesmo conjunto de dados. Conforme pode ser observado nas Figuras 7 e 8, percebe-se que os resultados apresentados vão de encontro ao que foram apresentados e discutidos em 5.1, sendo assim, considera-se robusto o Índice de Incerteza da Política Econômica construído neste trabalho.

Figura 7 – Choque de Incerteza: IIE-Br



Fonte: Elaboração própria

Figura 8 – Choque de Incerteza: EPU Brasil



Fonte: Elaboração própria

## 6 CONSIDERAÇÕES FINAIS

Como observado, os primeiros resultados demonstraram que o Índice de Incerteza da Política Econômica desenvolvido nesse estudo é capaz de capturar o sentimento de incerteza a partir das atas do Copom. Ao comparar com os demais indicadores, a conclusão em um primeiro momento mostrou ser satisfatória, claro que é possível buscar formas de refinar ainda mais o índice construído neste trabalho com o objetivo de melhorar a sua precisão em captar esse sentimento.

Um ponto que merece destaque é a eliminação de um possível viés do intérprete<sup>29</sup>, os indicadores estimados por meio de artigos de jornais muitas vezes pode refletir uma análise própria do jornalista ao interpretar as informações divulgadas nas atas das reuniões. Espera-se a existência de um viés de escrita por parte de quem divulga a ata, até como medida de reduzir o impacto sobre as expectativas do mercado. Por isso, ao optar por gerar o índice diretamente dos documentos oficiais, retira-se esse provável segundo viés.

Ao observar as variáveis em que não foram aplicada nenhuma restrição de sinal, um choque de incerteza afeta de forma significativa os setores analisados, é produzido um impacto negativo sobre a inflação, o investimento (em um primeiro momento), consumo e o PIB. Concluiu-se que a incerteza possui um comportamento contracíclico e choques de incerteza terão um efeito negativo sobre a atividade econômica. Os resultados alcançados neste trabalho estão de acordo com o que foi apresentado por Filho (2014) e Barboza e Zilberman (2018), no entanto, fez-se o uso de um conjunto maior de variáveis macroeconômicas e de apenas uma variável de incerteza. É possível estabelecer um elo entre os resultados advindos deste estudo e o desempenho macroeconômico do Brasil nos últimos 4 anos.

## 7 AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

<sup>29</sup> O Ferreira et al. (2017) justifica que ao utilizar apenas um jornal na sua análise o resultado poderia conter viés.

## REFERÊNCIAS

- AARLE, B. van; MOONS, C. Sentiment and uncertainty fluctuations and their effects on the euro area business cycle. *Journal of Business Cycle Research*, Springer, v. 13, n. 2, p. 225–251, 2017. Citado na página 3.
- AGUIAR, E. J. de et al. Análise de sentimento em redes sociais para a língua portuguesa utilizando algoritmos de classificação. In: SBC. *Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. [S.l.], 2018. Citado na página 2.
- ALCHIAN, A. A. Uncertainty, evolution, and economic theory. *Journal of political economy*, The University of Chicago Press, v. 58, n. 3, p. 211–221, 1950. Citado na página 2.
- ALUÍSIO, S. M.; ALMEIDA, G. M. de B. O que é e como se constrói um corpus? lições aprendidas na compilação de vários corpora para pesquisa linguística. *Calidoscópico*, v. 4, n. 3, p. 156–178, 2006. Citado na página 5.
- APEL, M.; GRIMALDI, M. The information content of central bank minutes. 2012. Citado na página 4.
- BAKER, S. R.; BLOOM, N.; DAVIS, S. J. Measuring economic policy uncertainty. *The quarterly journal of economics*, Oxford University Press, v. 131, n. 4, p. 1593–1636, 2016. Citado 7 vezes nas páginas 3, 8, 9, 10, 11, 12 e 14.
- BARBOZA, R. d. M.; ZILBERMAN, E. Os efeitos da incerteza sobre a atividade econômica no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 72, n. 2, p. 144–160, 2018. Citado 4 vezes nas páginas 2, 3, 14 e 16.
- BHOLAT, D. et al. Text mining for central banks. *Available at SSRN 2624811*, 2015. Citado na página 4.
- BRUNO, G. Text mining and sentiment extraction in central bank documents. In: IEEE. *2016 IEEE International Conference on Big Data (Big Data)*. [S.l.], 2016. p. 1700–1708. Citado na página 4.
- BRUNO, G. Central bank communications: Information extraction and semantic analysis. In: *The R User Conference, useR! 2017 July 4-7 2017 Brussels, Belgium*. [S.l.: s.n.], 2017. p. 253. Citado na página 4.
- CHISHOLM, E.; KOLDA, T. G. *New term weighting formulas for the vector space method in information retrieval*. [S.l.], 1999. Citado 2 vezes nas páginas 5 e 6.
- CORREA, R. et al. Sentiment in central banks' financial stability reports. *Available at SSRN 3091943*, 2017. Citado na página 3.
- FERREIRA, P. C. et al. Medindo a incerteza econômica no brasil. *Economia Aplicada*, 2017. Citado 3 vezes nas páginas 3, 12 e 16.
- FILHO, A. E. C.; ROCHA, F. Como o mercado de juros futuros reage à comunicação do banco central? *Economia aplicada*, SciELO Brasil, v. 14, n. 3, p. 265–292, 2010. Citado na página 4.
- FILHO, A. E. d. C. Incerteza e atividade econômica no brasil. *Economia Aplicada*, SciELO Brasil, v. 18, n. 3, p. 421–453, 2014. Citado 4 vezes nas páginas 2, 3, 14 e 16.
- GODEIRO, L. L.; LIMA, L. R. R. de O. Medindo incerteza macroeconômica para o brasil. *Economia Aplicada*, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto-USP . . . , v. 21, n. 2, p. 311, 2017. Citado na página 3.

- GRAMINHO, F. M. et al. Sentimento e macroeconomia: uma análise dos índices de confiança no brasil. *Brazilian Central Bank, trabalhos para discussão*, v. 408, 2015. Citado na página 2.
- GRIGNANI, F.; FONTANA, R. *Using machine learning and Bayesian networks to objectively analyze central bank statements and market sentiment*. Tese (Doutorado) — Politecnico di Torino, 2018. Citado na página 4.
- HANSEN, S.; MCMAHON, M. Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, Elsevier, v. 99, p. S114–S133, 2016. Citado na página 4.
- HENDRY, S.; MADELEY, A. Text mining and the information content of bank of canada communications. *Available at SSRN 1722829*, 2010. Citado na página 3.
- HUTTO, C. J.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth international AAAI conference on weblogs and social media*. [S.l.: s.n.], 2014. Citado na página 2.
- JURADO, K.; LUDVIGSON, S. C.; NG, S. Measuring uncertainty. *American Economic Review*, v. 105, n. 3, p. 1177–1216, 2015. Citado 2 vezes nas páginas 2 e 3.
- KEARNEY, C.; LIU, S. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, Elsevier, v. 33, p. 171–185, 2014. Citado na página 3.
- KNIGHT, F. H. *Risk, uncertainty and profit*. [S.l.]: Houghton Mifflin Company, New York, 1921. Citado na página 2.
- LI, J. et al. Topic popularity prediction with sentiment time series on short text based social media. 2019. Citado na página 2.
- LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, Wiley Online Library, v. 66, n. 1, p. 35–65, 2011. Citado 3 vezes nas páginas 8, 9 e 10.
- MACHADO, M. A. V.; SILVA, M. D. d. O. P. d. et al. Análise do sentimento textual dos relatórios de desempenho trimestral das indústrias brasileiras. *Sociedade, Contabilidade e Gestão*, v. 12, n. 1, 2017. Citado 2 vezes nas páginas 2 e 3.
- MONTES, G. C.; NICOLAY, R. T. d. F. Central bank’s perception on inflation and inflation expectations of experts: empirical evidence from brazil. *Journal of Economic Studies*, Emerald Group Publishing Limited, v. 42, n. 6, p. 1142–1158, 2015. Citado na página 4.
- NOPP, C.; HANBURY, A. Detecting risks in the banking system by sentiment analysis. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, 2015. p. 591–600. Disponível em: <<https://aclweb.org/anthology/D/D15/D15-1071>>. Citado na página 2.
- ORMEROD, P.; NYMAN, R.; TUCKETT, D. Measuring financial sentiment to predict financial instability: A new approach based on text analysis. *arXiv preprint arXiv:1508.05357*, 2015. Citado na página 2.
- PAO, M. L. Automatic text analysis based on transition phenomena of word occurrences. *Journal of the American Society for Information Science*, Wiley Online Library, v. 29, n. 3, p. 121–124, 1978. Citado na página 7.
- REDL, C. The impact of uncertainty shocks in the united kingdom. Bank of England Working Paper, 2017. Citado 2 vezes nas páginas 5 e 11.
- RYBINSKI, K. I. A machine learning framework for automated analysis of formal and informal central bank communication. the case of the national bank of poland. 2018. Citado na página 4.

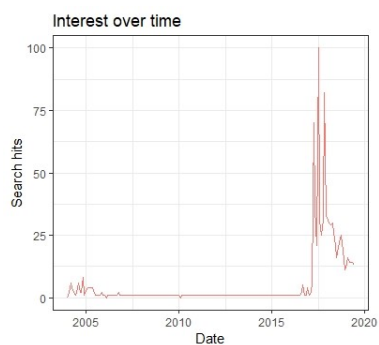


- SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information processing & management*, Elsevier, v. 24, n. 5, p. 513–523, 1988. Citado na página 5.
- SCHYMURA, L. G. A crescente importância de medir a incerteza e seus impactos no brasil de hoje. *Revista Conjuntura Econômica*, v. 71, n. 5, p. 6–9, 2017. Citado na página 2.
- SILVA, M. D. d. O. P. et al. O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro. Citado na página 2.
- UHLIG, H. What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, Elsevier, v. 52, n. 2, p. 381–419, 2005. Citado 2 vezes nas páginas 4 e 11.
- WILKINSON, L.; FRIENDLY, M. The history of the cluster heat map. *The American Statistician*, Taylor & Francis, v. 63, n. 2, p. 179–184, 2009. Citado na página 9.
- YOU, S.; DESARMO, J.; JOO, S. Measuring happiness of us cities by mining user-generated text in flickr. com: A pilot analysis. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*. [S.l.], 2013. p. 167. Citado na página 2.

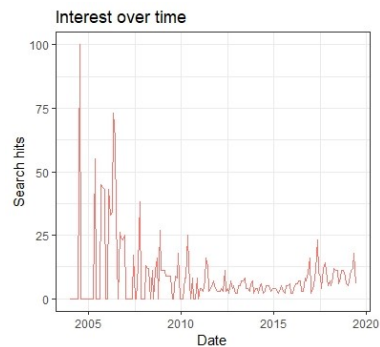
## 1 LISTA DE PALAVRAS INCERTEZA

congress; deficit; deficits; economic; economy; instabilities; instability; readjustment; readjustments; reform; reforms; regulation; regulations; risk; risked; riskier; riskiest; riskiness; risking; risks; risky; rumors; senate; shock; shocks; speculate; speculated; speculates; speculating; speculation; speculations; speculative; speculatively; uncertain; uncertainly; uncertainties; uncertainty; unexpected; unexpectedly; unpredictable; unpredictably; unpredicted; volatile; volatilities; volatility.

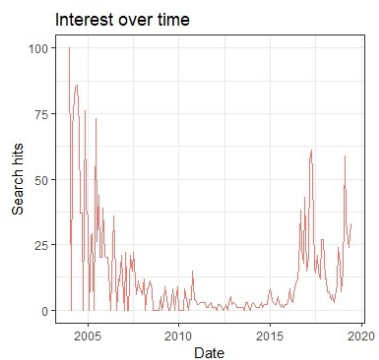
## 2 PESQUISAS GOOGLE TRENDS BRASIL



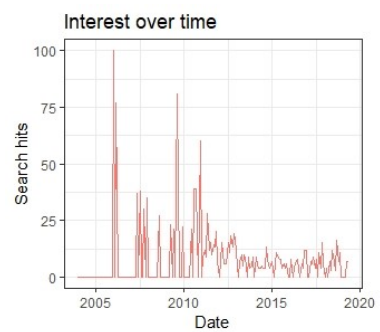
(a) Reforma Trabalhista



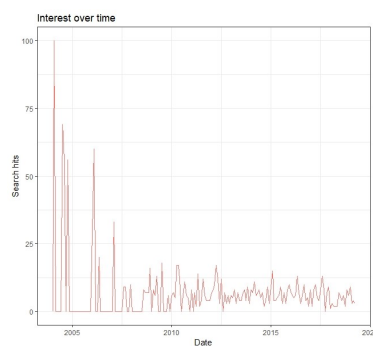
(b) Reforma Econômica



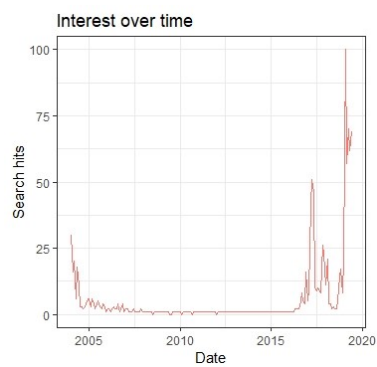
(c) Reforma Previdenciária



(d) Choque de Oferta



(e) Reajuste de Juros



(f) Reforma da Previdência

Figura 9 – Termos Próprios Google Trends Brasil

**Fonte: Elaboração Própria**

Nota: Os Termos são tratados como bigramas.