NOVA IMS Information Management School

MGI

Master Degree Program in
Information Management

The Macroeconomic Impact of Jair Bolsonaro

A Synthetic Control Approach

Lara Sousa Faria

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

THE MACROECONOMIC IMPACT OF JAIR BOLSONARO

A SYNTHETIC CONTROL APPROACH

By

Lara Sousa Faria

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Supervisors: Doctor Bruno Miguel Pinto Damásio

Doctor Flávio Luís Portas Pinheiro

February 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Frankfurt, 27th February 2023.

"Quando o teu chão ceder
A solidão entrar
Não duvides
A dor acaba por passar

Se um golpe te vencer
A força não chegar
Dá ouvidos
Ao que a vida quer ensinar

Sairás mais robusto e avisado
Do que vier, não esbanjes nem um bocado

Dá um passo atrás se for preciso
Observa, sê humilde, faz o teu melhor
Acredita isso basta

Se a rejeição te escolher
Ou a justiça te falhar
Não desistas
Tens uma luz e um lugar
Dá amor, nunca o tomes como certo
Confia em ti, sem seres cínico ou ingénuo

Dá um passo atrás se for preciso
Observa, sê humilde, faz o teu melhor
Acredita isso basta

Virão momentos bons e momentos maus
Importa encará-los com um sorriso
Mal chegue o Verão ir mergulhar no mar
E visitar os avós sempre que der

Nunca percas um bom amigo
Conta com a mãe
E conta comigo
Mas cuida de ti
Acredita isso basta"

A Balada, de Jorge Cruz

Para o Vicente. Que seja tudo aquilo que ele quiser

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have supported me throughout my journey to reach this point.

In particularly, I would like to say thank you to my mother, Deolinda, father, Fernando, and sister, Vânia, I am forever grateful for your love and support. Your unwavering belief in me kept me going through the toughest times. My cousin, André Faria, I thank you for all the conversations and invaluable contributions to the person I am today – you have fed my curiosity about the world, and I am grateful for that. To Iara Figueiras and João Vassal, my forever friends, with whom I've shared Lisbon and the burden (and blessing) of being far away from our first home. I owe so much of myself to you, and you were an essential part of this academic path. To Pedro M. Pereira, for the countless hours spent by my side during these past years – it would be unthinkable not to mention you here. To my university companions and life-long friends, Carolina Domingues, David Neves, Laura Meneses, Luana Carvalho, Mariana Vagos, Marisa Pereira, Marta Malato, Matilde Mota, and Miguel Nunes, I thank you for your friendship and support. To my thesis companions in Frankfurt, Alexandre Morais, Luísa Rodrigues, Madalena Santos, and Nirav Jivan, your support and camaraderie have been priceless. A special shout out to Luísa, my companion in many more life adventures. To my colleagues and friends at the DC, Andrea Poinelli, Circé Maillet, Clemens Lehner, Elisa Saporito and Kevin Castro. I am grateful for your readiness to help and your moral support during the last stage of the thesis. To Adriana Grasso and Sebastiaan Pool, I hold you in high admiration, for being an inspiration and for unknowingly giving me the push I needed to finish this journey. Lastly, I would like to thank my supervisors Bruno Damásio and Flávio Pinheiro for guiding me throughout this journey.

This journey would not have been possible without each and every one of you. Thank you.

ABSTRACT

This thesis aims to estimate the macroeconomic impact of the policies and actions of Brazilian President Jair Bolsonaro on the country's Gross Domestic Product (GDP) using the synthetic control method (SCM) developed by Abadie et al. (2003). This methodology is a powerful tool for causal inference as it allows for the estimation of a counterfactual outcome of a treatment group (in this case, Brazil under Bolsonaro) by constructing a synthetic control group that mimics the pre-treatment trends of the treatment group. The study finds that for the second part of the 2010 decade, the real outcome variable couldn't be replicated by the synthetic control. This suggests that there may have been unique factors at play in the Brazilian economy during this time that were not present in the donor pool and therefore couldn't be fully captured by the model. The study also finds that the recovery of the pandemic downturn in GDP was steeper for the synthetic Brazil compared to the real one. This could be somewhat related to how Bolsonaro handled it. However, despite the statistical significance of the results, it is not clear that the model can give us great insights in what it comes to have an answer to our research question – to give us a concrete answer to the question of how Brazil would have performed without Bolsonaro, further research may be needed.

KEYWORDS

Bolsonaro; Brazil; GDP; Synthetic Control Methods

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

SCM	Synthetic Control Methods
IMF	International Monetary Fund
WEO	World Economic Forum
WB	World Bank
GDP	Gross Domestic Product
GDP pc	Gross Domestic Product *per capita*
EMBI	Emerging Markets Bond Index
RMPSE	Root Mean Squared Prediction Error
DID	Difference-in-differences
MV	Missing Values
BPS	Basis Points
DP	Donor Pool

1. INTRODUCTION

Jair Bolsonaro is a Brazilian politician and retired military officer who served as the President of Brazil since January 1, 2019 up until the end of 2022. He is a member of the Liberal Party (PL). He is known for his controversial and polarizing statements and actions. Bolsonaro rose to prominence during the 2018 presidential elections, where he campaigned on a platform of economic liberalization and a hardline stance on crime and corruption. He won the election with 55% of the vote, defeating the Workers' Party (PT) candidate Fernando Haddad in the second round.

The Brazilian economy has gone through several stages of development in the last decades. In the 1990s, the country was facing high inflation. To tackle this issue, the government implemented the Real Plan, which pegged the Brazilian currency to the US dollar. This measure, along with others, was successful in curbing inflation and stabilizing the economy, resulting in a period of strong economic growth. This growth was driven by a combination of factors such as favorable commodity prices, increased foreign investment, and a growing domestic consumer market.

In the early 2000s, the Brazilian economy continued to experience strong growth, driven by high commodity prices, a growing domestic consumer market, and increased foreign investment. However, the global financial crisis of 2008 had a negative impact on the country's economy, leading to a period of slow growth and high unemployment.

The 2010s were marked by a period of economic instability and slow growth for Brazil. The country was hit by a combination of factors such as weak commodity prices, political instability, and a decline in business confidence. Despite the government's efforts to implement policies to stimulate economic growth, the country struggled to recover.

The Brazilian economy in the 2010s was marked by a period of economic instability and slow growth. In the early 2010s, the country was still recovering from the 2008 global financial crisis, which had a negative impact on the country's economy, leading to a period of slow growth and high unemployment. However, the country experienced a rebound in the latter half of the decade, driven by high commodity prices, a growing domestic consumer market, and increased foreign investment. However, this growth was short-lived as the latter half of the decade was marked by a combination of factors such as weak commodity prices, political instability, and a decline in business confidence. Additionally, the "Lava Jato" (Carwash) operation, a widespread corruption investigation that targeted state-run oil company Petrobras and other major Brazilian corporations, had a significant impact on the country's economy and political landscape. This led to a period of economic instability and slow growth for Brazil. Despite the government's efforts to implement policies to stimulate economic growth, the country struggled to recover.

In recent years, the Brazilian economy has been affected by the global economic crisis caused by the COVID-19 pandemic, which has led to a sharp decline in economic activity and high unemployment. The government has implemented a series of measures to try to mitigate the impact of the crisis, but the outlook for the economy remains uncertain. The country also has been facing internal and external challenges such as political instability, social unrest, and trade tensions

During Bolsonaro's tenure as President of Brazil, his administration implemented a number of economic policies aimed at stimulating growth and attracting foreign investment, such as deregulation, tax reductions, and labor market reforms, as well as privatizing state-owned companies in sectors such as energy, telecommunications, and infrastructure. During Bolsonaro's mandate, the COVID-19 pandemic had a significant impact on Brazil, as the country experienced one of the worst outbreaks in the world. Bolsonaro's handling of the pandemic was heavily criticized for its lack of a clear national strategy and for downplaying the severity of the virus. Additionally, many people criticized Bolsonaro for his handling of the economic consequences of the pandemic, accusing him of prioritizing the needs of large corporations over those of small businesses and regular citizens.

This study aims to quantify the impact of Jair Bolsonaro's presidency on the Brazilian economy using analytical research, similar to studies conducted on other populist leaders in the past. The research question being addressed is "What is the macroeconomic impact of Jair Bolsonaro's administration on Brazil?". The methodology used to conduct this study is a Synthetic Control Method (SCM). Developed by Abadie and Gardeazabal (2003) , SCM is a statistical method that creates a synthetic control group using a combination of similar units from a donor pool to estimate the counterfactual of a specific event or policy change. This synthetic counterfactual mimics the outcome of the target variable before the intervention, in this case, the election of President Bolsonaro. The study will specifically focus on evaluating the impact of Bolsonaro's administration on real GDP in Brazil by comparing the observed real GDP after his election with the counterfactual scenario of what the real GDP would have been if he had not been elected.

This research looked at two groups of countries - one without any restrictions and another that had similarities to the Brazilian economy. Findings indicate that it is challenging to pinpoint the precise effect of the Bolsonaro Administration on the Brazilian economy. The chosen model accurately predicted the real GDP up until 2014, but its accuracy decreased afterwards, making it difficult to assess the extent to which the Bolsonaro Administration contributed to the gap between the model's predictions and the actual results after taking office in 2019. The results, although statistically significant, do not allow for firm conclusions to be drawn. More research is required to fully understand the complex dynamics at play.

This document is structured as follows: Chapter 2 provides a literature review on the Synthetic Control Method and its application in the study of economic policy. Chapter 3 explains the methodology in detail, including its formal aspects, the assumptions, and data requirements. Chapter 4 presents the empirical study, in which we estimate the macroeconomic impact of the Bolsonaro administration using SCM. Chapter 5 presents and discusses the results of the study. Finally, Chapter 6 concludes, with a summary of the main findings and suggestions for future research.

2. LITERATURE REVIEW

2.1. EVALUATING A LEADER'S MACROECONOMIC IMPACT

There is a significant amount of literature on the macroeconomic impact of leaders in the macroeconomy, with varying perspectives and conclusions.

One viewpoint is that of "policy-based" leadership, which holds that the actions and decisions of leaders can have a direct and measurable impact on macroeconomic outcomes such as GDP growth, inflation, and unemployment. Research in this area examines how specific policies, such as increasing government spending affect macroeconomic performance. For example, a study by Alesina et al. (2010) found that when leaders implement fiscal consolidation policies such as reducing government spending and increasing taxes, it leads to better macroeconomic outcomes like lower deficits and debt-to-GDP ratios. On another study, Alesina et al. (2006) found that leaders who adopt more market-oriented policies tend to have better economic outcomes.

Another perspective is that of "leadership style" or "leadership traits," which theorizes that the personal characteristics or leadership styles of leaders can indirectly influence macroeconomic outcomes through their effect on the broader economic and political environment. Studies in this area have examined the relationship between leadership traits such as "charisma" or "risk-taking" and macroeconomic outcomes. Ferguson (2006) argues that the development of finance and financial institutions is closely tied to the rise of powerful leaders and states. He contends that throughout history, leaders who have been able to establish stable and efficient financial systems have been able to finance wars, stimulate economic growth, and create the conditions for long-term prosperity.

A third standpoint is that of "institutional" or "systemic" leadership, which emphasizes the role of broader political and economic institutions in shaping macroeconomic performance, and the limited ability of individual leaders to affect outcomes. Studies in this area have examined the relationship between institutional factors such as democratic governance, the rule of law, or the level of economic development, and macroeconomic performance. Jones et al. (2005) find that leaders are more influential in autocratic systems with few constraints on the executive. Tabellini (2008) examines the relationship between constitutional design and economic performance, arguing that certain constitutional features are more conductive to economic growth, regardless of the leadership profile (see also Martins and Damásio (2020)).

Overall, the literature on the macroeconomic impact of leaders is diverse and multifaceted, with different studies emphasizing different aspects of leadership and reaching different conclusions about the relationship between leadership and macroeconomic outcomes. It is also worth noting that there are some criticisms of the literature on leadership and macroeconomics, one of the main criticisms is that it is hard to separate the impact of leadership from the impact of other factors that affect macroeconomic outcomes, such as global economic conditions (Damásio, Louçã and Nicolau (2018)), structural features of the economy (Damásio and Nicolau (2020)), or random fluctuations (Damásio and Mendonça (2019) , Damásio and Mendonça (2022)). Rodrik (2015) discussed the limitations of economic models to explain the complexity of the real-world economies and has emphasized the need for more context-specific and historically informed analysis. The author also warns against the overgeneralization of cross-country studies and the overemphasis on the role of leaders in shaping economic outcomes.

2.2. SYNTHETIC CONTROL METHODS

The synthetic control method, a statistical technique developed by Abadie and Gardeazabal (2003), has gained widespread utilization in the estimation of the causal effect of policy interventions on particular outcome variables. The method entails the construction of a synthetic control group that mimics the trend of the outcome variable in the treated group prior to the intervention by combining observations from multiple untreated units that possess the most resemblance to the treated unit based on a set of pre-intervention covariates.

The emergence of this methodology is driven by the growing demand for quantitative methods that can complement and enhance the qualitative analysis of comparative studies, as emphasized by Tarrow (1995), Sekhon (2004), and Lieberman (2005). The synthetic control method has been applied to a wide range of issues, such as estimating the costs of terrorism in the Basque country (Abadie and Gardeazabal, 2003), where the authors used the method to construct a synthetic control group for Spain, which was the first country in Europe to experience a significant number of terrorist attacks. The study found that terrorist attacks had a detrimental effect on economic activity in Spain. Other applications of this methodology include estimating the impact of California's anti-tobacco policy on tobacco consumption (Abadie, Diamond, and Hainmueller, 2010), and estimating the economic impact of the 1990 reunification of Germany on West Germany (Abadie, Diamond, and Hainmueller, 2014).

However, other researchers have also applied this methodology to various problems, such as assessing the impact of natural disasters on economic growth (Cavallo et al., 2010), evaluating the impact of economic liberalization (Billmeier et al., 2013), analyzing the impact of inflation targeting in emerging countries (Wang-Sheng, 2011), and examining the value of bank's political connections during tumultuous times (Acemoglu et al., 2013) among others.

On the other hand, there are several shortcomings and limitations of the present model that have been pointed out by scholars. Abadie (2021) himself has acknowledged some of these limitations in his own research and proposed modifications to address them. One of the main criticisms of the method is that it may be sensitive to the choice of control units, leading to large changes in the estimated treatment effect with small changes in the selection of control units. Another limitation is that the method may not be able to account for time-varying confounding factors that could affect the comparison between the treated and control units. For example, in the paper "Terrorism and the World Economy" (2003), Abadie and Gardeazabal used the synthetic control method to estimate the economic impact of terrorism on Spain, but some researchers have argued that the method may not have fully accounted for other factors that could have affected the Spanish economy during the same period. Additionally, the synthetic control method may not be well-suited for cases where the treatment effect is heterogeneous across different subgroups of the population. Finally, there may be limitations in robust and efficient computation of synthetic controls.

Overall, the synthetic control method has proven to be a valuable tool in the estimation of causal effects, and it has been applied to a wide range of issues, providing insights into the impact of policy interventions on various outcome variables. Despite its limitations, researchers continue to use the synthetic control method, while also acknowledging it and working to improve the methodology for future research.

2.3. Synthetic control methods to evaluate leader's impact

In recent years, there has been a growing literature using SCM that aims to study the economic impact of countries' leaders, especially when it comes to populist governments.

For example, Grier and Maynard (2016) use the SCM to perform a case study in order to estimate the impact of Hugo Chavez on the Venezuelan economy. Their results indicate that his administration had a substantial negative impact on per capita income, but no significant effect on poverty, health, or inequality. In a related study, Absher et al. (2020) use the SCM to estimate the economic impact of left-populist regimes in Latin America, namely Venezuela, Nicaragua, Bolivia, and Ecuador. They find that these countries ended up over 20% poorer, on average, compared to the average of their synthetic counterfactual, whilst regarding income inequality or infant mortality there was no significant improvement.

Funke et al. (2020) engage in a comprehensive study, identifying 50 populist presidents and prime ministers from 1900 until 2018. They find that there is a negative impact on GDP per capita, relative to a synthetic non-populist counterfactual. Recently, Born et al. (2021) use the SCM to estimate the effect of President Donald Trump on some macroeconomic indicators for the US, such as GDP, unemployment and the labor force. They did not find significant evidence of a Trump effect compared to a synthetic counterfactual.

Overall, these studies demonstrate the potential of synthetic control methods to evaluate the impact of leaders, particularly populist leaders, on macroeconomic outcomes such as economic growth, public debt, and institutional quality. However, it's important to note that these studies provide evidence on the correlation between populist leaders and macroeconomic outcomes, but it is not clear if it is a causal relationship and other factors could be in play. Furthermore, synthetic control methods are based on assumptions that are testable, and there could be concerns about the validity of the synthetic control group, sensitivity to the choice of covariates, and the potential for bias. Therefore, it's crucial that researchers continue to acknowledge the limitations of the synthetic control method, while also working to improve the methodology in future research.

3. METHODOLOGY

SCM were first developed by Abadie and Gardeazabal (2003) as a means of evaluating the impact of specific events or interventions on a target variable of interest at national, regional, or sectoral levels. These methods are particularly useful in small-sample comparative studies, where traditional difference-in-differences (DID) estimations or other regression analyses may not be feasible due to the need for large sample sizes and multiple observed instances of the event of interest.

The fundamental question that SCM aims to answer is "what would have happened to the development of a target variable in the affected region if a particular event or intervention had not occurred?" - in other words, the goal is to estimate the causal impact of the intervention on the target variable.

The SCM approach involves constructing a control region, also known as a synthetic counterfactual, using a linear combination of several alternative regions where the intervention has not taken place. This synthetic control region is then compared to the real region over time, both before and after the intervention has occurred. The idea behind this approach is that a combination of unaffected units may provide a more appropriate comparison than any single unaffected unit alone.

The procedure for selecting comparison units is formalized through a data-driven process, which allows for the estimation of the impact of the intervention at some point in time. The difference between the synthetic control and the development of the real region over time is considered to be the impact of the intervention.

It is important to note that while SCM is a powerful method for inferring causality in small-sample comparative studies, it is based on certain assumptions and requires certain data and conditions to be met. Therefore, it is essential to carefully consider these assumptions, data requirements, and conditions when applying the method to ensure the validity of the results.

This chapter aims to provide a general overview of the formal aspects of the method, including its assumptions, data requirements, and conditions under which causality can be inferred, in order to gain a comprehensive understanding of SCM and its potential applications in evaluating the impact of specific events or interventions on a target variable of interest. The chapter will also discuss the process of constructing a control region, also known as a synthetic counterfactual, using a linear combination of several alternative regions where the intervention has not taken place, and comparing it to the real region over time. The procedure for selecting comparison units will be formalized through a data-driven process.

3.1.1. Formal Aspects

Formal descriptions of this methodology can be consulted in several previous published works, such as in Abadie, Diamond, and Hainmueller (2010), Abadie (2021), Adhikari (2021), or Cunningham (2021).
Here, only a brief formal overview will be provided.

Suppose that there are J units: $j = 1, 2, \dots, J$, in which the first unit ($j = 1$) is the treated one, that is, the one affected by the intervention in study. Then, the remaining $J - 1$ units correspond to ones which were not affected by the intervention – the control group. Moreover, assume that there is a total of T periods of time, T_p pre-intervention periods and $T - T_p$ post-intervention periods. T_o corresponds to the intervention period. Furthermore, for each unit j and time t, the outcome of a variable Y_{jt}, along with a vector of unit characteristics ($X_{1j}, \dots, X_{kj}$) are observed. Afterwards, for the unit affected by the intervention, the variable outcome observed with and without the intervention is defined as Y_{1t}^I and Y_{1t}^N, respectively. Given this, the effect of the intervention (α_{1t}) under analysis in period t (with $t > T_o$) is measured as:

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N \tag{1}$$

It is straightforward to understand that for $t > T_o$, given that unit 1 is affected by the intervention, then $Y_{1t} = Y_{1t}^I$. Now the challenge is to estimate the value of Y_{1t}^N, i.e., what would be the outcome of the affected unit in the absence of the intervention – the counterfactual outcome. To obtain this value, a synthetic equivalent is created using the units $j = 2, \dots, J$ in the control group. The synthetic control is estimated as the weighted combination of the control units. Given a set of weights (w), the synthetic control estimator of Y_{1t}^N is represented by:

$$\hat{Y}_{1t}^N = w_2 Y_{2t} + \dots + w_J Y_{Jt} \tag{2}$$

And the synthetic control estimator of α_{1t}:

$$\hat{\alpha}_{1t} = Y_{1t}^I - \hat{Y}_{1t}^N \tag{3}$$

Furthermore, SCM establish a restriction on weights, which must be nonnegative, and add up to one, with $w_j \in [0,1]$. This restriction may be relaxed, however, in such cases, the model may allow extrapolation, as thoroughly explained by Abadie, Diamond, and Hainmueller (2014).

The next step consists in obtaining the optimal weights value. Literature proposes that $w_2, \dots, w_J$ are chosen so that the resulting synthetic control best resembles the characteristics of the affected unit, prior to the interventions, i.e., the authors propose to choose a synthetic control, $w^* = \{w_2^*, \dots, w_J^*\}$ that minimizes:

$$\sum_{i=1}^{k} v_i \ (X_{i1} - \sum_{j=2}^{J} w_j X_{1j})^2 \tag{4}$$

Where $v_1, \dots, v_k$ represent the relative importance of the synthetic control reproducing the values of the unit characteristics $X_{11}, \dots, X_{k1}$ introduced in the model to calibrate the weights.

It is easy to see that the expression (4) above can be minimized, for a given set of weights, $v_1, \dots, v_k$, using a quadratic optimization, constraining the values of $w_2^*, \dots, w_J^*$ to be positive and sum to one. Yet, there is a remaining challenge: how should $v_1, \dots, v_k$ be chosen?

The answer lies in choosing $v_1, \dots, v_k$ so they produce the best fit in terms of how closely the synthetic control describes the outcome variable during the pre-intervention period, i.e., choosing $v_1, \dots, v_k$ so that the prediction error ($Y_{1t} - \hat{Y}_{1t}^N$) is minimized. To facilitate this procedure, Abadie, Diamond, and Hainmueller (2011) have developed an R package entitled *Synth*.

3.1.2. Assumptions

Abadie (2021) describes a list of contextual requirements under which SCM are a well-suited tool for policy evaluation. Most of the listed requirements regard not only synthetic control methods, but also other types of comparative study research design. The assumptions are as follows:

First, the estimated effect of the intervention must be large enough to differentiate it from the random shocks that affect the outcome's volatility. Additionally, the volatility of the outcome should be monitored, and if it is significant, it should be removed through time-series filtering. Second, a comparison group must be available. This group should consist of units that have not suffered similar interventions or have not been affected by idiosyncratic shocks as a result of the intervention under study. The comparison group should also have similar characteristics to the treated unit. Third, the possibility of anticipation by forward-looking agents must be taken into account. It may be useful to backdate the intervention to the moment where it is announced, rather than when it is implemented. Fourth, control units that have been significantly indirectly affected by the intervention through spillover effects should not be included in the comparison group. Last, the differences in the characteristics of the treated unit and the synthetic control should be small. The treated unit characteristics should fall within the convex hull of the control units' characteristics.

3.1.3. Data Requirements

Additionally, in the literature, there are certain data requirements that need to be met to conduct an effective analysis.

Firstly, it is necessary to collect data on the outcomes and predictors of the outcome for both the unit affected by the intervention and the set of units in the control group. Secondly, it is crucial to gather sufficient pre-intervention information, as synthetic control estimators tend to become more accurate with a long pre-intervention period. Thirdly, collecting enough post-intervention information is necessary to ensure that the effect of the intervention is observed.

3.1.4. Goodness of fit and Casual Inference

To check if the comparison unit created using SCM can be used as a valid counterfactual, we must compare it to the treated unit before the treatment to see how similar they are. Abadie et al. (2010) propose using the RMSPE of the outcome variable to evaluate whether the treated unit and its synthetic counterpart have a good fit or not. The RMSPE is defined as:

$$RMSPE = \sqrt{\frac{1}{T_P} \sum_{t=1}^{Tp} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j * Y_{jt}\right)^2} \tag{5}$$

If the RMSPE is 0, then the synthetic control perfectly matches the pre-intervention trajectory of the outcome variable. However, when that is not the case, it becomes harder to match the RMPSE with the synthetic control respective goodness of fit. Adhikari et al (2016) introduce a “pretreatment fit index”, which facilitates the assessments of the quality of the fit. This index is a ratio, defined as:

$$Fit\ Index = \frac{RMSPE}{benchmark\ RMSPE} \tag{6}$$

Where the benchmark RMSPE is defined as:

$$Benchmark\ RMSPE = \sqrt{\frac{1}{T_P} \sum_{t=1}^{Tp} (Y_{1t})^2} \qquad (7)$$

If the fit index is α, it means that the RMSPE is equivalent to the RMSPE obtained when the treated and synthetic unit differ by α percent in each year prior to the treatment. A fit index of 0 indicates a perfect fit, while a fit index of 1 indicates that the fit is similar to a synthetic control that is twice as large - or by construction half as small - as the treated unit. However, the fit index can be greater than 1 if the outcome variable of the treated unit is much larger (or smaller) than that of the synthetic unit, which usually suggests a poor fit.

Having a satisfactory goodness of fit is crucial, but it does not suffice to infer causality. The credibility of results obtained through the SCM is a challenging aspect, particularly when the sample size is small and only one or few units are affected by the intervention. According to the literature, the standard inference techniques with standard deviations and confidence intervals are inapplicable in most SCM applications. To address this, the literature suggests many approaches.

For the purpose of this study, the one proposed by Abadie et al. (2010) will be followed. The method consists in constructing p-values based on Fisher (1935) which is widely used in the literature. The method involves iteratively applying the SCM to each unit in the donor pool to obtain a distribution of placebo effects. Then, the treatment unit's ratio of post-treatment to pre-treatment root mean squared prediction error (RMSPE) is compared to this distribution to determine if it is extreme. The p-value is calculated by sorting the ratio in descending order and determining the treatment unit's position in the distribution. This method is used to confirm the statistical significance of the observed difference. The idea behind this method is that, if the observed difference between the two series is nothing more than prediction error, any model chosen would've done that, even if there was no treatment effect. The null hypothesis used in this test is the "no treatment effect whatsoever", which is the most common null used in the literature. However, this test relies on the large sample properties which may not be present in many small sample SCM applications. Therefore, this p-value should be considered as a gauge of the validity of the result, but it should be considered alongside other robustness checks and intuition of the researcher.

4. EMPIRICAL STRATEGY

The present chapter aims to provide a detailed description of the empirical study performed, including information on the data such as its sources, data cleaning and preparation, descriptive statistics, and the development of the synthetic control model. It comprises two sections.

The first section of the chapter will focus on the data - data sources used in the study, the data cleaning and preparation process, including the description of the necessary transformations and missing values treatment. Descriptive statistics will then be presented, to give an overview of the characteristics of the data used to model our study.

The second section will provide a detailed explanation of the empirical application of the synthetic control method, including information on how the control units – donor pool - were selected, the criteria used to weight the control units, and any assumptions made in the process.

4.1. DATA

The data used to develop this study draws upon two primary sources: the database from the World Economic Outlook (WEO) October 2022, published by the International Monetary Fund (IMF), and data from the World Bank repository. The WEO dataset contains information on economic growth, inflation, trade, and financial conditions in countries around the world, as well as projections for the near future. From the World Bank data set, we were able to collect information on consumption and exports, which were lacking from the WEO database.

The initial dataset comprises information on GDP, GDP growth, consumption, investment, savings, exports, exports growth and unemployment for 196 countries, for the period from 1980 until 2027. Descriptive statistics for the panel can be found in Appendix A.

Since these are aggregated official data, there are some steps regarding data quality with whom one does not have to worry about, and thus, for the purpose of this study, those steps will be skipped. However, there's still a substantial amount of missing values that must be studied. Additionally, in order to develop the method, there are some variable transformations that will be applied to the initial series. In this section, those procedures will be described.

4.1.1. Missing values treatment

In this subsection, the methodologies employed to address missing values within the dataset are outlined. The techniques utilized in this study include deletion and imputation.

Upon examination of the distribution of missing values, it was observed that there is a significant percentage of missing values for the years prior to 1990. As a result, all observations corresponding to this period were removed. Furthermore, there were still a considerable number of countries for which a substantial amount of values were missing for most years. These countries, whose missing values represented more than 20% for at least one of the variables under study, were also removed. This resulted in a final panel of 110 countries.

Another challenge encountered was the fact that the series were sourced from different databases. The WEO database includes estimates for the near future, which are based on projections performed by the IMF staff. These estimates begin after 2021 and extend until 2027. Conversely, the series obtained from the World Bank database for exports and consumption end in 2021.

To fill in the missing values for exports, the IMF WEO database's yearly export growth data - including projections - was utilized. The missing values were imputed by extending the latest data point and assuming the growth rate estimated by IMF staff projections. As for consumption it was assumed that this aggregate would grow at the same rate as GDP. It is acknowledged that this assumption may not always hold true, such as during an economic recession, when GDP may decrease while consumption remains stable or even increases. However, it generally makes sense to assume that consumption and GDP will follow similar trends, as consumption is a major component of GDP. Figure 4.1 below exhibits the pattern for consumption and the respective projections for Brazil, using this described approach.

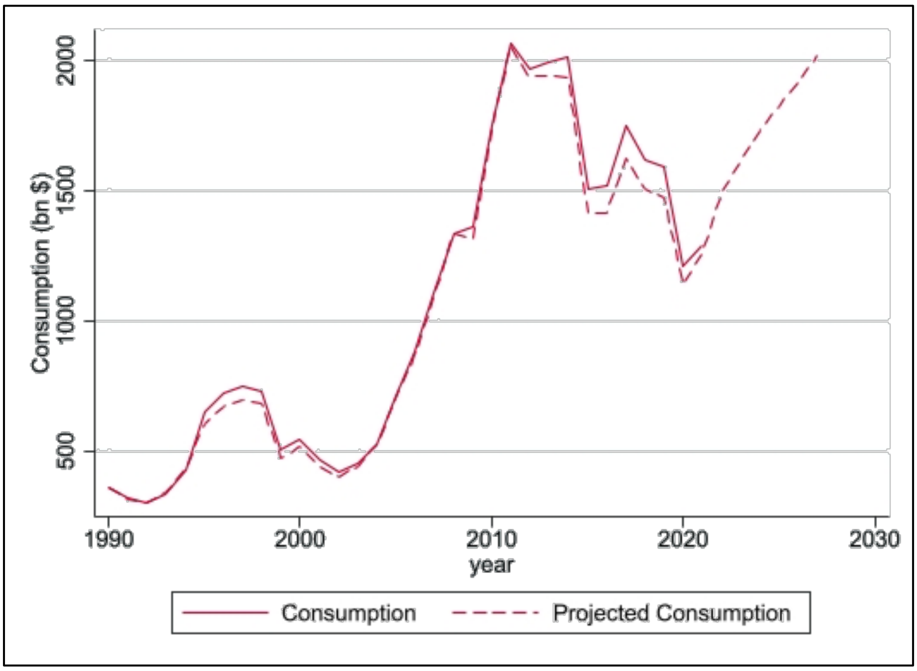

Figure 4.1 – Consumption vs Projected Consumption for Brazil

Despite the previously implemented procedures, a residual quantity of missing values persisted, as it can be observed in Table 4.1. Ultimately, it was decided to drop all variables for which the percentage of missing values exceeded 3%, specifically the variables *unemployment* and *gdp_growth*. Finally, the remaining residual missing values were then interpolated linearly. This completes the process of addressing missing values.

Table 4.1 – Missing values per variable

Variable	Missing	Total	% Missing
gdpnom	32	4,180	0.77
gdppcnom	32	4,180	0.77

gdppcrealppp	15	4,180	0.36
exportsgrowth	53	4,180	1.27
exports	34	4,180	0.81
consumption	34	4,180	0.81
inv	28	4,180	0.67
sav	33	4,180	0.79
pop	20	4,180	0.48
unemployment	1,449	4,180	34.67
gdp_growth	142	4,180	3.4

4.1.2. Variable creation and transformation

For the purpose of facilitating the development of the synthetic control model, which will be outlined in the following section, some variables were created, and some transformations were applied.

A variable *Region* was established to differentiate observations in the panel that pertain to countries within the same world region as Brazil - Latin America. The variable is assigned a value of 1 if the observation pertains to a country within this region and 0 if it does not. Similarly, a variable *EMBI* was created to classify observations from countries that are identified as emerging economies by the Emerging Markets Bond Index (EMBI). The variable is assigned a value of 1 for observations from countries deemed as emerging economies by the EMBI and 0 for observations from countries that are not.

Additionally, in accordance with the method followed by Born et al (2010), all variables were calculated as the percent deviation from their values in both the years 1990 and 2019 (represented as t'), except for exports growth which is in percent, as demonstrated below:

$$dev\_var_t\_t' = \frac{var_t - var_{t'}}{var_{t'}} * 100 \quad (8)$$

Finally, the desired outcome variable was GDP at constant prices (*gdpreal*). However, access to this data was not available. Instead, Real GDP per capita and population were utilized to calculate *gdpreal* through multiplication of the two variables.

4.1.3. Final dataset structure

This subsection presents an overview of the dataset used in this study by providing a summary of the variables and their characteristics in a table format. Table 4.2 below includes information on the variables' names, descriptions, types, and roles in the analysis. The main goal of this subsection is to present a clear and organized representation of the dataset structure and composition, to facilitate the understanding of how it will be used in the research.

Table 4.2 – Dataset structure

Variable	Variable Description	Variable Type	Variable Role
ISO	Encoded country's ISO code	Numeric - Integer	Panel id
country	Country's name	String	Panel id description
EMBI	Boolean to identify Emerging Markets	Boolean	Flag
Region	Boolean to identify Latin American countries	Boolean	Flag
year	year	Numeric - Integer	Time variable
gdpnom	Gross domestic product, current prices (Billions U.S. dollars)	Numeric - Float	Target variable (not in use)
gdppcnom	Gross domestic product per capita, current prices (U.S. dollars)	Numeric - Float	Target variable (not in use)
gdpreal	Gross domestic product, constant prices (2017 international dollar)	Numeric - Float	Target variable
gdppcrealppp	Gross domestic product per capita, constant prices (Purchasing power parity; 2017 international dollar)	Numeric - Float	Target variable (not in use)
exportsgrowth	Volume of exports of goods and services (Percent change)	Numeric - Float	Predictor variable (not in use)
exports	Volume of exports of goods and services (% of GDP)	Numeric - Float	Predictor variable (not in use)
consumption	Final consumption expenditure (% of GDP)	Numeric - Float	Predictor variable (not in use)
inv	Total investment (% of GDP)	Numeric - Float	Predictor variable (not in use)
sav	Gross national savings (% of GDP)	Numeric - Float	Predictor variable (not in use)
pop	Population (Millions of Persons)	Numeric - Float	Predictor variable (not in use)
dev_1990_gdpnom	% deviation from 1990 of *gdpnom*	Numeric - Float	Predictor variable
dev_1990_gdppcnom	% deviation from 1990 of *gdppcnom*	Numeric - Float	Predictor variable
dev_1990_gdpreal	% deviation from 1990 of *gdpreal*	Numeric - Float	Predictor variable
dev_1999_gdppcrealppp	% deviation from 1990 of *gdppcrealppp*	Numeric - Float	Predictor variable
dev_1990_exportsgrowth	% deviation from 1990 of *exportsgrowth*	Numeric - Float	Predictor variable
dev_1990_exports	% deviation from 1990 of *exports*	Numeric - Float	Predictor variable

dev_1990_consumption	% deviation from 1990 of *consumption*	Numeric - Float	Predictor variable
dev_1990_inv	% deviation from 1990 of *inv*	Numeric - Float	Predictor variable
dev_1990_sav	% deviation from 1990 of *sav*	Numeric - Float	Predictor variable
dev_1990_pop	% deviation from 1990 of *pop*	Numeric - Float	Predictor variable
dev_2019_gdpnom	% deviation from 2019 of *gdpnom*	Numeric - Float	Predictor variable
dev_2019_gdppcnom	% deviation from 2019 of *gdppcnom*	Numeric - Float	Predictor variable
dev_2019_gdpreal	% deviation from 2019 of *gdpreal*	Numeric - Float	Predictor variable
dev_2019_gdppcrealppp	% deviation from 2019 of *gdppcrealppp*	Numeric - Float	Predictor variable
dev_2019_exportsgrowth	% deviation from 2019 of *exportsgrowth*	Numeric - Float	Predictor variable
dev_2019_exports	% deviation from 2019 of *exports*	Numeric - Float	Predictor variable
dev_2019_consumption	% deviation from 2019 of *consumption*	Numeric - Float	Predictor variable
dev_2019_inv	% deviation from 2019 of *inv*	Numeric - Float	Predictor variable
dev_2019_sav	% deviation from 2019 of *sav*	Numeric - Float	Predictor variable
dev_2019_pop	% deviation from 2019 of *pop*	Numeric - Float	Predictor variable

4.2. SYNTHETIC CONTROL

This subsection aims to describe the application of the synthetic control method developed by Abadie et al. (2003) to estimate the impact of the policies and actions of Brazilian President Jair Bolsonaro on the country's real GDP.

The target variable for this study is real GDP, and the predictor variables used are consumption, investment, savings, exports, and population. The treatment unit is Brazil, and the treatment period is 2019. The pre-intervention period is 1990-2018, and the post-intervention period is 2020-2027.

The study was conducted through an iterative process, where various options were considered. However, in this section, only two results will be presented: one for an unrestricted pool (all 110 countries are present) and the restricted pool (20 countries with similarities to Brazil).

The parameters for this study are summarized in table 4.3 below.

Table 4.3 – SCM parameters

SCM parameters	**This study**

Target Variable (Y_{jt})	Real GDP
Predictor Variables (X_{kj})	Consumption, Investment, Savings, Exports, Population
Treatment Unit (j=1)	Brazil
Donor Pool ($J-1$)	Unrestricted sample[1] (110 countries), Restricted sample[2] (20 countries)
Treatment Period (T_o)	2019
Pre-intervention Period (T_p)	1990-2018
Post-intervention Period ($T-T_p$)	2020-2027

Each of the following subsections will explain the rationale behind the selection of these parameters.

4.2.1. Target variable

The target variable in this study is real GDP, which is the value of all goods and services produced within a country in a given period of time, adjusted for inflation.

Real GDP is a good indicator of the macroeconomic impact of the policies and actions of Brazilian President Jair Bolsonaro, in this case. It captures the overall level of economic activity in the country and reflects the level of consumption, investment, savings, exports, and population, all of which are important factors that can be affected by the policies and actions of the government. It is widely used, reported, and understood. Additionally, since it accounts for inflation, it allows for a fair comparison over time.

4.2.2. Predictor variables

The choice of predictor variables in this study is based on the average characteristics of the economy as suggested most literature on SCM and is adapted for data availability. These variables are chosen to capture the main economic trends of the treatment unit (Brazil) and the control units (other countries in the pool). This study follows the approach of Born et al. (2021), choosing predictors as described below.

The first predictor variable used in this study is consumption, which captures the overall spending of households, businesses, and government. Consumption is a key driver of economic growth and is one of the main components of GDP.

[1] Appendix B – Unrestricted Donor Pool
[2] Appendix C – Restricted Donor Pool

The second predictor variable used is investment, which captures the overall spending of businesses and government on fixed assets such as machinery, buildings, and infrastructure. Investment is another important driver of economic growth, as it leads to an increase in productivity and capacity.

The third predictor variable used is savings, which captures the overall level of savings of households, businesses and government. Savings are an important component of the economy as they provide funds for investment and also serve as a buffer against economic shocks.

The fourth predictor variable used is exports, which captures the overall level of exports of goods and services. Exports are important for the economy as they provide a source of foreign exchange and can lead to an increase in economic growth.

The fifth predictor variable used is population, which captures the overall size of the population. Population is an important economic variable as it is related to the size of the labor force, which is an important input into the production process.

4.2.3. Treatment period

The treatment period is an important aspect of the synthetic control method, as it determines the time period during which the intervention (in this case, the policies and actions of Brazilian President Jair Bolsonaro) took place. In this study, the treatment period is 2019.

Jair Bolsonaro was elected president of Brazil in 2018, but he only took office in 2019. Therefore, the treatment period is set to 2019, as this is the year when the policies and actions of the Bolsonaro government began to be implemented. It is assumed that the impact of the government's policies on the economy would be reflected in the GDP data for that year.

4.2.4. Pre-intervention period

The choice of the pre-intervention period is a crucial aspect of the synthetic control method as it establishes the time frame used to identify the trends and patterns of the treatment unit (Brazil) before the intervention (2019). In this study, the pre-intervention period chosen is 1990-2018.

According to Abadie et al, the pre-intervention period should be long enough to capture the underlying trends and patterns of the treatment unit and to provide a robust estimate of the counterfactual outcome. A longer pre-intervention period allows for a better identification of the trends and patterns in the treatment unit, which leads to a more accurate estimate of the counterfactual outcome.

The literature on synthetic control methods is not consistent in determining the adequate length of the pre-intervention period. Abadie, Diamond and Hainmueller (2010), a prominent source on synthetic control methodologies, uses 18 pre-intervention observations, which is a relatively low number. However, SCM are primarily applied in case studies, where even a small number of observations may be deemed adequate. In this study, the pre-intervention period of 1990-2018, spanning almost three decades, is considered long enough to capture the underlying trends and patterns of Brazil's GDP and the economic and political changes that have occurred in the country that can affect it.

4.2.5. Post-intervention period

The post-intervention period is the time period following the intervention (2019) and is used to compare the actual outcomes of the treatment unit (Brazil) to the counterfactual outcomes estimated by the synthetic control method. In this study, the post-intervention period is 2020-2027.

The choice of the post-intervention period is important as it determines the time period over which the actual outcomes of the treatment unit are compared to the counterfactual outcomes. A longer post-intervention period allows for a more comprehensive assessment of the impact of the intervention, but it also increases the uncertainty of the estimates due to the potential for unanticipated events.

Abadie et al. (2014) argue that a period of approximately 10 years following an event is a reasonable time frame to observe the effects. In this study, the post-intervention period of 2020-2027 is chosen to capture the potential impact of the intervention on Brazil's GDP, which can be considered a small period to draw conclusions – furthermore, it is important to note that values after 2021 are estimates made by IMF projections and should be interpreted with caution. It would make sense to repeat this study in the future when more data is available to have a more comprehensive analysis.

4.2.6. Donor pool – control group

In this subsection, the choice of the control group, also known as the donor pool, in the SCM used in the study will be described. The SCM uses a weighted average of a set of control units to estimate the counterfactual outcome of the treatment unit. The choice of the control group is an important aspect, as it affects the accuracy of the estimate of the counterfactual outcome.

Two models will be presented in this study: one with an unrestricted pool and another with a restricted pool. The unrestricted pool includes all countries available in the dataset after cleaning (110 countries) as potential control units, while the restricted pool includes only those countries that were selected, due to being similar to Brazil in terms of the predictor variables before the intervention. This approach is similar to the one followed by Billmeier et al. (2013), who also estimate the SCM twice, one using a non-restricted pool, and the other restricting the donor pool to countries from the same geographic region as the country under study.

The unrestricted pool model offers a broader perspective of the countries that could be used as control units, potentially capturing a wider range of patterns and trends in the data. There is literature that suggests that it may be acceptable not to restrict the donor pool in certain cases. SCM are a flexible technique and the choice of how to select the control group will depend on the specific research question, data availability and the characteristics of the treatment unit.

For instance, Abadie et al. (2010) argue that the synthetic control method can be applied to a wide range of cases and the restriction of the donor pool is not always necessary. They suggest that the synthetic control method can work well even when the control group is not perfectly matched to the treatment unit. This is because the method uses a weighted average of the control units to estimate the counterfactual outcome, which can help to mitigate the effects of any differences between the control units and the treatment unit. Moreover, in the particular case of this study, there is no need to exclude countries that have been affected by this treatment – no interference assumption, explained in section 3.1.2. – given that we are evaluating the impacts of Bolsonaro's administration on Brazilian GDP, and no other country has received the same treatment. As such, results will be presented for the case where the pool is not restricted in the following chapter 5.1.

However, it is important to note that most studies advocate for the restriction of the donor pool. The restriction of the donor pool is a common practice in SCM as it allows for a more accurate estimate of the counterfactual outcome by ensuring that the control units are as similar as possible to the treatment unit in terms of the predictor variables before the intervention. This is important because the method uses the average of the control units to estimate the counterfactual outcome of the treatment unit. If the control units are not similar to the treatment unit, then there's a violation of one of the assumptions – second assumption, described in section 3.1.2. – which can lead to interpolation biases, and overfitting. Nonetheless, it can be challenging to determine whether this assumption is satisfied, as it requires a balance between intuition and quantitative analysis during the selection of control units. There will always be a subjective element to this process, as one must make decisions about which control units to include.

Table 4.4 below contains all the countries that were included as part of the control group for the restricted analysis. These countries were selected because they are either from the same region as our treated unit – Latin America, or they were classified as an emerging market by the EMBI. There is a detailed individual description of similarities between these countries' economies and Brazil.

Table 4.4 – Restricted Donor Pool

ISO Code	Country	Similarities with Brazil
ARG	Argentina	Large Latin American countries with similar population sizes, GDP per capita, and GDP growth rates. Both countries have large agricultural and industrial sectors, and both have significant trade relations with China.
BOL	Bolivia	Latin American countries, with similar GDP per capita and GDP growth rates. Both countries have significant natural resources, including minerals and hydrocarbons.
CHL	Chile	Large Latin American countries with similar GDP per capita, GDP growth rates, and export-oriented economies. Both countries have significant trade relations with China.
COL	Colombia	Large Latin American countries with similar GDP per capita, GDP growth rates, and export-oriented

		economies. Both countries have significant trade relations with China.
CRI	Costa Rica	Latin American countries, with similar GDP per capita and GDP growth rates.
ECU	Ecuador	Latin American countries, with similar GDP per capita and GDP growth rates. Both countries have a relatively stable political and economic environment, and both have a relatively high level of human development.
EGY	Egypt	Emerging economies, large countries, with similar GDP per capita and GDP growth rates. Both countries have significant natural resources, including oil and gas, and both have large agricultural sectors.
HUN	Hungary	Emerging economies, countries with similar GDP per capita and GDP growth rates. Both have a relatively high level of human development.
IND	India	Emerging economies, large countries, with similar GDP per capita and GDP growth rates. Both countries have a significant agricultural sector and a large population.
JAM	Jamaica	Latin American countries, with similar GDP per capita and GDP growth rates. Both countries have a relatively stable political and economic environment, and both have a relatively high level of human development.
MAR	Morocco	Emerging economies, both countries with similar GDP per capita and GDP growth.
MEX	Mexico	Large Latin American countries with similar GDP per capita, GDP growth rates, and export-oriented economies. Both countries have significant trade relations with the United States.
PAN	Panama	both Latin American countries, with similar GDP per capita and GDP growth rates. Both countries have a relatively stable political and economic environment, and both have a relatively high level of human development.
PER	Peru	both Latin American countries, with similar GDP per capita and GDP growth rates. Both countries have a relatively stable political and economic environment, and both have a relatively high level of human development.
ROU	Romania	Emerging economies, both countries with similar GDP per capita and GDP growth rates. Both have a relatively high level of human development.
RUS	Russian Federation	Emerging economies, large countries, with similar GDP per capita and GDP growth rates. Both countries have

		significant natural resources, including oil and gas, and both have large agricultural sectors.
THA	Thailand	Emerging economies, countries with similar GDP per capita and GDP growth rates. Both have a relatively high level of human development.
TUR	Turkey	Emerging economies, countries with similar GDP per capita and GDP growth rates. Both have a relatively high level of human development.
ZAF	South Africa	Emerging economies, countries with similar GDP per capita and GDP growth rates, and both have a relatively high level of human development.

5. RESULTS AND DISCUSSION

This chapter presents the findings of the empirical application of the synthetic control method (SCM) to estimate the macroeconomic impact of Bolsonaro on Brazil's real GDP.

The chapter is divided into two sections. The first section presents the results of the SCM using an unrestricted pool of potential control units, including all the 110 countries available in the dataset. The second section presents the results of the SCM using a restricted pool of potential control units, which includes only those countries that are similar to Brazil in terms of the predictor variables before the intervention.

5.1. UNRESTRICTED DONOR POOL ANALYSIS

5.1.1. Results

The results presented in this section are the outcome of an iterative process, where different models were tested – changing the predictors and respective years - in order to identify the one that best mimics the evolution of the target variable - Real GDP. The model that was chosen as the one that presented the smallest RMSPE. It includes consumption, investment, savings and population as predictor variables. All of these variables are presented as percent deviations from their values in 1990.

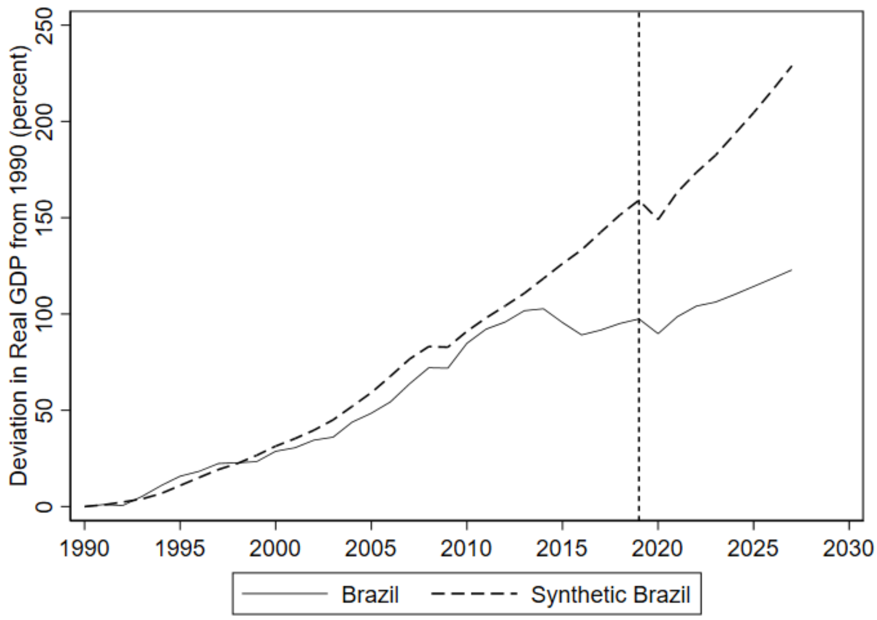

Figure 5.1 – Unrestricted: Brazil vs Synthetic Brazil

Figure 5.1 displays the time series for real GDP (solid line) and in the doppelganger/synthetic economy (dashed line). Table 5.1 shows how the synthetic control performs in terms of the targeted covariates, which are the variables used to minimize the distance between the two series – they determine how much weight each indicator variable receives in the minimization problem. In this case, even though these values may resemble regression coefficients, they are formulated in a different manner, and cannot be interpreted as either partial correlations or marginal effects. The values are in basis points.

Table 5.1 – Balance table

Predictor Variables	Treated	Synthetic	Differences
consumption	180.13	182.74	2.61
investment	1.75	1.75	0.00
savings	-13.54	-12.65	0.89
population	23.23	23.36	0.13

The contribution of the individual countries in the synthetic control ranges from 0% to 100%, and adds up to 100%. In this particular case, since the synthetic control is constructed using an unrestricted pool of donor countries, the weight of each individual country is very residual. Individual countries and respective contribution can be found in appendix B.

The chosen model was able to replicate the behavior of the target variable relatively well up until 2014, however, from that point onwards, a gap between the observed values and the synthetic values started to increase. Figure 5.2 bellows allows this visualization clearly.

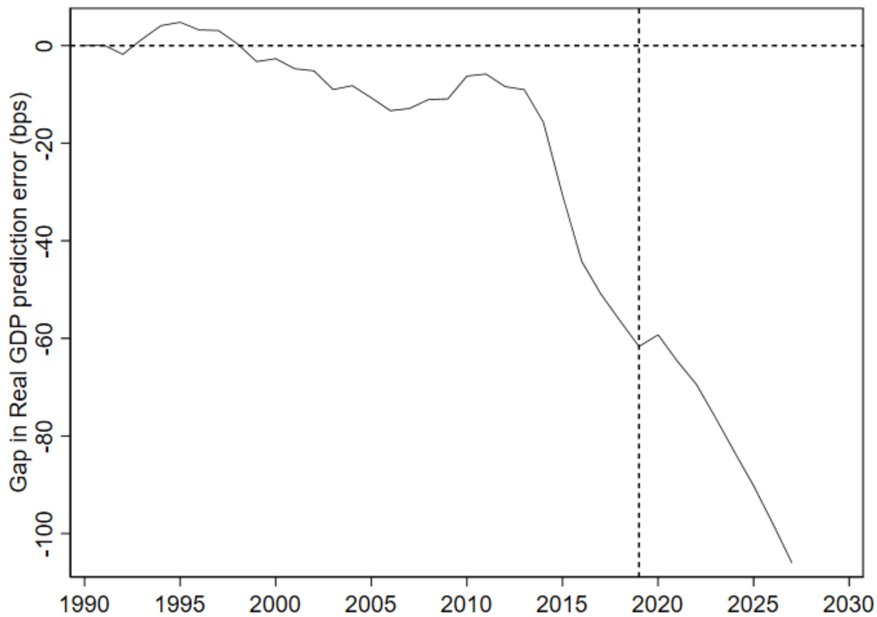

Figure 5.2 –Unrestricted sample: Gap in Real GDP prediction error

Despite several attempts to control for this discrepancy, it was not possible to fully replicate the economic dynamics that occurred in Brazil from 2014 onwards with this respective set up.

5.1.2. Goodness of fit and Causality

As previously discussed, the algorithm finds a synthetic control for Brazilian real GDP as a linear combination of countries in the donor pool, so that the synthetic control matched the values of the variable that present the highest predictive power and that RMSPE is minimized. As presented in table 5.3, the differences in the pretreatment average of the predictors between the treated unit and its synthetic counterfactual are relatively small. Moreover, the RMSPE is 0.182 and the fit index, which is calculated following the approach formally described in section 3.1.4, is 0.09. Thus, both the predictors and the RMSPE of the outcome variable show a relatively satisfactory pretreatment fit between Brazil and synthetic Brazil, despite the fact that for the second part of the 2010 decade the gap between both starts to widen.

After applying the methodology, statistical inference placebo tests were performed, as recommended by Abadie, Diamond, and Hainmueller (2010), and formally described in section 3.1.4. The process involved iteratively applying the synthetic control method to each country in the donor pool to obtain a distribution of placebo effects. This was done by calculating the RMSPE for each placebo for the pre-treatment period, the post-treatment period, and the ratio of the post- to pre-treatment RMSPE. The ratios were then sorted in descending order and the treatment unit's ratio was calculated as p=rank/total. This method was used to determine whether the treatment effect for Brazil was extreme compared to the donor pool's own placebo ratios.

As such, despite the limitations of the model's ability to accurately predict post-2014 GDP, the results of the placebo tests indicate a statistically significant difference between the synthetic and actual GDP with a rank of 6 out of 110 countries, and a p-value of 0.046, which is considered statistically significant – confronted with the arbitrary 5% regularly used as reference.

The next step is to perform a similar analysis, but for the sample with a restricted set of control units.

5.2. Restricted Donor Pool Analysis

5.2.1. Results

Similarly to the steps followed to implement the method in the unrestricted donor pool sample, this result is also a outcome of an iterative process, in which various models were tested to identify the one that best replicates the evolution of the target variable. The predictors are consumption, investment, savings, population, and in this model, we also added exports. The target variable is real GDP, and all variables are displayed as percentage deviation from their respective values in 1990.

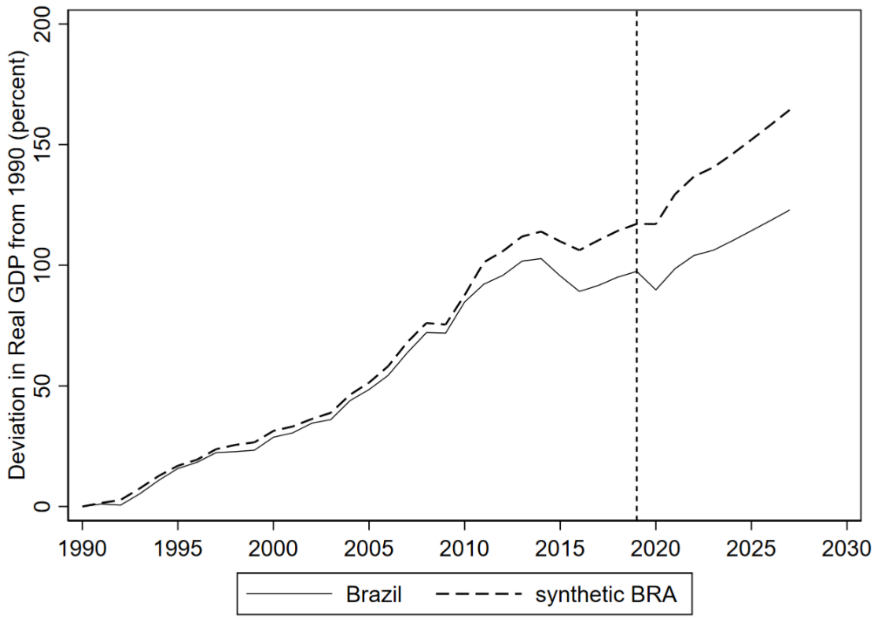

Figure 5.3 - Restricted sample: Brazil vs Synthetic Brazil

Figure 5.3 above depicts the target variable, real GDP (solid line) and its respective synthetic doppelganger (dashed line). Table 5.2 shows how the synthetic control performs in terms of the targeted covariates.

Table 5.2 – Restricted: Predictor Balance

Predictor Variable	**Treated** (bps)	**Synthetic** (bps)	**Differences** (bps)
consumption	180.13	179.35	-0.78
investment	1.75	1.75	0.00
savings	-13.54	-13.59	-0.05
population	23.23	23.18	-0.05
exports	287.4	287.2	-0.2

Table 5.3 below shows the contribution of the individual countries in the synthetic control. These figures are distributed as follows, and it means that the weights reported in the table indicate that

the trend of real GDP in Brazil prior to the election of Bolsonaro – i.e. prior to 2019 – is best represented by a combination of the following countries.

Table 5.3 – Weight distribution per control unit

Country	**Weight** (bps)
Argentina	2.1
Bolivia	1.0
Chile	2.0
Colombia	2.5
Costa Rica	**6.0**
Ecuador	1.3
Egypt	0.1
Hungary	1.8
India	1.7
Jamaica	**17.1**
Morocco	1.8
Mexico	1.8
Panama	1.5
Perú	2.3
Romania	1.8
Russia	1.4
Thailand	2.6
Turkey	1.8
South Africa	**49.2**

Figure 5.4 below shows that the results from this model follows a similar pattern to the one applied using the unrestricted sample - the model was able to replicate the behavior of the target variable relatively well up until 2014, however, from that point onwards, a gap between the observed values and the synthetic values started to increase – however, with a smaller magnitude.

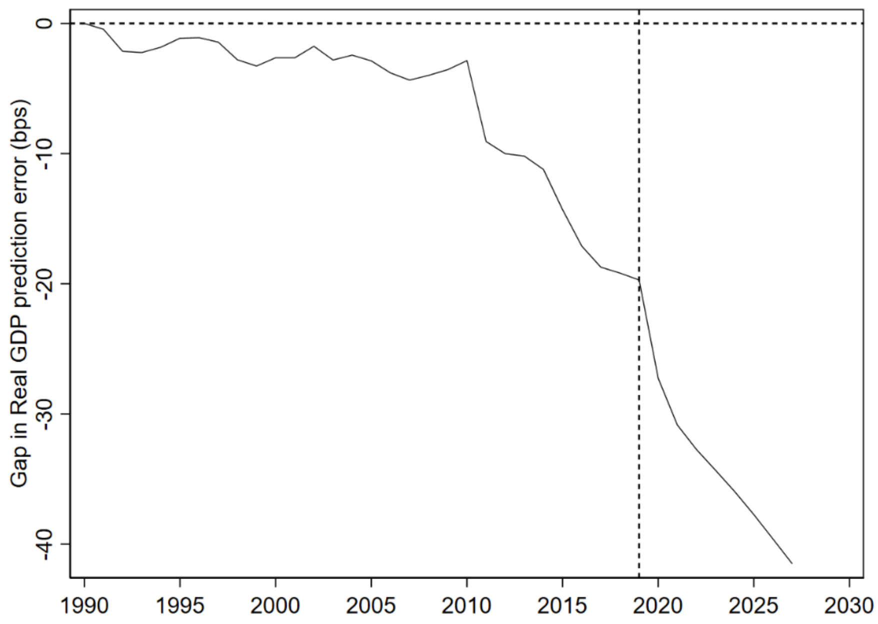

Figure 5.4 –Restricted sample: Gap in Real GDP prediction error

5.2.2. Goodness of fit and Causality

As previously explained, the algorithm seeks to identify a synthetic control for Brazilian real GDP by combining other countries in the donor pool in a linear fashion, ensuring the synthetic control matches the variable with the greatest predictive power and minimizing RMSPE. Table 5.3 shows that the differences in predictor variables between the treated and synthetic counterfactual units were relatively small before the treatment. Additionally, the RMSPE value was 0.033, and the fit index, calculated using the method described in section 3.1.4, was 0.01. This suggests that the synthetic control model exhibited an overall strong pretreatment fit for both the outcome variable and predictor variables.

Following the literature, statistical inference placebo tests were conducted – also similarly to the diagnostic performed to the model estimated using the unrestricted sample. The process involved applying the SCM iteratively to each country in the donor pool to generate a distribution of placebo effects. This involved calculating the RMSPE for each placebo during the pre-treatment and post-treatment periods, as well as the post-treatment to pre-treatment RMSPE ratio.

Results of the placebo tests demonstrate a statistically significant distinction between the actual and synthetic GDP for Brazil, with a rank of 2 out of 20 countries and a p-value of 0.018, which is regarded as statistically significant using the commonly used reference value of 5%.

Figure 5.5 illustrates the distribution of the placebo effects for each donor country.

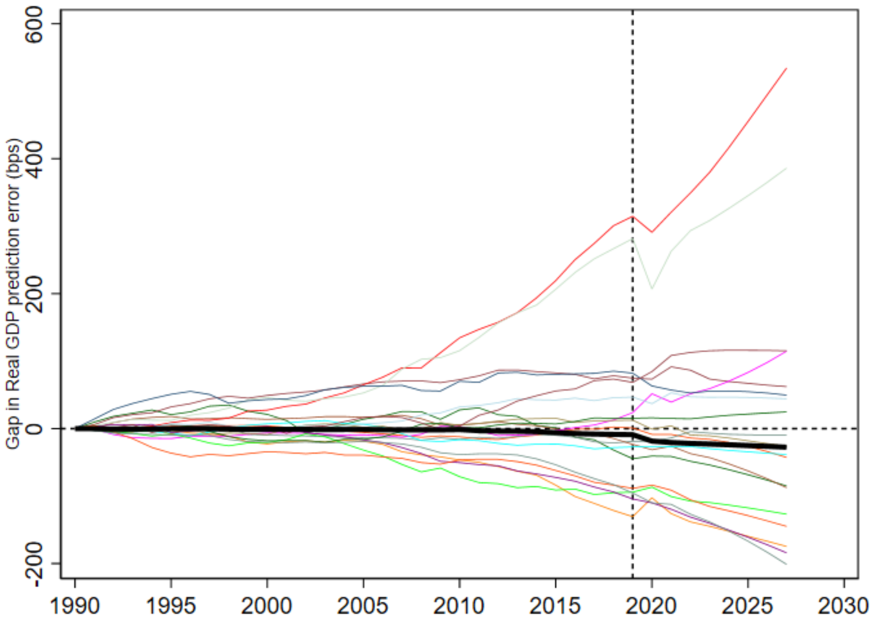

Figure 5.5 – Placebo Distribution. Brazil is the dark line.

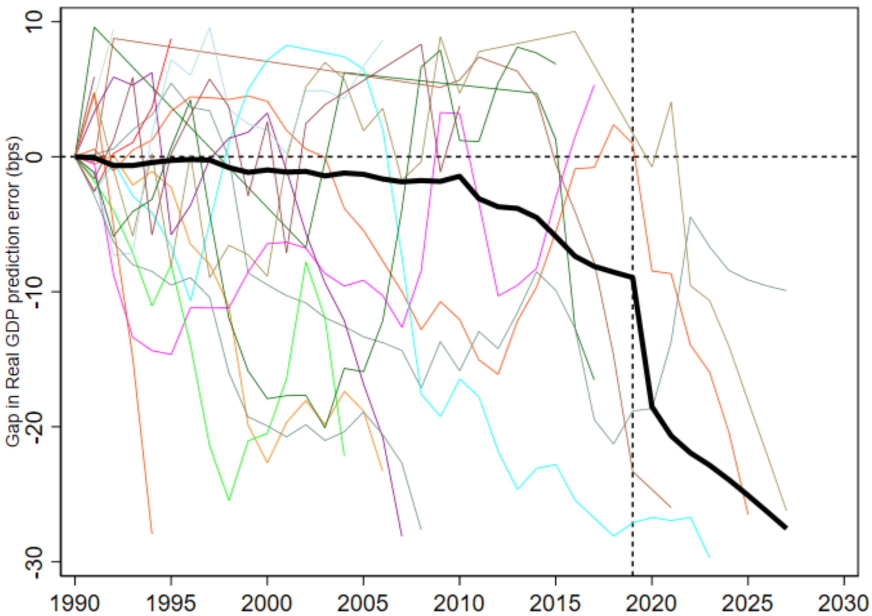

Figure 5.6 – Placebo Distribution. Brazil is the dark line – zoomed in.

A discussion confronting the results from both models follows.

5.3. Discussion

Focusing on the unrestricted donor pool analysis, the limitations of the model's ability to accurately predict post-2014 GDP may be explained by several factors – intuitively, there can be a conjunctural explanation. Indeed, it is important to note that the Brazilian economy experienced a significant growth during the commodities boom in 2010, which allowed for a high GDP growth rate. However, from 2014 onwards, the end of the commodities boom, coupled with political instability and the impact of the Lava Jato investigation, led to a substantial decrease in the GDP growth rate that couldn't be replicated by the synthetic control. This suggests that there may have been unique factors at play in the Brazilian economy during this time that were not present in the donor pool and therefore could not be fully captured by the model.

On the other hand, the underlying reason for this limitation can also reside on the technical part of the analysis – the implementation of the methodology. Indeed, the algorithm creates the synthetic control as a linear combination of countries in the donor pool, assigning weights to each of the predictors, by using the averages of all predictor variables over the entire preintervention period from all the countries in the panel. This can be suboptimal, since we are not restricting the sample. For this reason, the following step was to repeat the study, but using a restricted donor pool of countries with characteristics that somehow resemble Brazil – the followed rationale is described in section 4.2.6.

Looking at the latter analysis, indeed it can be observed that the results slightly improve. However, there is a similar pattern. This is evident by looking at the below figures 5.6 and 5.7.

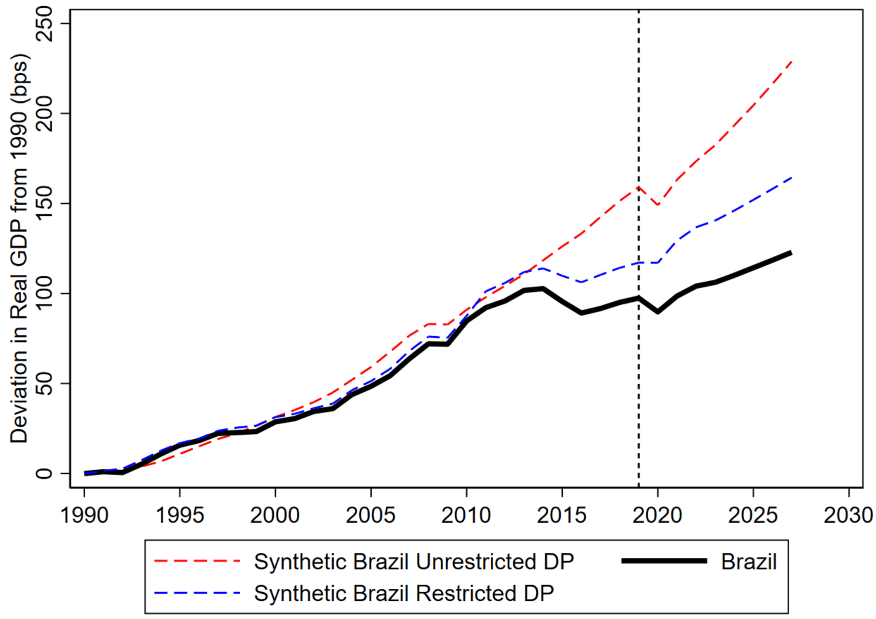

Figure 5.7 – Comparison between synthetic controls estimated using an unrestricted DP vs a restricted DP

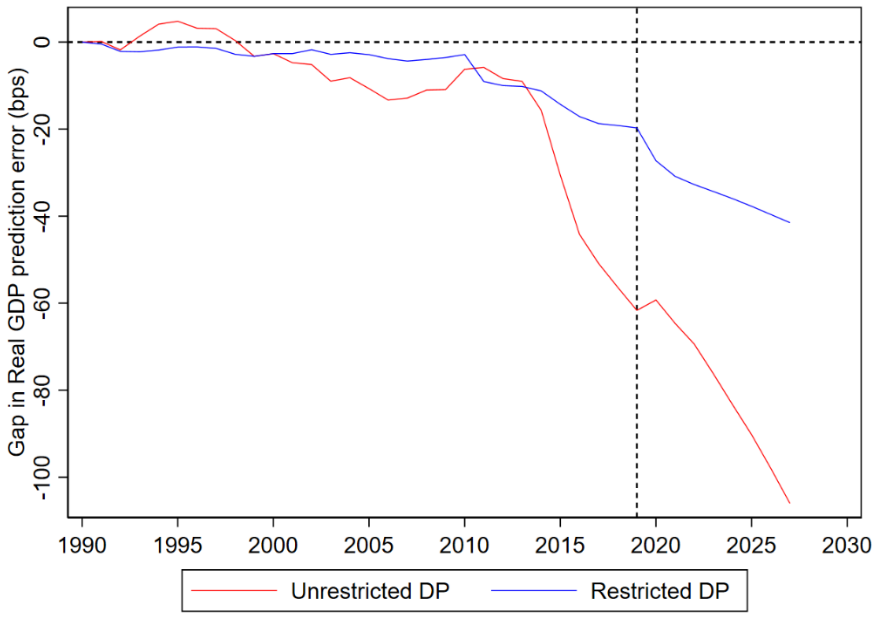

Figure 5.8 –Gap comparison between synthetic controls estimated using an unrestricted DP vs a restricted DP

As such, despite the fact that the results present statistical significance, they must be interpreted with caution. It is unclear if it can provide a concrete answer to our research question of how Brazil would have performed without the policies of President Bolsonaro. The gap between the real and synthetic GDP growth rates increases from 2014 onwards, possibly due to unique factors such as the end of the commodities boom, political instability, and the impact of the Lava Jato investigation, which were naturally not present in the donor pool and therefore not fully captured by the model.

For the unrestricted donor pool analysis, it is visible that the Brazilian economy presents a slower recovery after the pandemic period, compared to the synthetic one. For the restricted analysis, we observe that the counterfactual trend is more similar to the outcome variable, yet it is slightly steeper, and it’s evidenced a smaller drop during the pandemic.

These trends could potentially be related to how President Bolsonaro's administration handled the COVID-19 pandemic in Brazil. Indeed, his administration was severely criticised. However, while these factors may have played a role, it is hard to attribute it solely to Bolsonaro's administration and further research would be needed to draw more concrete conclusions.

6. CONCLUSIONS AND FUTURE WORKS

In conclusion, this study aimed to estimate the macroeconomic impact of President Jair Bolsonaro's policies on the Brazilian economy using the SCM. The results of the analysis using both restricted and unrestricted donor pools showed that while the RMSPE was better for the restricted sample, the model was not able to fully replicate the economic dynamics that occurred in Brazil from 2014 onwards. Despite the limitations of the model, the results of the placebo tests indicate a statistically significant difference between the synthetic and actual GDP. It is important to note that these results must be interpreted with caution as there may have been unique factors at play in the Brazilian economy during this time that were not present in the donor pool and therefore not fully captured by the model. Additionally, it is visible that the Brazilian economy presents a slower recovery after the pandemic period compared to the synthetic one, which could potentially be related to President Bolsonaro's administration's handling of the COVID-19 pandemic. However, further research would be needed to draw more concrete conclusions. The study suggests that it's hard to attribute the gap in real GDP between the synthetic and the real Brazil to Bolsonaro's policies. The study highlights the importance of further research to better understand the economic dynamics of Brazil and the impact of President Bolsonaro's policies.

Additionally, the findings of this research study indicate that further research is needed in order to fully understand the impact of President Bolsonaro's administration on the Brazilian economy. The synthetic control method used in this study is a powerful tool for evaluating the impact of a specific intervention, but it is not without its limitations.

One limitation is that the analysis is highly aggregated, making it difficult to control for all of the factors that may be at play. A more detailed and less aggregated analysis, such as a study of the impact of specific policies on sectoral GDP, would be more appropriate for evaluating the impact of President Bolsonaro's administration. Additionally, identifying a covariate that can control for what the model is not capturing would be beneficial.

Another limitation of this study is that it relies on predictions for the post-intervention period, rather than actual values. This makes it difficult to draw solid conclusions about the impact of President Bolsonaro's administration. A future repetition of the study using actual values would be more informative. Furthermore, repeating the study with less aggregated data, such as quarterly data, would also be beneficial.

Finally, it is important to note that studying the impact of President Bolsonaro's administration on the Brazilian economy from a different angle could be beneficial. For example, studying the heterogeneity of Brazil's states, and evaluating how states in which Bolsonaro supporters were more predominant performed, could provide more insight into the impact of his policies. In this case, a classic difference-in-differences approach may be more appropriate than the synthetic control method used in this study.

Ultimately, this study provides a preliminary estimate of the impact of President Jair Bolsonaro's policies on the Brazilian economy. However, further research is needed to draw more solid conclusions and address the limitations of the study.

BIBLIOGRAPHICAL REFERENCES

Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. The American Economic Review, 93(1), 113-132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association, 105(490).

Abadie, A., Diamond, A., & Hainmueller, J. (2011). Synth: An R Package for Synthetic Control Methods in Comparative Case Studies. Journal of Statistical Software, 42(13).

Abadie, A., Diamond, A., & Hainmueller, J. (2014). Comparative Politics and the Synthetic Control Method. American Journal of Political Science.

Absher, S., Grier, K., & Grier, R. (2020). The economic consequences of durable left-populist regimes in Latin America. Journal of Economic Behavior & Organization.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., & Mitton, T. (2013). The Value of Connections in Turbulent Times: Evidence from the United Stated. National Bureau of Economic Research.

Adhikari, B. (2021). A Guide to Using the Synthetic Control Method to Quantify the Effects of Shocks, Policies, and Shocking Policies.

Adhikari, B., & Alm, J. (2016). Evaluating the economic effects of flat tax reforms using synthetic control methods. Southern Economic Journal, 83(2), 437–463.

Alesina, A., & Ardagna, S. (2010). Large changes in fiscal policy: Taxes versus spending. NBER Macroeconomics Annual, 24, 1-68.

Alesina, A., & Spolaore, E. (2006). The size of countries: Does it matter? Journal of the European Economic Association, 4(2-3), 427-436.

Billmeier, A., & Nannicini, T. (2013). Assessing Economic Liberalization Episodes: A Synthetic Control Approach. The Review of Economic and Statistics, 95(3), 983-1001.

Born, B., Müller, G. J., Sedláček, P., & Schularick, M. (2021). The macroeconomic impact of Trump. Policy Studies.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2010). Catastrophic Natural Disasters and Economic Growth. Research Department Publications - Inter-Amercian Development Bank.

Cunningham, S. (2021). Causal Inference. Yale University Press.

Damásio, B., & Mendonça, S. (2019). Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition. Applied Economics Letters, 26(10), 843-849.

Damásio, B., & Mendonça, S. (2022). Leader-follower dynamics in real historical time: a Markovian test of non-linear causality between sail and steam (co-) development. Applied Economics, 1-11.

Damásio, B., & Nicolau, J. (2014). Combining a regression model with a multivariate Markov chain in a forecasting problem. Statistics & Probability Letters, 90, 108-113.

Damásio, B., & Nicolau, J. (2020). Time inhomogeneous multivariate Markov chains: detecting and testing multiple structural breaks occurring at unknown dates.

Damásio, B., Louçã, F., & Nicolau, J. (2018). The changing economic regimes and expected time to recover of the peripheral countries under the euro: A nonparametric approach. Physica A: Statistical Mechanics and its Applications, 507, 524-533.

Ferguson, N. (2006). The Ascent of Money: A Financial History of the World. Penguin Press.

Fisher, R. A. 1935. The Design of Experiments. Edinburgh: Oliver; Boyd

Grier, K., & Maynard, N. (2016). The economic consequences of Hugo Chavez: A synthetic control analysis. Journal of Economic Behavior & Organization.

Jones, B., Olken, B., Aug. 2005. Do leaders matter? National leadership and growth since World

War II. The Quarterly Journal of Economics, 120 (3), 835–864.

Lieberman, E. (2005). Nested Analysis as a Mixed-Method Strategy for Comparative Research. American Political Science Review, 99, 436-452.

Martins, D., & Damásio, B. (2020). One Troika fits all? Job crash, pro-market structural reform and austerity-driven therapy in Portugal. Empirica, 47(3), 495-521.

Rodrik, D. (2015). Economics Rules: The Rights and Wrongs of the Dismal Science. WW Norton & Company.

Sekhon, J. (2004). Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals. Perspectives on Politics, 2(2), 281-293.

Tabellini, G. (2008). The Economic Effects of Constitutions. MIT Press., 282-284.

Tarrow, S. (1995). Bridging the Quantitative-Qualitative Divide in Political Science. American Political Science Review, 89(2), 471-474.

Wang-Sheng, L. (2011). Comparative case studies of the effects of inflation targeting in emerging economies. Oxford Economic Papers.

APPENDIX A – SUMMARY STATISTICS

Variable	Variable Description	Mean	Std. dev.	Min	Max	n	N	MV	Source
gdpnom	Gross domestic product, current prices (Billions U.S. dollars)	303	1474	0	30282	192	8313	711	IMF
gdppcnom	Gross domestic product per capita, current prices (U.S. dollars)	10811	17105	17	151082	192	8313	711	IMF
gdppcrealppp	Gross domestic product per capita, constant prices (Purchasing power parity; 2017 international dollar)	17973	20422	353	170219	191	8240	784	IMF
inv	Total investment (% of GDP)	24	10	-11	116	170	7223	1801	IMF
sav	Gross national savings (% of GDP)	20	12	-236	121	169	7252	1772	IMF
exportsgrowth	Volume of exports of goods and services (Percent change)	6	20	-91	649	175	7285	1739	IMF
unemployment	Unemployment rate (Percent of total labor force)	9	6	0	70	112	4361	4663	IMF
pop	Population (Millions of Persons)	35	132	0	1469	192	8294	730	IMF
consumption	Final consumption expenditure (% of GDP)	81	18	12	237	179	6217	2807	WB
exports	Exports of goods and services (% of GDP)	39	29	0	229	181	6402	2622	WB

APPENDIX B – UNRESTRICTED DONOR POOL

ISO Code	Country	Synthetic Weight (bps)
ALB	Albania	2.70
ARG	Argentina	0.80
ARM	Armenia	0.30
AUS	Australia	0.90
AUT	Austria	1.10
AZE	Azerbaijan	0.60
BDI	Burundi	1.00
BEL	Belgium	1.20
BEN	Benin	0.70
BGD	Bangladesh	0.70
BGR	Bulgaria	1.00
BHR	Bahrain	0.60
BHS	Bahamas	0.90
BLR	Belarus	1.40
BLZ	Belize	0.70
BOL	Bolivia	0.70
BRA	Brazil	1.20
BRB	Barbados	0.80
BTN	Bhutan	0.80
BWA	Botswana	1.00
CAF	Central African Republic	0.90
CAN	Canada	1.20
CHE	Switzerland	0.80
CHL	Chile	2.00
CIV	Côte d'Ivoire	0.80
CMR	Cameroon	0.60
COG	Congo	0.80
COL	Colombia	0.80
COM	Comoros	0.80
CRI	Costa Rica	0.90
CYP	Cyprus	1.20
CZE	Czechia	1.50
DEU	Germany	1.10
DNK	Denmark	0.70
DOM	Dominican Republic	0.80

DZA	Algeria	0.70
ECU	Ecuador	0.90
EGY	Egypt	1.10
ESP	Spain	1.60
FIN	Finland	1.20
FRA	France	0.70
GAB	Gabon	1.40
GBR	United Kingdom	0.70
GEO	Georgia	0.80
GHA	Ghana	0.80
GIN	Guinea	1.10
GNB	Guinea-Bissau	1.40
GRC	Greece	0.70
GTM	Guatemala	0.90
HKG	Hong Kong	0.50
HND	Honduras	0.70
HTI	Haiti	1.40
HUN	Hungary	0.90
IDN	Indonesia	0.80
IND	India	0.80
IRL	Ireland	0.90
IRN	Iran	0.90
ISL	Iceland	1.50
ITA	Italy	0.90
JAM	Jamaica	0.70
JOR	Jordan	1.70
JPN	Japan	0.80
KEN	Kenya	0.70
KGZ	Kyrgyzstan	1.00
KOR	Korea, Republic of	0.90
LUX	Luxembourg	0.90
MAR	Morocco	0.80
MDG	Madagascar	0.80
MEX	Mexico	0.80
MLI	Mali	1.20
MLT	Malta	0.60
MNG	Mongolia	0.80
MOZ	Mozambique	0.70
MRT	Mauritania	1.00

MUS	Mauritius	0.80
MYS	Malaysia	0.80
NAM	Namibia	0.60
NER	Niger	0.60
NGA	Nigeria	1.10
NLD	Netherlands	1.00
NOR	Norway	0.90
NZL	New Zealand	0.60
OMN	Oman	0.80
PAK	Pakistan	0.60
PAN	Panama	0.70
PER	Peru	0.80
PHL	Philippines	1.50
PRT	Portugal	0.80
PRY	Paraguay	1.90
ROU	Romania	0.90
RUS	Russian Federation	0.70
RWA	Rwanda	0.60
SAU	Saudi Arabia	0.50
SEN	Senegal	0.70
SGP	Singapore	0.70
SLB	Solomon Islands	0.70
SLE	Sierra Leone	0.80
SLV	El Salvador	0.90
SVK	Slovakia	1.30
SWE	Sweden	1.10
SWZ	Eswatini	0.40
TCD	Chad	0.80
TGO	Togo	1.10
THA	Thailand	0.90
TUN	Tunisia	0.90
TUR	Turkey	0.70
TZA	Tanzania	0.60
UGA	Uganda	0.90
USA	United States of America	0.90
ZAF	South Africa	0.90