European Central Bank Speeches: A Case Study

By

Gustavo de Oliveira Vital

Master Thesis in Economics

Supervised by:

João Manuel Portela da Gama

Faculdade de Economia

Universidade do Porto

2022

Abstract

This work intends to relate the press conferences of the European Central Bank, for a period from January 2005 to December 2020, with the macroeconomic scenario and with real European macroeconomic variables. As exposed by (Shapiro, Sudhof, & Wilson, 2020; Shapiro & Wilson, 2021) and (Barsky & Sims, 2012) it is possible to better understand what happens in the real economy from sentiment indices. This work follows Shapiro et al. (2020) when the author exposes the possibility of an economic index obtained by Natural Language Processing techniques through minutes and central bank reports, allowing a correlation with different macroeconomic scenarios. The methodology used includes processing techniques of Natural Language Processing and Sentiment Analysis. To relate the indices with macroeconomic variables, variable selection models are used (LASSO, Adaptive Lasso and Elastic Net). Furthermore, impulse response functions are obtained for a shock to the index and it is analyzed how the macroeconomic variables respond to a shock in the economic sentiment index.

Keywords: Sentiment Analysis, Central Bank, Vector Autoregressive, Variable Selection Models.

Resumo

Este trabalho pretende relacionar as conferências de imprensa do Banco Central Europeu, para um período de janeiro de 2005 a dezembro de 2020, com o cenário macroeconómico e com variáveis macroeconómicas reais europeias. Conforme exposto por (Shapiro et al., 2020; Shapiro & Wilson, 2021) e (Barsky & Sims, 2012) é possível entender melhor o que acontece na economia real a partir de indices de sentimentos. Este trabalho segue Shapiro et al. (2020) quando o autor expoe a possibilidade de um indice economico obtido por tecnicas de Natural Language Processing atraves de atas e relatórios de bancos centrais, permitindo uma correlação com diferentes cenários macroeconômicos. A metodologia utilizada engloba técnicas de processamento de Natural Language Processing e Sentiment Analysis. Para relacionar os indices com variaveis macroeconomicas, se faz o uso de modelos de seleção de variaveis (LASSO, Adaptive Lasso e Elastic Net). Ainda, são obtidas as funções de resposta ao impulso para um choque no indice e analisa-se como as variaveis macroeconomicas respondem a um choque no indice de sentimentos economico.

Palavras-chave: Análise de Sentimentos, Banco Central, Vetor Autoregressivo, Modelos de Seleção de Variáveis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Economic decisions, in particular monetary policies, are based on resolutions provided by central banks. These remedies, however, are not always obvious, especially for lay people. For example, central bank minutes often include much more information than appears at first glance, especially when a group of texts is taken into account for a conjunctive or structural analysis. Understanding a collection of texts can occasionally be very difficult for a human being; depending on the purpose of understanding, the analysis can involve thousands or even millions of pages. However, thanks to developments in computing, questions like these are increasingly likely to be resolved. For a better understanding and knowledge of what a text, or a series of texts (corpus), actually says or expresses, techniques and tools like Natural Language Processing (NLP) and text mining are being heavily used.

It is already possible to discover trends in search engines such as Google and Yahoo using text analysis to anticipate macroeconomic indicators or even increase the understanding of consumer behavior. In recent study, Bholat, Hansen, Santos, and Schonhardt-Bailey (2015) considers that, although the advances made in the field of computing are significant, the applicability of text mining in the economy is still not used as it could be. In the same study, the authors present cases of applicability of text mining in the economic field, with examples in the job market or even explaining how Natural Language Processing techniques can be applied to economics.

An alternative approach of application still in the field of Natural Language Processing is the sentiment analysis. Basically, it is "the task of identifying positive and negative opinions, emotions, and evaluations" (Wilson, Wiebe, & Hoffmann, 2005). Sentiment analysis allows extracting information that normally a human being would not be able to, due to the amount of text, or even the difficulty of recognizing unstructured patterns. Through sentiment lexicons, it is possible to classify whether a text, phrase or words has a positive, neutral or negative expression (polarity-based lexicon) – or even assign scores to a text, phrase or word (valence-based lexicon). When working with a set of economic texts, it is possible, then, to assign scores so that feelings can be monitored in a temporal way.

Another article written by Nyman, Kapadia, and Tuckett (2021) considered the possibility of a sentiment index, based on social media and networks, for a measure of excitement or anxiety about the financial and economic situation. The index proposes to act as a proxy for market sentiment: bullish or bearish. The ratio of the index would then be compared against historical events and other financial indicators (Nyman et al., 2021). The idea of sentiment indices does not come from Nyman et al. (2021) – authors have already presented the possibility of modeling the economy by including non-observable variables that can represent economic sentiment. Even though in studies such as Akerlof and Shiller (2010); Angeletos and La'O (2013); Barsky and Sims (2012); Bholat et al. (2015); Gennaioli and Shleifer (2018) the idea of economic sentiments has already been presented; Previous studies Shapiro et al. (2020); Shapiro and Wilson (2021) presented the idea of an index of sentiments coming from Natural Language Processing techniques, where the index presents a boost to economic activity.

Gradually, Natural Language Processing techniques and tools have been increasingly used with implementations in the economic field. It is not difficult to find correlations between macroeconomic variables and sentiment analyzes extracted from the media, economic newspapers or magazines (Nyman et al., 2021; Ostapenko et al., 2020) or even from speeches or parliamentary hearings of central banks (Fraccaroli & Giovannini, 2020; Shapiro et al., 2020; Shapiro & Wilson, 2021).

The examination and investigation of how textual patterns supplied by the European Central Bank can point to and connect with conjunctural and structural moments of the economy is the major goal of this work. From the speeches of the European Central Bank, an index of economic sentiment was elaborated that can reflect what the economic scenario goes through – that is, the reflection of the economy from the point of view of the economic speeches of the European Central Bank. The use of Natural Language Processing techniques in the economics field can be very advanced and useful for a better understanding of what central banks are actually indicating (through textual documents), and not just reduce the central bank's indication based on its own resolutions. Language Processing techniques are being used more and more to capture patterns that are practically imperceptible to the human eye.

The greatest relevance and contribution of this work is the use of the speeches of the European Central Bank as an indicator of economic sentiments. From this point, it will be possible to compare the results obtained with previous correlated works in the literature so that an "asymptotic" relationship is expected in terms of results – in terms of counter intuition, the objective of this work is to corroborate the indicated literature.

The chapter 2 presents the bibliographic review of this work and exposes the main works used as a reference for this bibliography. The chapter exposes the point of view of Shapiro et al. (2020); Shapiro and Wilson (2021) articles that made use of Natural Language Processing and Sentiment Analysis techniques. Also, other articles Angeletos and La'O (2013); Barsky and Sims (2012) are exposed that work with the *concept* of feelings in an economic scenario, even when considering a general equilibrium scenario.

The 3 chapter presents the review and introduction of the reference methodology, related to the Natural Language Processing parts and lexicons. Two fundamental lexicons are exposed and used in this work – VADER (Hutto & Gilbert, 2014) and LM-SA-2020 (Terblanche & Vukosi, 2021), the first lexicon being a valence-based lexicon and the second a polarity- based lexicon, economically grounded. Still, concepts of Natural Language Processing and Sentiment Analysis are defined.

The 4 chapter presents the result of the present work. This chapter addresses the case study carried out. Using the speeches of the European Central Bank as a basis, two studies are carried out: in the first, models of selection of variables are estimated to infer whether indices of economic sentiments based on NLP would be statistically significant for the estimation of economic variables; and in the second study we estimate the responses of the impulse response functions given a shock in sentiment indexes using an Autoregressive Vector – this would make it possible to understand how the estimated economic variables would behave when a "variation" of sentiments occurs in the economic scenario.

The work ends with the fifth chapter, highlighting the final considerations of the dissertation and the conclusions regarding what was done.

Chapter 2

Literature review

According to Chowdhury (2003) "Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things". In order to build the right tools and techniques to enable computer systems to comprehend and manipulate natural languages to carry out specified tasks, researchers in natural language processing seek to learn about how people interpret and use language.

Working with quantitative and qualitative data simultaneously is one of the benefits of using documents, texts, and minutes (Bholat et al., 2015, p. 1). Even while these types of studies have a wide range of applications, they have only recently been employed in economics despite the fact that working with both types of data makes it possible to conduct research and draw statistical inferences that are not achievable with simply structured data. In this way, it would be possible to have a better understanding of what happens in the economic scenario behind models that we work with NLP – understanding feelings and emotions of economic agents has always been a goal of science, whether for a better understanding of its uses or even in terms of behavior of the consumer.

2.1 Measuring News Sentiment

In a recent study, Shapiro et al. (2020) presents new evidence that incorporate NLP techniques into economic science. According to the authors, it was possible to obtain a sentiment index that matched the Michigan Consumer Sentiment Index (MCSI)[1], from economic articles and assessed financials. In terms of methodology, it would then be possible to obtain an index of consumer sentiment through selected journals and textual sentiment analysis techniques.

The authors start from the lexicons GI (Heston & Sinha, 2017), LM (Loughran & McDonald, 2011), and HL (Hu & Liu, 2004) and test their predictive capacity against human ratings

[1] http://www.sca.isr.umich.edu/

and from there, classify the models of according to goodness-of-fit measures. The LM and HL lexicons obtained similar and superior results compared to GI. Once this is done, the authors also consider the union of the three lexicons, so that human ratings are now tested against "GI + LM + HL lexicon" (Shapiro et al., 2020, p. 13). The authors still use VADER (Hutto & Gilbert, 2014) due to its predictive ability and its "negation rule" (Potts, 2010), unfortunately, the authors point out "Vader is designed for the social-media domain, not the economics/ finance domain" (Shapiro et al., 2020, p. 14).

For a solution around the predictability of lexicons, the authors decide to create their own lexicon, through their composition (and the inclusion of VADER's negation terms). Having done that, the article then proposes the creation of a sentiment index based on this lexicon, so that a monthly sentiment index was estimated from fixed effects ($\hat{f}_{t(a)}$), given the regression:

$$s_a = f_{p(a),j(a)} + \varepsilon_a \quad ,$$

where s_a is the net positivity score for article a and $f_{t(a)}$ is a sample-month (t) fixed effect. Newspapers are indexed by j and article type – either editorial or regular article – is indexed by p (Shapiro et al., 2020, p. 20). So, $f_{p(a),j(a)}$ is the fixed effect of newspaper*type: this ensures that the index is independent of changes over time, relative to its composition of newspapers and editorials, when compared to regular articles. This might be significant, as the tone of articles varies greatly between newspapers and between editorials and ordinary articles within a newspaper.

Two main studies were proposed: first, it would be analyzed whether the sentiment index elaborated by the authors would be a good predictor variable for real economic variables. For this exercise, the authors used three variable selection models as a methodological approach: a LASSO, an Adaptive LASSO, and a Group LASSO. Variable selection models showed that the sentiment index is significant in the predictor aspect, even when compared to the Michigan Consumer Sentiment Index, the LASSO Group points out the authors' index (based on sentiment techniques) as superior in some models ' 'Specifically, at least one of the versions of LASSO prefers the measure of news sentiment in forecasts of employment, output (IP), inflation, real rate (FFR) and S&P 500" (**?**, p. 26). The economic variables used for this study were: employment, output, inflation, real rate, consumption, S&P 500, MCSI and Conference Board Consumer Confidence Index, in addition to the authors' sentiment index.

In a second exercise carried out by the authors, it was analyzed how economic activity would react to a positive shock in the sentiment index. Unlike the conventional method (from an autoregressive vector), the impulse response functions were obtained through local projections Jordà (2005) (similar to VAR, but with a less restrictive approach). The results obtained demonstrate that a positive shock in the sentiment index would lead to a slight increase in consumption, in the economy's output, in the real interest rate and in the price level. In addition to these results, impulse responses were also estimated for the Michigan Consumer Sentiment Index and the Conference Board's Consumer Confidence index (CBCI). When analyzing the responses of economic activity to the MCSI, the economic variables (with the exception of the interest rate that responds positively) were not statistically significant. When analyzing the responses of the economic variables to a CBCI choue, the results are similar to the results obtained with the authors' sentiment index – the responses of the economic variables are statistically significant and point to a slight increase.

From economic articles, then, it is possible to analyze and better understand the economic situation. The extraction of a sentiment index through NLP techniques also saves the time and funds needed to obtain an index of this content. With the passage of time, the NLP techniques will evolve, allowing, thus, an improvement in the computational field and in the part of sentiment analysis applied to the economy.

2.2 Taking the fed at its word: A New Approach to Estimating Central Bank Objectives Using Text Analysis

In another article, Shapiro and Wilson (2021) analyzes Federal Reserve deliberations from NLP techniques to estimate central bank preferences. In other words, it would be possible to estimate the objective function of some generic central bank from its speeches or internal meetings.

Based on the estimation of the central bank's loss function, the results showed that the inflation target for the analyzed period (2000-2011) was approximately 1.5%. With this result, it is possible to perceive that, as everything indicates, the value of the inflation target would be, therefore, significantly below what the surveys of inflation expectations indicated in the period – for the long term. However, it has been documented (Shapiro & Wilson, 2021, p.32) that the position of certain members of the Federal Open Market Committee[2] have sometimes stated that an inflation target of 1.5% seemed consensus at least until 2009 – after this period, the inflation target value would be 2.0%.

The article also finds that the "negativity" of the FOMC is most affected by economic growth and financial conditions. This statement is supported by Walsh (2003), which addresses how central banks end up focusing more on economic growth and Coibion and Gorodnichenko (2011) which raises this hypothesis based on empirical estimations of the Taylor rule. In a previous work, Thornton (2011) has already pointed out that the FOMC ends up focusing more on the issue of "'growth in output', and not sustainable employment, the unemployment rate, or any concept of slack, as part of their policy directive from 1979 through 2008" (Shapiro & Wilson, 2021, p.34). A question also raised by the article was how financial variables would behave in the face of FOMC speeches. Bernanke and Gertler (2001) points out that monetary policy should not, for example, respond to a change in asset prices. Having an opinion that supports this study, former Fed Vice President Don Kohn says that the Fed does not respond to asset prices Kohn (2006, 2009).

[2]Where the speeches come from.

On the other hand, a study by Peek, Rosengren, and Tootell (2015) presents financial variables as good predictors of the Fed's interest rate, when these are incorporated into a Taylor rule that takes into account financial instabilities. Finally, Cieslak and Vissing-Jorgensen (2021) points out that both a stock market crash and a real negative stock market return affect FOMC discourses, and thus become predictive variables of the Fed's interest rate.

2.3 Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence

In another article, to better investigate consumer behavior and sentiment, Barsky and Sims (2012) developed two fundamental strategies: the estimation of a New Keynesian Dynamic Stochastic General Equilibrium DSGE, and the estimation of several VAR models. Through the answer of a question, "Turning to economic conditions in the country as a whole, do you expect that over the next five years we will have mostly good times, or periods of widespread unemployment and depression, or what?" (Barsky & Sims, 2012, p.1347), the authors created a variable (E5Y) so that this is the percentage of favorable answers to the question minus the percentage of negative answers to the question plus 100. The authors did the same, then, to a horizon of 12 months ahead, and called the variable E12M, in order to better understand how expectations would vary in different future timeframes. Also, they create the expected personal financial (PFE) variable, according to the answer "Now looking ahead – do you think that a year from now you (and your family living there) will be better off financially, worse off, or just about the same as now?" (Barsky & Sims, 2012, p.1371) and finally create a consumer expectation index (ICE) according to the equation:

$$ICE = \frac{PFE + E12M + E5Y}{4.1134} + 2.0 \tag{2.1}$$

another index, the consumer sentiment index (ICS) is developed from a variation of the equation (2.1):

$$ICE = \frac{PFE + E12M + E5Y + DUR + PFP}{6.7558} \tag{2.2}$$

where in the equation (2.2) DUR represents "wheater or not it is currently a good time to buy 'large household items'" (Barsky & Sims, 2012, p.1372) and PFP is identical " to PFE, except that PFP respondents to compare their current financial situation relative to one year ago"(Barsky & Sims, 2012, p.1372).

Based on this interest in a possible sentiment index, the authors estimated VAR models in order to understand how this index would behave if modeled against variables of economic activity. The authors demonstrate from impulse response functions that economic variables such as consumption would have, given a positive shock on the sentiment index, a temporary positive change. The same happens when analyzing the output of the economy.

The authors go further and implement a new Keynesian DSGE model incorporating into the model an “animal spirit” effect interpreted as “noise innovations in the signal about productivity growth”(Barsky & Sims, 2012, p.1353) and trust, represented by E5Y. Assuming that agents observe the technology level from period to period, a noisy signal of the growth rate can be observed:

$$s_t = g_t + \varepsilon_{s,t}$$

where g is an AR(1) stationary process of growth and the shock $\varepsilon_{s,t}$ is interpreted as the animal spirit shock. E5Y in turn is modeled according to the autoregressive process:

$$E5Y_t = (1 - \rho_e) + \rho_e E5Y_{t-1} + u_t$$

where u_t represents the innovation in confidence[3].

The objective of the DSGE estimation was to compare the results of the impulse responses of an anial spirit shock with the results obtained in the VAR models. This ends up having repercussions on another question: “why are confidence innovations prognostic of future movements in economic activity”(Barsky & Sims, 2012, 1356)? The estimation results corroborate previous studies Christiano, Eichenbaum, and Evans (2005); Rotemberg and Woodford (1997) on the topic.

The theoretical impulse responses obtained from the DSGE present interesting results in relation to the shock in the animal spirit variable. First, given a shock in this variable, the response of the economy’s output would be a decrease with an ever growing in the first post-shock periods, with a gradual return to its steady state in about 10 periods. The inflation response to a shock in the animal spirits variable would be a decrease in the first period with a gradual return to its steady state around 10 periods; and finally, the interest rate would undergo a slight positive variation in the first 3 periods, with a gradual return to its steady state in about 7 periods. The DSGE also analyzes the response of the consumer confidence index to a positive shock in animal spirit: this variable would respond positively to a positive shock in animal spirit, unlike the other variables, the temporary effect of consumer confidence would be present until the return from this variable to its steady state, in just over 20 periods.

The DSGE model assumes, though, that the animal spirit variable would affect the economy from its technology level, thus affecting the product. While the impact on the product is only slightly significant, this may be due to the calibration of the model. The authors calibrated the DSGE so that it followed a certain level of price rigidity: “in other words, prices are almost perfectly rigid and the central bank does not adjust interest rates to output fluctuation”(Barsky &

[3] u_t is also modeled as an autoregressive process of order 1. For further explanation, see (Barsky & Sims, 2012, p.1354)

Sims, 2012, p.1365). The results obtained through the DSGE were similar to the results obtained in L'Huillier, Lorenzoni, Blanchard, et al. (2009).

2.4 Sentiments

In another article, Angeletos and La'O (2013) investigates, also through a DSGE model, how the issue of "market sentiments" and "animal spirit" affects the equilibrium of the economy. The model also investigates what would be the effect of limited communication within a neoclassical economy, where commerce is random and decentralized. It is shown, then, that the business cycle responds to exogenous shocks of sentiment through technology, and the propagation of these shocks depends directly on the communication condition of the economy.

The model developed by the authors considers the economy divided into several "islands" Lucas Jr (1972), so that it is heterogeneous in terms of exchange opportunities, productivity factor and information. The model incorporates the sentiment shock as a random shock into the factor of productivity equation, so that:

$$x_{it} = x_{jt} + \xi_t + i_{it} \tag{2.3}$$

where ξ_t would represent the sentiment variable in the productivity factor equation (2.3). This variable ξ is affected when, for example, a firm makes an important production or employment decision and does not have the option of warning or communicating to consumers who it will find or exchange goods or even whose decision will be determine the profitability of the firm (Angeletos & La'O, 2013, p.741).

The results found by the authors indicated that the exogenous shocks of technology were propitiated, in the authors' view, by the shocks of "feelings". "These shocks impact the information that is available to each island, without, however, affecting the latter's beliefs either about the aggregate fundamentals or about the idiosyncratic fundamentals of its trading partner" (Angeletos & La'O, 2013, p.742). According to the authors, these clashes can also be called "higher-order beliefs". The specific type of higher-order uncertainty is then eliminated by requiring the aggregate fundamentals to be defined and well known, as previously pointed out by Morris and Shin (2002); Woodford (2001). "Nevertheless, by introducing trading frictions and imperfect communication, we open the door to higher-order uncertainty at the micro level: when two islands are matched together, they are uncertain, not only about each other's productivity, but also about each other's beliefs of their productivity, each other's beliefs of their beliefs of their productivity, and so on. The fluctuations we document reflect correlated variation in this kind of higher-order beliefs" (Angeletos & La'O, 2013, p.742).

The authors thus tried to explain the origins of exogenous shocks in economic fluctuations through sentiment shocks. The question that still remains open is whether these shocks could be shifts in preference and shifts in technology or even waves of pessimism and optimism in the economy. Despite using an approach that does not use NLP techniques, the search for measuring sentiments in the economy seems to reveal the importance of this abstract variable for a better understanding of the behavior of economic agents.

2.5 Macroeconomic Expectations: News Sentiment Analysis

Trying to extract sentiments from newspaper news and scientific articles to identify whether they affect macroeconomic variables, Ostapenko et al. (2020) obtains a sentiment index from "topic models" and a representation of the documents grouped by clusters. Each type of article ended up impacting a specific area or sector of economic activity: "Articles about the economy are found to be the most important for their expectations for interest rates, articles about loans matter most for in action expectations, and articles about housing are the most relevant for unemployment expectations" (Ostapenko et al., 2020, p.2). To understand how the sentiment index obtained relates to the real activity of the economy, the author chose to estimate a VAR: "The VARs for real economic activity also include the logarithm of industrial production, the logarithm of real personal consumption expenditures, the logarithm of the PCE price index, the federal funds rate, and the consumer sentiment index. The VARs for monetary policy use the logarithm of industrial production, the logarithm of the consumer price index, the one-year constant-maturity Treasury yield as a monetary policy indicator, and the excess bond premium as an indicator of financial conditions" (Ostapenko et al., 2020, p.2).

The results found by the author indicate that the output of the economy suffered a variation of 20% for a long-term horizon given a shock in the sentiment index. This clash of sentiments is also related to an increase in consumption, inflation and interest rates, albeit as a temporary change. Furthermore, the study pointed out that in terms of magnitude, a shock to the sentiments variable is more important, in long-term terms, than a shock to consumer sentiments. "The study nds that a soft news shock accounts for about 20% of the forecast error variance of output at long horizons. decomposing the main component by soft news shocks to separate topics accounts for about 7-27% of the forecast error variance of output and about 5% of the variance in consumption at long horizons in different models" (Ostapenko et al., 2020, p.40). Compared with the other models estimated by the author, a shock in the sentiment index was similar to the results estimated by Barsky and Sims (2012), in terms of responses of economic variables to the shock. The results, when analyzing a shock to the Fed news, corroborate the previous literature Jarociński and Karadi (2020); Smith and Becker (2015) which points out that the Fed news may contain information that anticipates the outlook of the general economy – the same happens when considering the shock of monetary policy news.

Text mining, sentiment analysis, and other tools enable the area of economics gradually gain a better understanding of what is going on and how structural or conjunctural events relate to social expressions. Understanding consumer behavior and how the media may affect or even alter an economic situation are both made possible by sentiment analysis. From a text or corpus, text mining makes it possible to extract both qualitative and quantitative information. NLP will eventually be employed more and more to improve our understanding of the world and economic science.

Chapter 3

Methodology

The traditional approach in terms of economic sentiment is the construction of sentiment indices through surveys (Shapiro et al., 2020, p.4) – "Usually, these surveys are monthly, with interviews and even verification of the interviewees' personal finances" (Shapiro et al., 2020, p. 5). The idea of this work is to obtain a sentiment index through sentiment analysis and Natural Language Processing techniques.

3.1 Natural Language Processing

According to Liddy (2001), Natural Language Processing (NLP) is a computational approach to textual analysis that is based on both a set of theories and a set of technologies. In other words, by the formal definition, "Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications" (Liddy, 2001, p. 2). The purpose of this tool is, therefore, "to accomplish human-like language processing". In general, the classification of a text, phrase, or even word is given categorically according to its positivity (positive, negative, neutral), or through valence scores, from 1 to 5. "The sentiment of text is a measure of the speaker's tone, attitude, or evaluation of a topic, independent of the topic's own sentiment orientation (e.g., a horror movie can be 'delightful')" (Shapiro et al., 2020, p. 5).

The present work will use lexical-based NLP techniques. "At this level, humans, as well as NLP systems, interpret the meaning of individual words. Several types of processing contribute to word-level understanding – the first of these being assignment of a single part-of-speech tag to each word. In this processing, words that can function as more than one part-of-speech are assigned the most probable part-of-speech tag based on the context in which they occur" (Liddy, 2001, p.7). Still, "Additionally at the lexical level, those words that have only one possible sense or meaning can be replaced by a semantic representation of that meaning. The nature of the representation varies according to the semantic theory utilized in the NLP system. The following representation of the meaning of the word launch is in the form of logical predicates. As can be observed, the single lexical unit is decomposed into its more basic properties. Given that there is a set of semantic primitives used across all words, these simplified lexical representations make it possible to unify meaning across words and to produce complex interpretations, much the same as humans do" (Liddy, 2001, p.7).

3.2 Sentiment Analysis and Lexicons

In the scope of sentiment analysis, it is necessary to use a dictionary of sentiments to detect polarities of sentiments and positive/negative scores. From the scores and polarities, it is possible to obtain a sentiment index that generates a possible correlation with the macroeconomic scenario. Basically, a sentiment dictionary works by indicating a specific punctuation for each word, taking into account punctuation and connectives. That is, depending on how a sentence or sentence is written, its polarity (score) varies.

3.2.1 Sentiment Lexicons

One of the most adopted and valuable resources for sentiment analysis is the use of sentiment lexicons (Ahire, 2014; Cambria, Schuller, Xia, & Havasi, 2013; Kaity & Balakrishnan, 2020; Nusko, Tahmasebi, & Mogren, 2016). A sentiment lexicon is a collection of words (sometimes referred to as polar or opinion words) linked to their sentiment orientation, that is, positive or negative (Kaity & Balakrishnan, 2020; Medhat, Hassan, & Korashy, 2014).

3.2.2 Polarity-based Lexicons

Polarity-based Lexicons are lexicons that allow you to automatically evaluate a text based on its polarity. These lexicons are "a basic resource for analyzing the sentiments and opinions expressed in texts" (San Vicente & Saralegi, 2016, p.938). When analyzing a text from the perspective of sentiment analysis, a polarity-based lexicon, words and phrases are commonly classified as positive, neutral and negative. If we take the example of the phrase "Good people sometimes have bad days", a polarity-based lexicon would possibly classify the word "Good" being a *positive* word and the word "bad" would probably be classified as *negative* – the other words in the sentence would possibly be classified as "neutral" (Berthold et al., 2007).

In terms of score classification, Berthold et al. (2007) shows that in proportional and popular values a positive score can be presented as:

$$score = \frac{\text{number of positive words - number of negative words}}{\textit{number of positive words} + \textit{number of negative words}}$$

where *number of positive words* would be the total number of positive words in a sentence or text and *number of negative words* would be the total number of negative words in a sentence or text.

A possible barrier observed is in relation to the elaboration of a polarity-based lexicon and the way in which a lexicon is created (P. J. Stone, Dunphy, & Smith, 1966). The vast majority of polarity-based lexicon is aimed at generic texts (without a specificity) – when extracting sentiments from specific texts (such as economic and financial) a lexicon could present spurious results for not taking into account words that in the economic field can be considered positive or negative (the vast majority of lexicons do not include words and economic terms such as "inflation", "recession", or other such terms) (Loughran & McDonald, 2011).

Table 3.1: Examples of polarities in a polarity-based lexicon – LM-SA-2020

Word	Polarity	Word	Polarity	Word	Polarity	Word	Polarity
Abundance	Positive	Abandon	Negative	Inspirational	Positive	Defensive	Negative
Abundant	Positive	Abdicated	Negative	Invented	Positive	Dever	Negative
Acclaimed	Positive	Abdicates	Negative	Inventor	Positive	Deficit	Negative
Accomplish	Positive	Aberrant	Negative	Leadership	Positive	Defraud	Negative
Advances	Positive	Aberrations	Negative	Leading	Positive	Defunct	Negative
Achieves	Positive	Abrupt	Negative	Lucrative	Positive	Degradation	Negative

Source: Words and polarities taken from (Terblanche & Vukosi, 2021)

Table 3.1 presents examples of a lexicon based on polarities. Note, however, that this lexicon was constructed based on economic terms and words, based on over 10,000 economic and financial articles (Terblanche & Vukosi, 2021, p.1).

In general, when a polarity-based lexicons is applied to a text, more advanced computational techniques are not necessary, except for an interactive algorithm that computes the polarity of the text (Berthold et al., 2007).

3.2.3 Valence-based Lexicons

A valence-based lexicon can be justified by the need that when analyzing a text or phrase the expected results can be not only binary (positive and negative), but determined by the "intensity" of each word or phrase (Hutto & Gilbert, 2014). When considering a valence-based lexicon, polarity is the main focus in determining the scores of (Cambria, Havasi, & Hussain, 2012) words and texts. The authors develop a valence-based lexicon so that texts and words have values such that the $score_i$ (with i being a word, text or even phrase) ranges from -1 to 1, so that $score_i \subset (-1, 1) \quad \forall \quad score_i \in \mathbb{R}$.

Even though Cambria, Li, Xing, Poria, and Kwok (2020) is a valence-based lexicon implemented in Python, compared to other valence-based lexicons (Hutto & Gilbert, 2014) gives inferior results to other valence-based lexicon.

3.2.4 VADER – Valence Aware Dictionary for sEntiment Reasoning

VADER, or Valence Aware Dictionary for sEntiment Reasoning is a lexicon initially created as a parsimonious lexicon for social media text. However, it has been used in general cases of textual sentiment analysis given it's benchmarks compared to other lexicons or even machine learning oriented techniques "relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms" (Hutto & Gilbert, 2014, p.216). Differently of most part of lexicons, VADER was created taking into account a combination of qualitative and quantitative methods to empirically validates and produces a *golden-standard* sentiment lexicon Hutto and Gilbert (2014).

Due to the fact that VADER is an open-source lexicon, it is relatively simple to modify – even if it is not what was done in this work, it would be possible, if necessary, merging VADER with some other lexicons, with the objective of creating a more complex and dense lexicon focused on economic science and finance. This lexicon has about 7520 words and textual forms with a classified score compound which after normalized varies from -1 to 1 such that:

$$score_{=}\begin{cases} positive \quad if \quad compound > 0.05 \\ neutral \quad if \quad 0.05 \geq compound \geq -0.05 \\ negative \quad if \quad compound < 0.05 \end{cases} \qquad \forall \quad compound \in (-1, 1) \tag{3.1}$$

The positive, neutral and negative scores are ratios for each category that the text or expression fells on:

> "These are the most useful metrics if you want to analyze the context & presentation of how sentiment is conveyed or embedded in rhetoric for a given sentence. For example, different writing styles may embed strongly positive or negative sentiment within varying proportions of neutral text – i.e., some writing styles may reflect a penchant for strongly flavored rhetoric, whereas other styles may use a great deal of neutral text while still conveying a similar overall (compound) sentiment. As another example: researchers analyzing information presentation in journalistic or editorical news might desire to establish whether the proportions of text (associated with a topic or named entity, for example) are balanced with similar amounts of positively and negatively framed text versus being "biased" towards one polarity or the other for the topic/entity" Hutto (2021).

Even when VADER excels when in social media, it's scores benchmarks when considered newspaper editorials are higher above the other lexicons or machine learning techniques (Table 3.2) – "Surprisingly, when we further inspect the classification accuracy, we see that VADER (F1 = 0.96) actually even outperforms individual human raters (F1 = 0.84) at correctly classifying the sentiment of tweets into positive, neutral, or negative classes" (Hutto & Gilbert, 2014, p.216).

3.2.5 Loughran-McDonald: LM-SA-2020

The other lexicon used in this work is the LM-SA-2020 and was the same provided by Loughran and McDonald (2011). Fundamentally, the difference between this one is the com-

Table 3.2: VADER 3-class classification performance as compared to individual human raters and 7 established lexicon baselines

	Correlation to ground truth (mean of 20 humans raters)	Classification Accuracy Metrics: Overall Precision	Overall Recall	Overall F1 score
NY Times Editorials (5,190 article snippets)				
Ind. Humans	0.745	0.87	0.55	0.65
VADER	0.492	0.69	0.49	0.55
Hu-Liu04	0.487	0.70	0.45	0.52
SCN	0.252	0.62	0.47	0.38
GI	0.362	0.65	0.44	0.49
SWN	0.262	0.57	0.49	0.52
LIWC	0.220	0.66	0.17	0.21
ANEW	0.202	0.59	0.32	0.35
WSD	0.218	0.55	0.45	0.47

Source: (Hutto & Gilbert, 2014, p. 223)

position: the authors developed a dictionary with the purpose of revising the traditional lexicons in which certain words are or are not considered positive or negative in the economic and financial sphere (Loughran & McDonald, 2011, p. 35):

> "The motivation for building the LM-SA-2020 word list was based on an experiment using the above-mentioned original lists to detect sentiment-carrying words in South African financial article headlines"(Terblanche & Vukosi, 2021, p. 1)

This lexicon uses 808 financial articles and only about 37% of the headlines actually corresponded to the expected sentiments (either in terms of words or expressions) given the articles verified by the authors(Loughran & McDonald, 2011). In terms of benchmark, with adding economic words and removing others in terms of polarity, sentiment detection and prediction increased by about 29% when added to NLTK's WordNet[1].

The results obtained by the authors were based on an analysis of two samples of reference articles: first, the authors considered a sample of 10 thousand files related to firms subject to shareholder litigation under Rule 10b-5. The other sample used by the authors considers Doyle, Ge, and McVay (2007), between August 2002 and November 2005, companies disclosed at least one material deficiency in internal control (Loughran & McDonald, 2011, p. 41). The authors estimated different models[2] to reach the final conclusion that the lexicon accuracy increases with

[1]https://www.nltk.org/howto/wordnet.html

[2]In fact, 28 different Logit models were estimated. The economic variables used were The number of shares outstanding times the price of the stock as reported by CRSP on the day before the file date; Book-to-market (Derived from the Compustat and CRSP data items as specified in Fama and French (2001). The variable is based on the most recent Compustat data no more than 1 year before the file date. After eliminating observations with

the addition or change of economic terms.

The lexicon created by the authors also allows for a more comprehensive classification in which, in addition to classifying certain words and terms as positive and negative, it also classifies them as "uncertainty, litigious, strong modal, and weak modal words"(Loughran & McDonald, 2011, p.62):

> "The paper finds evidence that some word lists are related to market reactions around the 10-K filing date, trading volume, unexpected earnings, and subsequent stock return volatility. [...] we show that financial researchers should be cautious when relying on word classification schemes derived outside the domain of business usage. Applying nonbusiness word lists to accounting and finance topics can lead to a high misclassification rate and spurious correlations"(Loughran & McDonald, 2011, p.62)

---

negative book-to-market, we winsorize the book-to-market variable at the 1% level); The volume of shares traded in days [−252, −6] prior to the file date divided by shares outstanding on the file date. At least 60 observations of daily volume must be available to be included in the sample; The prefile date Fama–French alpha based on a regression of their three-factor model using days [−252, −6]. At least 60 observations of daily returns must be available to be included in the sample; The percent of institutional ownership reported in the CDA/Spectrum database for the most recent quarter before the file date. The variable is considered missing for negative values and winsorized to 100% on the positive side; The average volume of the 4-day event window [0, 3], where volume is standardized based on its mean and standard deviation from days [−65, −6]; The root-mean square error from a Fama–French three-factor model for days [6, 252], with a minimum of 60 daily observations; Standardized unexpected earnings for the quarterly earnings announced within 90 days after the 10-K file date. The actual earnings and the analyst forecast consensus (mean) are from I/B/E/S unadjusted files, which are used to avoid the rounding issue. The unexpected earnings are standardized with stock price; The standard deviation of analysts' forecasts in the most recent period prior to the earnings announcement used to calculate SUE, scaled by the stock price at the end of the quarter; The monthly change in the mean of analysts' forecasts, scaled by the stock price in the prior month; and a dummy variable set equal to one for firms whose shares are listed on the NASDAQ stock exchange, else zero(Loughran & McDonald, 2011, p.63)

Chapter 4

European Central Bank Speeches: A Case Study

Given the need to corroborate and expand research in the area of sentiment analysis applied to the economic sciences – and based on the examples already observed in chapter 2, an application in a case study is proposed here.

4.1 Problem and Data Description

Following the example of Shapiro et al. (2020), Barsky and Sims (2012) and **?**, two practical exercises are proposed.

Firstly, in order to understand whether a sentiment index can be used for forecasting and estimating economic variables, and following the example of Shapiro et al. (2020), we propose the estimation of two LASSO models (L1 norm) in order to identify possible relevant variables in scenarios macroeconomic models – the first model is the conventional LASSO and the second the Adaptive LASSO. Still, in order to expand the field of estimations, an elastic net model is estimated, in order to capture improvements in model specifications and improve the selection of variables when the coefficients are close to zero when:the coefficients are close to zero, elastic net tries to take advantage of the information provided by the variables without necessarily forcing the coefficients to zero, and without considering a scenario where all variables are used in order to improve a model (L2 norm).

From there, the estimation of an autoregressive vector (VAR) is considered, also following the indications of the literature (Barsky & Sims, 2012; Shapiro et al., 2020) to better understand how a sentiment index relates to a macroeconomic scenario (and its variables). For this, the main focus of the second exercise is the exploration of impulse and response functions obtained through VAR.

The proposed exercises work fundamentally with observable European economic variables[1], with the exception of the proposed sentiment index (bag of words), made from two lexicons (VADER and LM-SA).

The other series used for the exercises were obtained from the website of FRED - Federal Reserve Bank of St. Louis - and they are: Consumer Price Index: Harmonized Prices: Total All Items for the Euro Area; Consumer Opinion Surveys: Consumer Prices: Future Tendency of Inflation: European Commission and National Indicators for the Euro Area[2]; Real Gross Domestic Product (Euro/ECU series) for Euro area; Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for the Euro Area; Harmonized Unemployment Rate: Total: All Persons for the Euro Area. Two dummies were also included in the exercise, the first referring to the Subprime mortgage crisis; and the second referring to the economic crisis generated by the COVID-19 pandemic.

Given that both series referring to sentiment indices are also not observable, they were obtained as proxies for the speeches of the European central bank[3]. The methodology discussed here is based on Loughran and McDonald (2011), taking into account its contribution to the literature when referring to the applicability of sentiment analysis applied to economics.

According to Loughran and McDonald (2011) it is possible to measure, as well as classify an economic text from the negative words in a way that the tone of these presents a correlation with economic and financial variables, since "The results to date indicate that negative word classifications can be effective in measuring tone, as reflected by significant correlations with other financial variables" (Loughran & McDonald, 2011, p. 35).

As stated by (Shapiro & Wilson, 2021, p. 13) "There is a large and growing literature aimed at quantifying sentiment from text. We use a method known as the 'Bag of Words' or 'lexical' approach, which relies on predefined dictionaries of words that are associated with particular sentiments" – in this work we also consider the polarity when taking into account the VADER – analysis from valence, given a score. Unlike VADER, where the polarity is displayed[4] the polarity of each text taking into account the LM-SA-2020 and extracting the composition of negative words in order to consider the weight of each term as a function of the total terms:

> "In the context of information retrieval [...] note that term weighting 'has an enormous impact on the effectiveness of a retrieval system.' Essentially, term weighting acknowledges that raw word counts are not the best measure of a word's information content"(Shapiro et al., 2020, p. 42)

[1] it is also worth mentioning the inclusion of the output gap, obtained from the Hodrick–Prescott filter (Hodrick & Prescott, 1997)

[2] Following the guidance of Shapiro et al. (2020)

[3] Speeches are available at https://www.ecb.europa.eu/press/key/date/html/index.en.html

[4] The polarity was obtained from the Natural Language Toolkit (Bird, Klein, & Loper, 2009) module for Python from positive, negative, neutral and compound words.

So, given the occurrence of a negative word $W_{Negative}$, its frequency is computed so that the index, based on the negative terms, is given by:

$$I_{Negative,i} = \frac{W_{Negative,i}}{W_{Total,i}} \quad ,$$

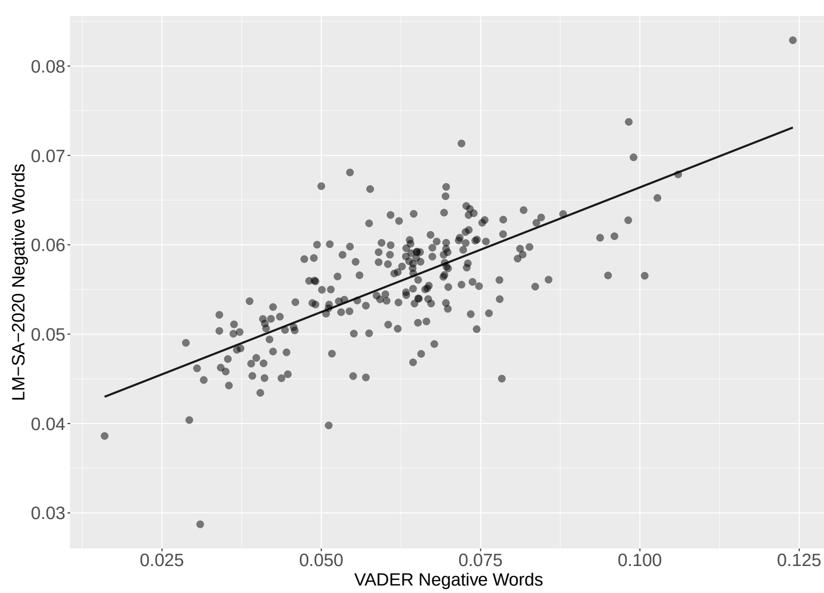
Where, $I_{Negative,i}$ represents the index score given the discourse i of the corpus; $W_{Negative,i}$ represents the number of negative words given the speech i of the corpus; and $W_{Total,i}$ represents the total number of words (positive, negative and neutral) given the speech i of the corpus. As the European central bank usually carries out more than one speech per month and the objective is the value of the monthly index, it was necessary to group the values obtained per speech around its monthly average, so that:

$$I_t = \frac{1}{n}\sum_{i=1}^{n} I_{Negative,i}$$

That is, the monthly value of the index is given by the average of the scores of the indexes of each speech in the same month.

The Figure 4.1 presents the scatter plot between the two formulated indices: VADER (valence) and LM-SA-2020 (polarity). Even though the polarity calculation methodology is different for both lexicons, the Pearson correlation coefficient between them is, as expected, positive (69.76%).

Figure 4.1: Correlation between the polarities obtained from the VADER and LM-SA-2020 lexicons (negative words)



Both indices are used in the case studies. At first, it would be recommended to (Loughran & McDonald, 2011, p. 62) the main use of a lexicon focused on economic and financial terms and words (LM-SA-2020) – however, VADER is used due to the excellent results that this lexicon presents. Furthermore, it is considered an exercise to compare both lexicons: in consideration of the fact that the LM-SA-2020 focuses on economic and financial terms and words, in contrast to VADER, which, even being aimed at media analysis, has excellent results against similar lexicons.

4.2 Experimental Setup

Since the case studies consider certain statistical and econometric techniques, it is necessary to consider some assumptions in relation to the methodologies adopted and to analyze the behavior of the time series.

The time horizon adopted was from January 2005 to December 2020 – it was not possible to extend this work until the year 2021 due to lack of data at the time of writing this: moreover, the choice of period was arbitrary, considering , however, the availability of the economic series and the availability of the speeches of the European Central Bank used. In all, the original database is composed of 132 observations so that two of the variables are sentiment indices (VADER and LM-SA-2020); two dummies are considered for the estimation of the variable selection models, namely: 1st- COVID, so that if the time point is February, March, or April 2020, COVID = 1, otherwise COVID = 0; 2nd - Debt Crisis in Europe, so that if the time period is from September 2011 to January 2013, the dummy assumes the value of 1, otherwise 0; all other variables are observable economic variables – with the exception of the output gap, obtained from the Hodrick-Prescott filter.

In addition, it is noteworthy that, for convenience in relation to the estimations and the number of observations, it was chosen to use series of monthly frequency.

Also regarding the behavior of the series, two unit root tests were performed to verify the stationarity condition: the Augmented-Dickey-Fuller (Cheung & Lai, 1995) and the Phillips & Perron Unit Root Test (Phillips & Perron, 1988) – both were estimated from the URCA package (Pfaff, 2008a) from the R software (R Core Team, 2021). Since in one of the case studies the stationarity condition is necessary for the exercise to be carried out, the series that are not zero-order I(0) integrated must be treated.

Basically, the Augmented-Dickey-Fuller and Phillips & Perron tests consist of verifying whether the time series have a unit root. In the case of the Augmented-Dickey-Fuller test, given the equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t \tag{4.1}$$

where α is the model constant, β the coefficient related to the trend, p is the order of the autoregressive process and Δ represents the difference in the series. If both α and β are not significant for the model, that is $\alpha = 0$ and $\beta = 0$, the variation of y in t will be explained only by ε, that is, by a white noise.

The Phillips & Perron unit root test, in turn, considers a model with an autoregressive process of order 1, so that:

$$y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t \tag{4.2}$$

where α is the model constant, β the coefficient related to the trend, ϕ is the coefficient related to a first order autoregressive process and ε is a white noise. Again, if $\alpha = 0$ and $\beta = 0$ we are dealing with a situation where the series y_t is explained only by a white noise, thus characterizing a random walk – that is: the series is non-stationary.

Both tests were performed for all series, with the exception of the indices of sentiments in difference, since both indices were stationary at level, these were only tested for level. In relation to the other variables tested (Table 4.1), only the Product Gap (statistically significant at 1% with trend ADF test); Producer Prices Index (statistically significant at 5% PP test) and Interest Rate (statistically significant at 10% ADF) were identified as stationary.

Table 4.1: Unit Root Tests: Augmented-Dickey-Fuller and Phillips & Perron

	Augmented-Dickey-Fuller						Phillips & Perron			
	None		Drift		Trend		Constant		Trend	
Consumer Opinion Surveys	-0.69189		-2.46604		-2.57411		-2.51257		-2.59686	
Unemployment Rate	-1.12625		-0.16789		-1.60657		-0.12221		-1.71892	
Interest Rate	-1.90764	*	-0.99368		-2.52323		-0.95965		-2.54834	
Consumer Price Index	-0.9864		-1.16925		-1.96056		-1.23476		-1.958	
Gap Gross Domestic Product	0.576651		-2.03914		-4.33739	*	-1.6031		-2.97149	
Producer Prices Index	0.726163		-3.11724	*	-2.92142		-2.99263	*	-2.64357	
Negative Sentiment (VADER)	-1.15748		-7.23406	*	-8.13027	*	-9.89749	*	-10.5938	*
Negative Sentiment LM-SA-2020)	-0.57929		-7.39231	*	-7.36532	*	-9.93006	*	-9.89294	*

Note: * symbolizes that the series do not reject the null hypotheses of the tests, in the absence of a unit root. The test statistics, for the Augmented-Dickey-Fuller test for none, drift and trend are respectively for 10%, 5% and 1%:-2.58, -1.95, -1.62; -3.46, -2.88, -2.57; -3.99, -3.43 and -3.13. For Phillips & Perron test for constant and trend are respectively for 10%, 5% and 1%: -3.48, -2.88, -2.58; -4.03, -3.44 and -3.15.

In order to follow the economic literature, and with the exception of the Product Gap variable, the series are treated as difference and, thus, the unit root tests are redone[5].

4.3 Pre-Processing Data

In order to treat the variables in order to obtain the stationarity condition, we now proceed to treat the series in difference, since the series that are not stationary are first-order I(1) integrated,

[5]The sentiment indices and Product Gap are, therefore, , worked at level

we work with the series from the first difference.

When running the tests in difference all series, both tests reject the null hypothesis of unit root at 1% (Table 4.2). The need to verify the stationarity condition becomes essential for one of the proposed case studies: the estimation of an Vector Autoregressive is not possible if any of the variables does not fulfill this condition.

Table 4.2: Unit Root Tests for the Series in Difference: Augmented-Dickey-Fuller and Phillips & Perron

	Augmented-Dickey-Fuller						Phillips & Perron			
	None		Drift		Trend		constant		trend	
Consumer Opinion Surveys	-7.38867	*	-7.36197	*	-7.37293	*	-9.22024	*	-9.2483	*
Unemployment Rate	-4.59511	*	-4.65701	*	-4.78258	*	-7.60278	*	-7.70628	*
Interest Rate	-7.35245	*	-7.56178	*	-7.53145	*	-9.97176	*	-9.9337	*
Consumer Price Index	-7.03739	*	-7.04257	*	-7.03008	*	-11.1771	*	-11.2587	*
Gap Gross Domestic Product	-6.87391	*	-6.9118	*	-6.89305	*	-6.57819	*	-6.55389	*
Producer Prices Index	-6.07706	*	-6.11674	*	-6.21462	*	-6.08898	*	-6.19294	*

Note: * symbolizes that the series do not reject the null hypotheses of the tests, in the absence of a unit root. The test statistics, for the Augmented-Dickey-Fuller test for none, drift and trend are respectively for 10%, 5% and 1%:-2.58, -1.95, -1.62; -3.46, -2.88, -2.57; -3.99, -3.43 and -3.13. For Phillips & Perron test for constant and trend are respectively for 10%, 5% and 1%: -3.48, -2.88, -2.58; -4.03, -3.44 and -3.15.

Given that the series are stationary in difference, for the VAR case study, we will use them like this, for the other exercise the series will remain level.

4.4 Modeling

Two exercises are proposed, as previously mentioned. First, a case study will be carried out taking into account the inclusion of sentiment indices when considering economic models when utilizing Variable Selection Models criteria. For this, the LASSO, Adaptive LASSO and Elastic Net models are used.

The second case study refers to an Autoregressive Vector (VAR) – impulse response functions are used to better understand the behavior of macroeconomic variables when a shock is given to one of the sentiment indices (VADER and LM- SA-2020).

4.4.1 Variable Selection Models

Variable selection models have as main objective the identification and statistically significant variables for a model. Unlike information criteria such as AIC (Akaike, 1974) and BIC (Schwarz, 1978), the variable selection models used here are variations of the conventional LASSO model. That is, the selection of variables is done by forcing the coefficients related to each variable to 0 when not significant, so that when significant, the coefficients are greater than 0.

The first method of Variable Selection Models used in this first case study is the conventional LASSO. The LASSO estimator (least absolute shrinkage and selection operator) is a linear estimation method presented in 1996 in a paper entitled "Regression Shrinkage and Selection via the lasso" (Tibshirani, 1996). For regressions and generalized regressions, this estimator was proposed as a shrinkage and selection method.

Assuming a data set such that (x^i, y_i) and $i = 1, 2, \ldots, N$, and assuming $x^i = (x_{i1}, \ldots, x_{ip})^T$ are the predictor variables and y_i is the dependent variable. Also, since $\hat{\beta} = (\beta_1, \ldots, \beta_p)^T$ the estimate LASSO ($\hat{\beta}^{lasso}$) (Tibshirani, 1996, p. 268) is given in its Lagrangian equivalent by:

$$\hat{\beta}^{lasso} = \frac{1}{2}(y - Xb)'(y - Xb) + \lambda \sum_{j=1}^{p} |b_j| \tag{4.3}$$

the penalty on $\sum_1^p |\beta_j|$ is called the L1 lasso penalty (Hastie, Tibshirani, Friedman, & Friedman, 2009, p.68) and this last constraint makes the solution nonlinear in y_i. The LASSO, however, "does not focus on subsets but rather defines a continuous shrinking operation that can produce coefficients that are exactly 0"(Tibshirani, 1996, p.286). Thus, it can be seen that the greater λ, the greater the number of coefficients defined as 0. The value of λ in this case study is obtained by cross validation following the literature(Hoornweg, 2018, p. 136).

Another estimator used in this case study as a variable selection model is the first variation of the LASSO presented here. The Adaptive lasso estimator was first described by Zou (2006) and, in contrast to lasso, augments the penalty with a vector of weights. The adaptive LASSO estimator ($\hat{\beta}^{alasso}$), then, is given by:

$$\hat{\beta}^{alasso} = \frac{1}{2}(y - Xb)'(y - Xb) + \lambda \sum_{j=1}^{p} \hat{w}_j |b_j| \tag{4.4}$$

Where $\hat{w}_j$ is a vector of weights and can "be given by $\hat{w}_j = \frac{1}{|b_{ols,j}|^{\gamma}}$, for some $\gamma > 0$" (Hoornweg, 2018, p. 115).

The third variable selection model used here is the Elastic Net model (Zou & Hastie, 2005). Despite the fact that there are more choices for other LASSO models, like the random LASSO (Wang, Nan, Rosset, & Zhu, 2011) and the group LASSO (Yuan & Lin, 2006), Elastic Net was used for being a well-known reference in the literature and for combining the L2 penalty[6]:

[6]The L2 penalty is also known as Ridge regression (Owen, 2007, p.2).

$$\hat{\beta}^{enet} = \frac{1}{2}(y - Xb)'(y - Xb) + \lambda \sum_{j=1}^{p} \left(\frac{\alpha}{2} b_j^2 + (1 - \alpha)|b_j| \right) \tag{4.5}$$

The formulation presented in (4.5) is the same formulation presented by Hastie et al. (2009). The Elastic Net model introduces a parameter α into the model so that when $\alpha = 0$ the model only incorporates the L1 LASSO penalty; and when $\alpha = 1$ the model only incorporates the L2 Ridge penalty; when $0 < \alpha < 1$ the Elastic Net model incorporates properties from both the LASSO model and the Ridge model. In this case, both λ and α were obtained from cross validation.

4.4.2 Vector Autoregression

In order to consider a deeper analysis in terms of economic simulation, and taking into account the endogeneity characteristic, a Vector Autogression (VAR) model is estimated as a second case study. In mathematical terms, we can represent a VAR(1) in a matrix form, that is, a first-order Autoregressive Vector as follows:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{4.6}$$

Where y_1 and y_2 are endogenous variables and ε_1 and ε_2 are the error terms for each equation. ϕ_{12} represents, in turn, the linear dependence of $Y_{1,t}$ on $y_{2,t-1}$ given the presence of $y_{1,t-1}$. Thus, if $\phi_{12} = 0$, then $y_{1,t}$ does not depend on $y_{2,t-1}$ when $y_{2,t-1}$ is given.

Also, according to Verbeek (2008), we can represent a first-order Autoregressive Vector as follows:

$$\overrightarrow{Y_t} = \phi + \theta \overrightarrow{Y_{t-1}} + \overrightarrow{\varepsilon_t} \quad ,$$

where, for a VAR(1) with two variables: $\overrightarrow{Y_t} = [y_{1,t}, y_{2,t}]'$ and $\overrightarrow{\varepsilon_t} = [\varepsilon_{1,t}, \varepsilon_{2,t}]$. Thus, in general a VAR(P) can be written as follows:

$$\overrightarrow{Y_t} = \phi + \theta_1 \overrightarrow{Y_{t-1}} + \cdots + \theta_p \overrightarrow{Y_{t-p}} + \overrightarrow{\varepsilon_t} \quad ,$$

where each θ_j is a matrix $k \times k$ and $\overrightarrow{\varepsilon_t}$ is a vector of length k of white noises, with a covariance matrix defined by $\sum$.

For each of its parts, the VAR model implies an ARMA model. Since the information set is expanded to additionally include the history of other variables, evaluating the components simultaneously has the advantages of being more parsimonious, containing fewer lags, and allowing for more accurate predictions (Verbeek, 2008, p.322). From a different angle, Sims (1980) asserts that using VAR models rather than simultaneous structural equations is advantageous since no 'arbitrary' limitations or previous assumptions need to be made regarding the separation between exogenous and endogenous variables.

In the same way that we can represent an Autoregressive from its moving average component, we can also write a VAR as a moving average vector. Considering the equation (4.6, we can represent it as follows:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \bar{y_1} \\ \bar{y_2} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^i \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix} \tag{4.7}$$

Where $y_{1,t}$ and $y_{2,t}$ are expressed in terms of $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{1,t}\}$ respectively. According to Enders (2008), it is possible to rewrite the previous equation in terms of $\{\varepsilon_{y1,t}\}$ and $\{\varepsilon_{y2,t}\}$. Thus, the error vector can be written as follows:

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} \end{bmatrix} = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{y1,t} \\ \varepsilon_{y2,t} \end{bmatrix} \tag{4.8}$$

combining the equations (4.10) and (4.8), we obtain:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \bar{y_1} \\ \bar{y_2} \end{bmatrix} + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{y1,t} \\ \varepsilon_{y2,t} \end{bmatrix} \quad ,$$

replacing the matrix 2×2 with a generic term to simplify the notation:

$$\Gamma_i = \frac{A_1^i}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \tag{4.9}$$

therefore, the moving media representation presented in (4.10) and (4.8) can be expressed as:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \bar{y_1} \\ \bar{y_2} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \Gamma_{11}(i) & \Gamma_{12}(i) \\ \Gamma_{21}(i) & \Gamma_{22}(i) \end{bmatrix}^i \begin{bmatrix} \varepsilon_{y1,t} \\ \varepsilon_{y2,t} \end{bmatrix} \quad , \tag{4.10}$$

or in a more compact form:

$$x_t = \epsilon + \sum_{i=0}^{\infty} \Gamma_i \varepsilon_{t-1}$$

In terms of analysis, the final equation offers a very helpful tool to look at how the variables y_1 and y_2 interact with one another. For time periods of y_1 and y_2, the effects of $\varepsilon_{y1,t}$ on shocks of

$\varepsilon_{y1,t}$ can be generated using the coefficients of Γ_i. As a result, it is clear that the four components of $\Gamma_{jk}(0)$ are effect multipliers (Enders, 2008, p.295).

By correctly adding the coefficients of the *impulse response functions*, it is possible to determine the overall effects of an impulse on $\varepsilon_{y1,t}$ and/or $\varepsilon_{y2,t}$. For instance, we can observe that the impact of $\varepsilon_{y2,t}$ on the value of y_{1+n} after n periods is $\Gamma_{12}(n)$. The cumulative sum of $\varepsilon_{y2,t}$ effects on y_1 after n periods is as follows:

$$\sum_{i=0}^{n} \Gamma_{12}(i) \quad ,$$

if y_1 and y_2 are stationary, the values of $\Gamma_{jk}(i)$ converge to zero, the larger i is. Shocks cannot, therefore, permanently alter stationary series. Thus, it follows:

$$\sum_{i=0}^{\infty} \Gamma_{jk}^2(i) \text{ is finite,}$$

the four sets of coefficients $\Gamma_{11}(i)$, $\Gamma_{12}(i)$, $\Gamma_{21}(i)$ and $\Gamma_{22}(i)$ are called the impulse response function.

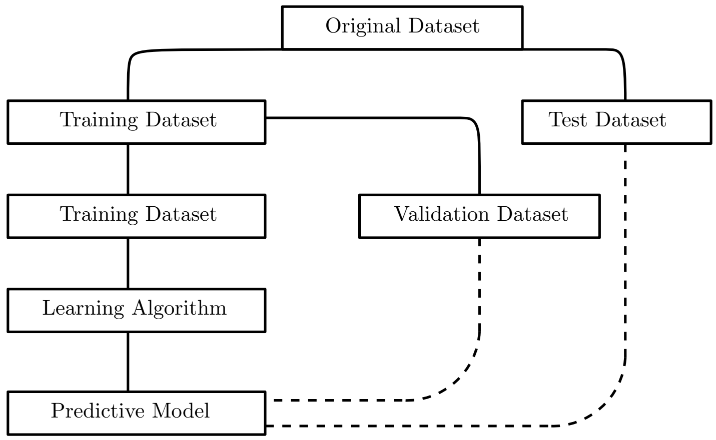
4.5 Cross Validation and Learning Algorithms

Cross Validation is a technique widely described in the literature (Breiman & Spector, 1992; Hastie et al., 2009; Hoornweg, 2018; M. Stone, 1974) that aims to obtain optimal parameters from a predefined model. In Cross Validation, the dataset is divided between a training dataset and a test dataset, these being composed of 80% and 20% of the original dataset (Breiman & Spector, 1992, 291). The model, then, "is estimated with a training sample and these estimates are used to 'predict' the outcomes of the validation sample. By varying the choice of a tuning parameters, one can select the set of configurations that leads to the best pseudo-out-of-sample forecasts"(Hoornweg, 2018, p.136).

Figure 4.2 presents a flowchart explaining how Cross Validation works. The same procedure was performed for the models estimated in this work (LASSO, Adaptive LASSO, and Elastic Net). Both the training, validation, and test dataset division were done randomly using the texttttMatrix (Bates, Maechler, & Jagan, 2022) package of the R software.

Four evaluation metrics criteria were taken into consideration when performing the cross validation procedure for the variable selection models. The R^2 Heinisch (1962), the MAE Willmott and Matsuura (2005), the MSE Bickel and Doksum (2015), and the RMSE Hyndman and Koehler (2006). Based on these criteria, the values of λ were selected for the LASSO and Adaptive LASSO models, and the values of λ and α for the Elastic Net model. For the VAR models, in

Figure 4.2: Cross Validation Flowchart



addition to the AIC and BIC selection criteria, the Hannan–Quinn information criterion Hannan and Quinn (1979) was also used.

From there, the selected models were analyzed and the results will be presented in the next section 4.6.

4.6 Experimental Results

Two exercises were performed in this section. First, the variable selection models are analyzed, and then in the second part of the section, the impulse response of the VAR models is analyzed. The R software and the Python language (Van Rossum & Drake Jr, 1995) were used to carry out this work, the codes can be found at Repository of this Dissertation.

4.6.1 Variable Selection Models

The first case study presented here concerns the estimation of three variable selection models, namely LASSO, Adaptive LASSO and Elastic Net[7]. The specification of the estimated model takes into account the addition of contemporary variables and a one-year lag in relation to the regressors and the response variable. The only variables that were not considered dependent variables in this study were the sentiment indices (VADER and LM-SA-2020) and the control dummies used (Subprime mortgage crisis and COVID-19 pandemic).

The optimal values of the λ and α parameters were obtained through Cross Validation. First, the dataset was divided between training dataset and test dataset and, finding the optimal values of the parameters, the model was evaluated against the test dataset – from evaluation metrics

[7]The calculations and estimations were made using the package glmnet (Simon, Friedman, Hastie, & Tibshirani, 2011).

such as MAE, MSE, and RMSE it is possible to identify of the best models, among the variable selection models.

For the estimation of models and "given that the LASSO model need not be parsimonious"(Shapiro et al., 2020, p. 25), all series presented here were used in the model. So the LASSO[8] equation (4.3) for representation purposes can be expressed as:

$$\hat{\beta}^{lasso} = \arg\min \sum_{i=1}^{n} \left(y_{i,t+12} - \sum_{j=1}^{p} x_{i,j} b_j \right)^2 + \lambda \sum_{j=1}^{p} |b_j| \tag{4.11}$$

where $y_{i,t+12}$ represents the interest variable i for a period of time $t+12$. In addition to the series initially presented by Barsky and Sims (2012) and with the exception of the consumption series, we introduce in this study a proxy for unemployment (Shapiro et al., 2020), and the Consumer Opinion Surveys, as in the model presented by Shapiro et al. (2020). Thus, $x_{i,j}$ represents the series i with its respective value in the periods t and $t+12$. If the variable selection models select one of the sentiment indices (VADER or LM-SA-2020), then this is an indication that the indices would help in out-of-sample predictions.

Table 4.3 and Table 4.4 present the selections of variables in each of the models, taking into account that each of the Tables presents one of the sentiment indices as a regression variable. In relation to Table 4.3, it is possible to verify that the sentiment index based on VADER is not significant when in a conventional LASSO model, in terms of unemployment estimation. Even so, when considering the estimations of the Adaptive LASSO and Elastic Net models, the sentiment index based on VADER presents a coefficient different from zero.

In terms of evaluation metric, when considering the VADER model where the dependent variable is unemployment, the smallest MAE found was that of the Elastic Net model, with a value of 0.2071393, in contrast to the value of the LASSO model, 0.209345; in relation to the MSE, the lowest was that of the Adaptive LASSO model, with a value of 0.06461785, in contrast to the LASSO, whose value was 0.0754084; finally, when we consider the RMSE metric evaluation, the lowest value obtained was that of the Adaptive LASSO model, with a value of 0.2542004, and the conventional LASSO value was 0.2746059. Thus, it is still possible to argue in relation to the choice of model, since, when considering the metric evaluation criteria, both the Adaptive LASSO and Elastic Net models proved to be superior to the estimated conventional LASSO model, considering that among the analyzed criteria the models that maintained the VADER sentiment index stood out better in relation to the model with the exclusion of the sentiment index.

When comparing the three models estimated in relation to the VADER sentiment index in this study, it is possible to notice that the results are in line with the results aligned with the liter-

[8]The model presented is the same for the Adaptive LASSO and Elastic Net models, only the penalty of each model changes, according to the specificity of each one.

ature. In the models estimated by Shapiro et al. (2020), when unemployment is considered as a dependent variable, only the LASSO model does not exclude the index developed by the author as an important variable in the model. The other two models estimated by the author, Adaptive LASSO and Group Lasso, end up not including this variable in the model[9].

When analyzing Table 4.4, the results are better. The LM-SA-2020 sentiment index as a regressor variable is present in all estimated models, LASSO, Adaptive LASSO and Elastic Net.

If comparing the evaluation metrics for each analyzed model (VADER and LM-SA-2020), the results are as follows: with the VADER sentiment index as a regressor variable, the best predictive model for Consumer Opinion Surveys was LASSO, for the Unemployment Rate was Adaptive LASSO, for the Interest Rate it was LASSO, for the Consumer Price Index it was LASSO, for the Gross Domestic Product it was Adaptive LASSO, and finally for the Producer Prices Index it was also the Adaptive LASSO.

On the other hand, when the reference regressor variable is the LM-SA-2020 sentiment index, for the Consumer Opinion Surveys, the best model analyzed was Elastic Net, for the Unemployment Rate, it was LASSO, for Interest Rate was the Adaptive LASSO, for the Consumer Price Index the LASSO and for both the Gross Domestic Product and the Producer Prices Index, the models chosen were the Adaptive LASSO.

When comparing the models with each other, that is, the models where one has the VADER sentiment index as a regressor variable and the other has the LM-SA-2020 sentiment index as a regressor variable, the models that had the VADER sentiment index slightly outperformed the models whose regression variable was the LM-SA-2020 sentiment index. The only model where the metric evaluation methods did not point out incisively was in relation to the model where the predictor variable was Consumer Opinion Surveys. With the exception of this one, the evaluation metrics indicated that the VADER index performed better on the following models with the following predictor variables: Unemployment Rate, Interest Rate and Gross Domestic Product. The LM-SA-2020 sentiment index stood out when the analyzed models had the following series as predictor variables: Consumer Price Index and Producer Prices Index.

In general, the models proved to be significant to highlight the functionality of how an economic sentiment index affects macroeconomic variables. Both indices showed significant for the estimation of the study variables in at least one of the models, LASSO, Adaptive LASSO or Elastic Net. The results obtained also corroborate the economic literature presented in this work. It is then analyzed how the economic variables would respond to a shock in sentiment indices.

[9]For clarification, when analyzing the models in which the dependent variable is "Consumption", these also did not select the sentiment indices developed by the author.

Table 4.3: Variable Selection Models: VADER Sentiment

	Consumer Opinion Surveys			Unemployment Rate			Interest Rate		
	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net
Consumer Opinion Surveys	x		x	x		x	x	x	x
Unemployment Rate	x	x	x	x	x	x	x	x	x
Interest Rate	x	x	x			x	x	x	x
Consumer Price Index	x	x	x	x	x	x	x	x	x
Gross Domestic Product	x		x		x	x	x	x	x
Producer Prices Index	x	x	x	x	x	x	x	x	x
VADER Sentiment	x	x	x		x	x	x	x	x
	Consumer Price Index			Gross Domestic Product			Producer Prices Index		
	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net
Consumer Opinion Surveys	x	x		x		x	x	x	x
Unemployment Rate			x	x	x	x	x	x	x
Interest Rate	x	x	x	x	x	x	x	x	x
Consumer Price Index			x		x	x	x	x	x
Gross Domestic Product			x			x	x	x	x
Producer Prices Index	x	x	x	x	x	x	x	x	x
VADER Sentiment	x	x	x	x	x	x	x	x	x

Note: the checkmarks symbolize that at least one of the regressors (contemporary or 12 lags) was estimated with non-zero coefficients by the LASSO, Adaptive LASSO or Elastic Net estimator indicated by the column headers. The dependent variable for each model is listed above the column headings.

Table 4.4: Variable Selection Models: LM-SA-2020 Sentiment

	Consumer Opinion Surveys			Unemployment Rate			Interest Rate		
	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net
Consumer Opinion Surveys	x		x	x		x	x	x	x
Unemployment Rate	x	x	x	x	x	x	x	x	x
Interest Rate	x	x	x			x	x	x	x
Consumer Price Index	x	x	x		x	x	x	x	x
Gross Domestic Product	x		x	x	x	x	x	x	x
Producer Prices Index	x	x	x		x	x	x	x	x
LM-SA-2020 Sentiment	x	x	x	x	x	x	x	x	x
	Consumer Price Index			Gross Domestic Product			Producer Prices Index		
	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net	Lasso	Adaptive Lasso	Elastic Net
Consumer Opinion Surveys	x	x	x	x		x	x	x	x
Unemployment Rate				x	x	x	x	x	x
Interest Rate	x	x	x	x	x	x	x	x	x
Consumer Price Index						x	x	x	x
Gross Domestic Product			x			x	x	x	x
Producer Prices Index	x	x	x	x	x	x	x	x	x
LM-SA-2020 Sentiment	x	x	x	x	x	x	x	x	x

Note: the checkmarks symbolize that at least one of the regressors (contemporary or 12 lags) was estimated with non-zero coefficients by the LASSO, Adaptive LASSO or Elastic Net estimator indicated by the column headers. The dependent variable for each model is listed above the column headings.

4.6.2 Response of Economic Activity to Sentiment Indices

To verify the relationship and behavior of the economic variables analyzed in response to a variation in sentiment indices, a Vector Autoregressive was estimated. The estimated model takes into account five different variables that represent economic activity in the euro area. The economics variables, as well as in the previously estimated variable selection models (subsection 4.4.1) were selected from the literature Barsky and Sims (2012); Shapiro et al. (2020) and are the gap of Gross Domestic Product; the Unemployment Rate; the Consumer Price Index; the Produce Price Index; and the Long-Term Government Bond Yields. In addition, the sentiment index was also included in the model – two models were estimated, each with a different index: in the first, the index based on VADER was used; and in the second, the index based on the LM-SA-2020.

Two VAR(2) models were estimated, so the model order was chosen according to the vars Pfaff (2008b, 2008c) package according to the AIC, BIC, HQ criteria. The estimated model can be represented mathematically through the equation:

$$\overrightarrow{Y_t} = \phi + \theta_1 \overrightarrow{Y_{t-1}} + \theta_2 \overrightarrow{Y_{t-2}} + \overrightarrow{\varepsilon_t} \tag{4.12}$$

Where $\overrightarrow{Y_t}$ represents the vector of endogenous variables in the period t, and $\overrightarrow{\varepsilon_t}$ represents the vector of errors in the period t. It is important to note that, for this exercise, the series that were not considered stationary according to the unit root tests (Table 4.1) are worked in first difference.

From the equation 4.12, the impulse response functions were obtained in order to visualize the behavior of the series. The error bands for the impulse response coefficients were computed by bootstrap[10] and the confidence interval used for the results follow recommendations from the literature (Shapiro et al., 2020, p. 40).

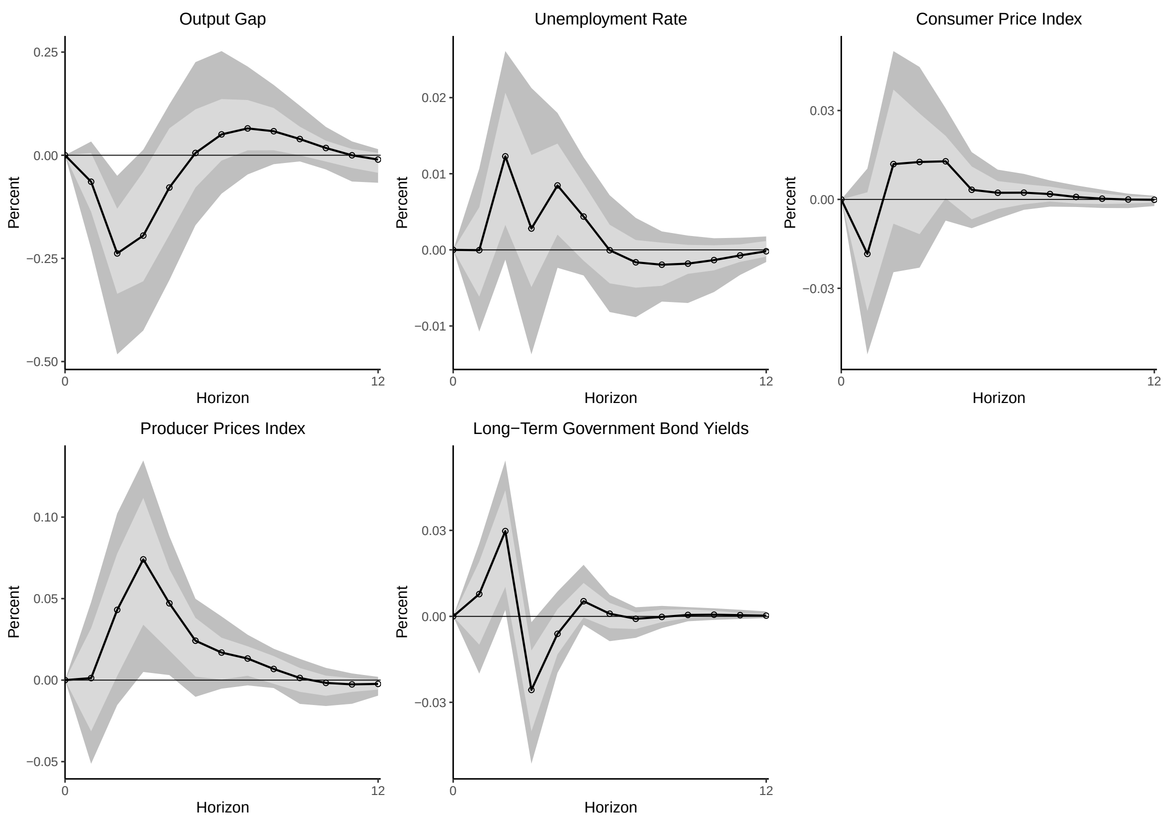
Figure 4.3 presents the results obtained from a shock on the VADER sentiment index. Result values are presented in percentage points.

From the graphs produced, it is initially perceived that the results that present statistical significance (Output Gap, Producer Price Index, and Long-Term Government Bond Yields) corroborate the results of the economic literature presented before. First, it can be seen that the Output Gap undergoes a significant negative variation given a positive shock on the VADER index – that is, given that the index is based on negativity, the higher the index score, the more negative, so that it affects the output gap negatively. The shock in the VADER index also demonstrates that a positive variation in the index would represent a positive variation in the Producer Prices Index, in a way that would generate inflationary pressure on prices for producers. Finally, the interest rate, represented by the Long-Term Government Bond Yields, presents both a positive

[10] 100 models were simulated for each variable under analysis to generate confidence interval values. Confidence intervals are determined by $CI_s = [s^*_{\alpha/2}, s^*_{1-\alpha/2}]$, where $s^*_{\alpha/2}$ and $s^*_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the normal distribution(Pfaff, 2008c, p. 17).

variation and a negative variation given a shock to the VADER index. In the first moment (period $t+2$) the variation is positive (as expected according to the literature) and in the period $t+3$ the variation is negative. The tendency is for the series to return to their stationary states in about 12 periods.

Figure 4.3: Impulse Response of a Sentiment Index (VADER) Shock on Economic Activity

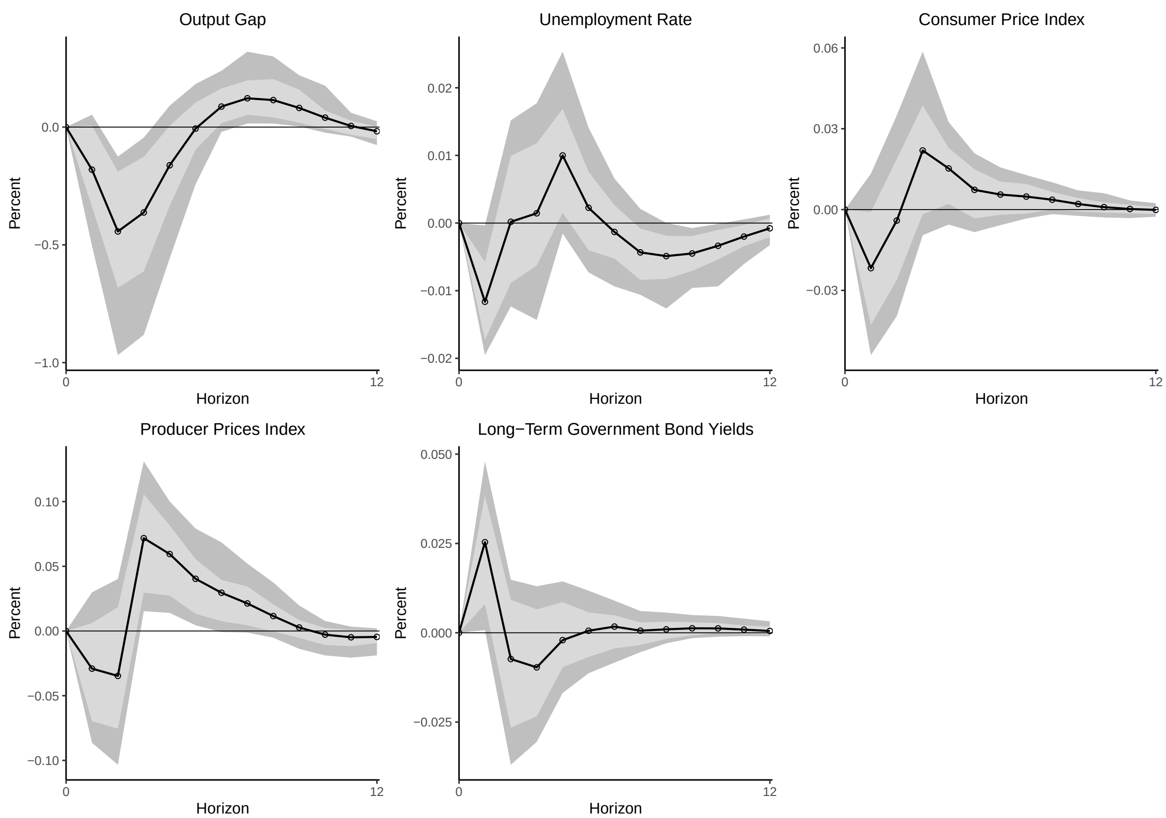


Note: Impulse Response from a sentiment index shock (VADER). The variables are described as follows: Output gap – obtained from a HP filter from Real Gross Domestic Product (Euro/ECU series) for Euro area; Unemployment Rate – Harmonized Unemployment Rate: Total: All Persons for the Euro Area; Consumer Price Index – Harmonized Prices: Total All Items for the Euro Area; Producer Prices Index – Economic Activities: Total Industrial Activities for the Euro; and Long-Term Government Bond Yields: – 10-year for the Euro Area. Following the literature, "plotted are the point estimates, 68 (light grey), and 90 (dark grey) percent confidence bands" (Shapiro et al., 2020, p. 40).

Figure 4.4 presents the results when considering the VAR(2) model (equation (4.12)) with the LM-SA-2020 index as an endogenous variable. The results obtained also corroborate the literature described earlier in this work.

A caveat is made to the results presented and in accordance with the LM-SA-2020 index, as this index has an economic and financial basis. Thus, in methodological terms, the results presented here differ when compared with the results referring to the VADER index. Also, it is worth mentioning that the model estimated with the LM-SA-2020 index has more significant impulse responses when considered statistically significant. The variables that showed statistical significance in relation to a shock in the LM-SA-2020 index are: Output Gap; Unemployment Rate; Producer Price Index; and Long-Term Government Bond Yields.

Figure 4.4: Impulse Response of a Sentiment Index (LM-SA) Shock on Economic Activity



Note: Impulse Response from a sentiment index shock (LM-SA-2020). The variables are described as follows: Output gap – obtained from a HP filter from Real Gross Domestic Product (Euro/ECU series) for Euro area; Unemployment Rate – Harmonized Unemployment Rate: Total: All Persons for the Euro Area; Consumer Price Index – Harmonized Prices: Total All Items for the Euro Area; Producer Prices Index – Economic Activities: Total Industrial Activities for the Euro; and Long-Term Government Bond Yields: – 10-year for the Euro Area. Following the literature, "plotted are the point estimates, 68 (light grey), and 90 (dark grey) percent confidence bands" (Shapiro et al., 2020, p. 40).

The results presented from Figure 4.4 also corroborate the literature presented in Barsky and

Sims (2012); Shapiro et al. (2020). Emphasizing an economic foundation in the impulse response graphs generated for the economic variables.

The results indicate that, considering a positive shock in the LM-SA-2020 sentiment index, the Output Gap would have a significant decrease in the second post-shock period, which would propagate for at least 3 more periods. When considering the behavior of the Unemployment Rate, it would suffer a slight decrease in the first and approximately in the tenth period post shock in the LM-SA-2020 sentiment index. The variable that represents the Consumer Price Index does not have statistical significance and, when we analyze the Producer Price Index, we realize that the model indicates that a positive shock in the LM-SA-2020 sentiment index would lead to an increase in the Consumer Price Index. Producer prices from the third to the fourth post-shock period. Finally, when analyzing the Interest Rate, represented by Long-Term Government Bond Yields, there is a slight increase (although statistically significant) in the first period after the sentiment index shock.

It is possible to verify that, in terms of statistical significance, when the LM-SA-2020 sentiment index (with economic foundation) is incorporated into the model, the variables begin to correspond better to it than to an index where the incorporated economic foundation is not considered.

The economic variable Consumer Price Index did not show statistical significance in either model (on the other hand, this same variable is not significant when (Shapiro et al., 2020, p.46) considers the Michigan Consumer Sentiment Index as a control variable in the VAR models estimated by the author).

The VAR models were able to show how economic activity would react to a positive shock in a sentiment index based on European Central Bank speeches. In addition, when considering an index elaborated from a lexical dictionary based on economic terms, the results are promising and corroborate the literature, pointing out that it is possible to extract information from a semi-structured database and for them to better understand how the economic scenario works or responds to a change in the relevant economic discourse of a central bank.

4.7 Discussion

From the variable selection models, it was possible to perceive that economic indices may be relevant for estimating economic variables and economic scenarios.

As much as the VADER index has no economic basis in its lexicons, both this and the LM-SA-2020 index showed statistical significance in terms of projection when considering variables of economic activity, through the LASSO, Adaptive LASSO and Elastic Net models. The results obtained corroborate the results obtained by Shapiro et al. (2020) indicating that essentially an unobservable variable representing an index of sentiments can help in the prediction of scenarios and economic activity.

When analyzing the VAR models, the results obtained were similar to those in the literature. With the exception of the Consumer Price Index variable, the other economic variables taken into account in this case study showed statistical significance when in the presence of a shock of economic sentiments.

It is also notable that when considering an index that is based on economic lexicons, the results are slightly better as exposed by Loughran and McDonald (2011). The VAR models performed better when considering the LM-SA-2020 Index, highlighting the importance of an economic lexicon when analyzing economic texts and articles.

In general, the case study proposed here corroborates the economic literature in a broad way, as well as presents interesting results for the economic analysis considering economic indices of sentiments and Natural Language Processing techniques.

Chapter 5

Conclusions

In this dissertation, it is presented how through Natural Language Processing and sentiment analysis techniques it is possible to better understand what happens in the economic scenario. Taking for example previous studies, the previous results obtained in Shapiro et al. (2020), through the variable selection models and in Barsky and Sims (2012); Shapiro et al. (2020), through the responses of the impulse response functions, were corroborated.

The sentiment indexes created in this work were elaborated from the speeches of the European Central Bank for a stipulated period from January 2005 to December 2020, the Repository of dissertation presents the algorithms and codes developed for the elaboration of this work.The sentiment indices obtained are based on two main lexicons: VADER (Hutto & Gilbert, 2014) and LM-SA-2020 (Terblanche & Vukosi, 2021). The first being a lexicon based on valence and the second a lexicon based on polarity, in order to incorporate in this work two "options" of different indices that relate to economic activity.

When in relation to the variable selection models, both indices were significant when estimated as predictors of economic activity variables. The LASSO, Adaptive LASSO and Elastic Net models showed that indices can be effectively incorporated into economic models in order to contribute to a greater predictive capacity of the model. For the VADER sentiment index, the variable selection model that stands out is the LASSO, in accordance with the metric evaluation measures; On the other hand, when analyzing the LM-SA-2020 sentiment index, the model for selecting variables that stands out is the Elastic Net. Through these models it was possible to capture the possibility of the predictive capacity of the indices in real economic variables. The variables that did not show statistical significance for the VADER sentiment index and LM-SA-2020 were only Unemployment Rate and Consumer Price Index, even so, at least one of the models (LASSO, Adaptive LASSO and Elastic Net) considers the inclusion of indices when one of these variables is an independent variable.

The estimated VAR models showed statistical significance when considering the responses of economic variables to a shock in sentiment indices. When considering a shock to the VADER index, the responses of the Unemployment Rate and Consumer Price Index variables were not significant; when considering the LM-SA-2020 index, only the response of the Consumer Price Index variable proved to be non-significant – Output gap; Unemployment Rate; Producer Price Index; and Long-Term Government Bond Yields showed a significant response to a shock in the sentiment index variable LM-SA-2020.

Even with a high correlation between the indices, the variables behaved differently when compared to the different indices. This can be explained by the fact that only one of the indices has an economic basis (Loughran & McDonald, 2011). This fact can also corroborate the point raised by Loughran and McDonald (2011); Shapiro et al. (2020), where an index of economic sentiments that was not based on an economic lexicon could bring spurious results for economic estimations.

In general, this work presents promising results for the applicability of Natural Language Processing techniques when applied to the field of economic science. Still, the estimated sentiment indices were able to predict and even relate to macroeconomic variables. Economic activity, as shown in this work, or even in previous works (Barsky & Sims, 2012; Shapiro et al., 2020), appeared to respond intuitively to shocks in the VADER and LM-SA-2020 sentiment indices. As much as sentiment analysis and text mining techniques have begun to appear in the economic literature in recent years, this field shows to be highly promising: with the advancement of computer technology and with an advance in NLP techniques, economic science would gain much in an area of research still little explored.

Bibliography

Ahire, S. (2014). A survey of sentiment lexicons. *Computer Science and Engineering IIT Bombay, Bombay*.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Akerlof, G. A., & Shiller, R. J. (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton university press.

Angeletos, G.-M., & La'O, J. (2013). Sentiments. *Econometrica*, *81*(2), 739–779.

Barsky, R. B., & Sims, E. R. (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review*, *102*(4), 1343–77.

Bates, D., Maechler, M., & Jagan, M. (2022). Matrix: Sparse and dense matrix classes and methods [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=Matrix (R package version 1.4-1)

Bernanke, B. S., & Gertler, M. (2001). Should central banks respond to movements in asset prices? *american economic review*, *91*(2), 253–257.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., … Wiswedel, B. (2007). KNIME: The Konstanz Information Miner. In *Studies in classification, data analysis, and knowledge organization (gfkl 2007)*. Springer.

Bholat, D., Hansen, S., Santos, P., & Schonhardt-Bailey, C. (2015). Text mining for central banks. *Available at SSRN 2624811*.

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes i-ii package*. Chapman and Hall/CRC.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

Breiman, L., & Spector, P. (1992). Submodel selection and evaluation in regression. the x-random case. *International statistical review/revue internationale de Statistique*, 291–319.

Cambria, E., Havasi, C., & Hussain, A. (2012). Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *Twenty-fifth international flairs conference*.

Cambria, E., Li, Y., Xing, F. Z., Poria, S., & Kwok, K. (2020). Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In *Proceedings of the 29th acm international conference on information & knowledge management* (pp. 105–114).

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, *28*(2), 15–21.

Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented dickey–fuller

test. *Journal of Business & Economic Statistics*, *13*(3), 277–280.
Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, *37*, 51–89.
Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, *113*(1), 1–45.
Cieslak, A., & Vissing-Jorgensen, A. (2021). The economics of the fed put. *The Review of Financial Studies*, *34*(9), 4045–4089.
Coibion, O., & Gorodnichenko, Y. (2011). Monetary policy, trend inflation, and the great moderation: An alternative interpretation. *American Economic Review*, *101*(1), 341–70.
Doyle, J. T., Ge, W., & McVay, S. (2007). Accruals quality and internal control over financial reporting. *The accounting review*, *82*(5), 1141–1170.
Enders, W. (2008). *Applied econometric time series*. John Wiley & Sons.
Fraccaroli, N., & Giovannini, A. (2020). Central banks in parliaments: A text analysis of the parliamentary hearings of the bank of england, the european central bank and the federal reserve.
Gennaioli, N., & Shleifer, A. (2018). A crisis of beliefs. In *A crisis of beliefs*. Princeton University Press.
Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(2), 190–195.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
Heinisch, O. (1962). Steel, rgd, and jh torrie: Principles and procedures of statistics.(with special reference to the biological sciences.) mcgraw-hill book company, new york, toronto, london 1960, 481 s., 15 abb.; 81 s 6 d. *Biometrische Zeitschrift*, *4*(3), 207–208.
Heston, S. L., & Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, *73*(3), 67–83.
Hodrick, R. J., & Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.
Hoornweg, V. (2018). *Science: Under submission*. Hoornweg Press.
Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth acm sigkdd international conference on knowledge discovery and data mining* (pp. 168–177).
Hutto, C. (2021). *VADER-Sentiment-Analysis*. https://github.com/cjhutto/vaderSentiment. (Accessed: 2022-05-01)
Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international aaai conference on web and social media* (Vol. 8, pp. 216–225).
Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679–688.
Jarociński, M., & Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, *12*(2), 1–43.
Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, *95*(1), 161–182.
Kaity, M., & Balakrishnan, V. (2020). Sentiment lexicons and non-english languages: a survey.

Knowledge and Information Systems, *62*(12), 4445–4480.
Kohn, D. L. (2006). Monetary policy and asset prices. *Monetary Policy*, 43.
Kohn, D. L. (2009). Monetary policy and asset prices revisited. *Cato J.*, *29*, 31.
L'Huillier, J.-P., Lorenzoni, G., Blanchard, O. J., et al. (2009). News, noise and fluctuations: An empirical exploration. In *2009 meeting papers.*
Liddy, E. D. (2001). Natural language processing.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, *66*(1), 35–65.
Lucas Jr, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, *4*(2), 103–124.
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093–1113.
Morris, S., & Shin, H. S. (2002). Social value of public information. *american economic review*, *92*(5), 1521–1534.
Nusko, B., Tahmasebi, N., & Mogren, O. (2016). Building a sentiment lexicon for swedish. In *Digital humanities 2016. from digitization to knowledge 2016: Resources and methods for semantic processing of digital works/texts, proceedings of the workshop, july 11, 2016, krakow, poland* (pp. 32–37).
Nyman, R., Kapadia, S., & Tuckett, D. (2021). News and narratives in financial systems: Exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control*, *127*, 104119. Retrieved from https://www.sciencedirect.com/science/article/pii/S0165188921000543 doi: https://doi.org/10.1016/j.jedc.2021.104119
Ostapenko, N., et al. (2020). *Macroeconomic expectations: news sentiment analysis* (Tech. Rep.). Bank of Estonia.
Owen, A. B. (2007). A robust hybrid of lasso and ridge regression. *Contemporary Mathematics*, *443*(7), 59–72.
Peek, J., Rosengren, E. S., & Tootell, G. M. (2015). Should us monetary policy have a ternary mandate? In *Federal reserve bank of boston 59 th economic conference: Macroprudential monetary policy.*
Pfaff, B. (2008a). *Analysis of integrated and cointegrated time series with r* (Second ed.). New York: Springer. Retrieved from http://www.pfaffikus.de (ISBN 0-387-27960-1)
Pfaff, B. (2008b). *Analysis of integrated and cointegrated time series with r* (Second ed.). New York: Springer. Retrieved from https://www.pfaffikus.de (ISBN 0-387-27960-1)
Pfaff, B. (2008c). Var, svar and svec models: Implementation within R package vars. *Journal of Statistical Software*, *27*(4). Retrieved from https://www.jstatsoft.org/v27/i04/
Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.
Potts, C. (2010). On the negativity of negation. In *Semantics and linguistic theory* (Vol. 20, pp. 636–659).
R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/
Rotemberg, J. J., & Woodford, M. (1997). An optimization-based econometric framework for the evaluation of monetary policy. *NBER macroeconomics annual*, *12*, 297–346.

San Vicente, I., & Saralegi, X. (2016). Polarity lexicon building: to what extent is the manual effort worth? In *Proceedings of the tenth international conference on language resources and evaluation (lrec'16)* (pp. 938–942).

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of econometrics*.

Shapiro, A. H., & Wilson, D. (2021). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis..

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, *39*(5), 1–13. Retrieved from https://www.jstatsoft.org/v39/i05/ doi: 10.18637/jss.v039.i05

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Smith, A. L., & Becker, T. (2015). Has forward guidance been effective? *Economic Review-Federal Reserve Bank of Kansas City*, *57*.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, *36*(2), 111–133.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Terblanche, M., & Vukosi, M. (2021). *Data statement for the lm-sa-2020 sentiment word list.* https://doi.org/10.25403/UPresearchdata.14401178. (Accessed: 2022-05-01)

Thornton, D. L. (2011). What does the change in the fomc's statement of objectives mean? *Economic Synopses*, *2011*(2011-01-01).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam.

Verbeek, M. (2008). *A guide to modern econometrics.* John Wiley & Sons.

Walsh, C. (2003). Speed limit policies: the output gap and optimal monetary policy. *American Economic Review*, *93*(1), 265–278.

Wang, S., Nan, B., Rosset, S., & Zhu, J. (2011). Random lasso. *The annals of applied statistics*, *5*(1), 468.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, *30*(1), 79–82.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347–354). Vancouver, British Columbia, Canada: Association for Computational Linguistics. Retrieved from https://aclanthology.org/H05-1044

Woodford, M. (2001). *Imperfect common knowledge and the effects of monetary policy.* National Bureau of Economic Research Cambridge, Mass., USA.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), *68*(1), 49–67.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.