

# Economic Impacts of the Brumadinho Dam Rupture

João Marrey Mendonça

**Final Paper Advisor:** Prof. Solange Ledi Gonçalves

**Final Paper Coordinator:** Prof. Rafael Pucci

Final Course Paper presented to the Faculty of Economics, Administration,  
Accounting and Actuarial Science  
of the University of São Paulo, in partial fulfillment of the requirements for the  
Bachelor's Degree in Economics.

São Paulo, 2025

# Contents

<b>Abstract</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Context . . . . .	2
1.2 Motivation . . . . .	4
<b>2 Methodology</b>	<b>5</b>
2.1 Overview . . . . .	5
2.2 Database and Variables . . . . .	5
2.3 Matching . . . . .	8
2.4 Model . . . . .	9
2.5 Inference . . . . .	12
<b>3 Results</b>	<b>13</b>
3.1 Overview . . . . .	13
3.2 GDP Per Capita . . . . .	14
3.2.1 Overview . . . . .	14
3.2.2 Effects on <i>GDPPerCapita</i> . . . . .	14
3.2.3 Total <i>GDP</i> transformation . . . . .	16
3.3 Wage . . . . .	17
3.4 Log Wage . . . . .	19
3.5 Prop Employed . . . . .	21
<b>4 Heterogeneity</b>	<b>23</b>
4.1 GVA Per Capita . . . . .	23
<b>5 Downstream Robustness</b>	<b>27</b>
<b>6 Conclusion</b>	<b>31</b>
<b>References</b>	<b>34</b>
<b>Appendix</b>	<b>35</b>
<b>A Mean of Predictors for Brumadinho and Synthetic Control</b>	<b>35</b>
<b>B Synthetic Control weights</b>	<b>41</b>

## List of Figures

1	River and affected municipalities map (MG state). . . . .	2
2	Tailing dams structures. . . . .	3
3	GDP Per Capita treated and synthetic trends. . . . .	15
4	GDP Per Capita gaps. . . . .	15
5	GDP Per Capita placebo gaps. . . . .	16
6	GDP Per Capita placebo distribution. . . . .	16
7	Wage treated and synthetic trends. . . . .	17
8	Wage gaps. . . . .	17
9	Wage placebo gaps. . . . .	18
10	Wage placebo distribution. . . . .	18
11	Log Wage treated and synthetic trends. . . . .	19
12	Log Wage gaps. . . . .	20
13	Log Wage placebo gaps. . . . .	20
14	Log Wage placebo distribution. . . . .	21
15	Proportion of Employment treated and synthetic trends. . . . .	22
16	Proportion of Employment gaps. . . . .	22
17	Proportion of Employment placebo gaps. . . . .	23
18	Proportion of Employment placebo distribution. . . . .	23
19	Sectorial GVA Per Capita treated and synthetic trends. . . . .	24
19	Sectorial GVA Per Capita treated and synthetic trends. (Continued) . .	24
20	Sectorial GVA Per Capita gaps. . . . .	25
21	Sectorial GVA Per Capita placebo gaps. . . . .	26
22	Sectorial GVA Per Capita placebo distributions. . . . .	26
23	GDP Per Capita matched donor pool maps. . . . .	27
24	GDP Per Capita trends and gaps no downstream. . . . .	28
25	Wage matched donor pool maps. . . . .	28
26	Wage trends and gaps no downstream. . . . .	29
27	Log Wage matched donor pool maps. . . . .	29
28	Log Wage trends and gaps no downstream. . . . .	29
29	Proportion of Employment matched donor pool maps. . . . .	30
30	Proportion of Employment trends and gaps no downstream. . . . .	30

## List of Tables

1	Original variables, their time periods and respective data sources. . . . .	7
2	Compound variables, their time periods and formulas. . . . .	8
A.1	GDP Per Capita estimation mean of predictors. . . . .	35
A.2	Wage estimation mean of predictors (Part 1). . . . .	36
A.3	Wage estimation mean of predictors (Part 2). . . . .	37
A.4	Log Wage estimation mean of predictors (Part 1). . . . .	38
A.5	Log Wage estimation mean of predictors (Part 2). . . . .	39
A.6	Proportion of Employment estimation mean of predictors. . . . .	40
B.1	GDP Per Capita weights (Only municipalities with at least 0.1% weight)	41
B.2	Wage weights (Only municipalities with at least 0.1% weight) . . . . .	42
B.3	Log Wage weights (Only municipalities with at least 0.1% weight) . . . . .	42
B.4	Proportion of Employment weights (Only municipalities with at least 0.1% weight) . . . . .	43

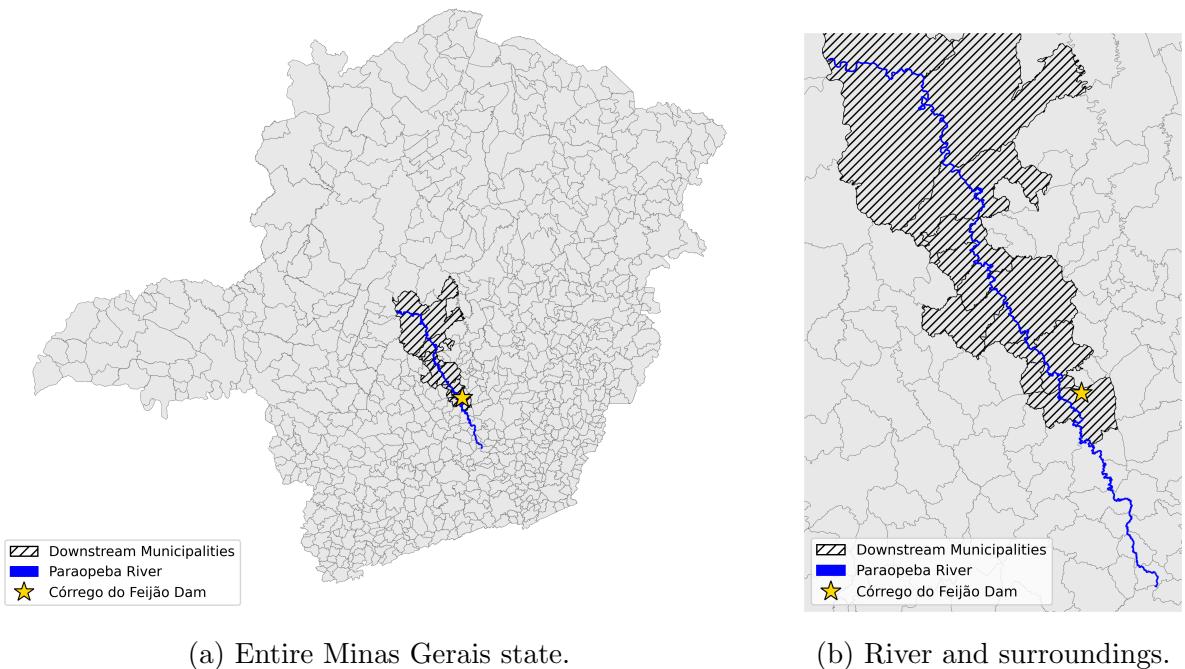
## **Abstract**

In January 2019, a tailings dam from the mining company Vale ruptured in the city of Brumadinho, Brazil. It was a major disaster, with the dam's mud causing the death of 272 people, besides a series of environmental, social and economic impacts. This paper uses a matching algorithm along with the synthetic control framework to evaluate the economic repercussions of this incident on the city of Brumadinho. It focuses on the effect on GDP per capita, wages and employment, also analyzing heterogeneity to determine which economic sectors were most affected among Agriculture, Industry, Services and Public Administration by using the framework on their sectoral GVA per capita. The results show a clear negative impact on GDP per capita, which was initially small, but increased over time to reach more than R\$100,000.00 in estimated effects. The heterogeneity analysis revealed that the affected sectors were Industry and Services, with negative impacts throughout most treatment periods, except for 2019, when Services presented positive and Industry small negative effects. This could possibly be explained by the increase in activity generated by the need to rebuild and remediate the impacts after the incident. The results did not provide enough evidence to identify any effects of the accident on sectoral GVA per capita for Agriculture or Public Administration. For wages and employment, the model also does not show evidence of impact.

# 1 Introduction

## 1.1 Context

On the 25th of January 2019, one of the largest socio-environmental catastrophes in Brazil's history happened, the collapse of the dam at Córrego do Feijão's iron mine, in the city of Brumadinho, Minas Gerais state (MG). Figure 1 shows the course of the river, the dam location and the downstream municipalities. The accident led to 272 deaths, a number approximately 18 times greater than that of the Mariana disaster in 2015, a similar dam rupture incident. More than 11.7 million  $m^3$  of toxic mud raged through the region and contaminated the Paraopeba River.



(a) Entire Minas Gerais state. (b) River and surroundings.

Figure 1: River and affected municipalities map (MG state).

The Brumadinho dam was operated by the Brazilian mining company Vale S.A. since 1976. It was built as an upstream tailings dam, a frailler type of structure compared to centreline or downstream tailings dam designs (Figure 2).

The last *in loco* inspection of the dam by the ANM (Brazilian National Mining Agency) had been done in 2016, with several reports stating that the agency was underfunded and did not have the resources necessary to keep track of all the mines under its responsibility. Because of this, many companies had the responsibility of inspecting their own activities, delegating this task to the auditor of their choice<sup>1</sup>.

The main contracted auditor was TÜV Süd, which was responsible for the official inspections and the Dam Condition Stability Declarations document. In 2018, months

1. See Senra (2019) for further details (text in portuguese).

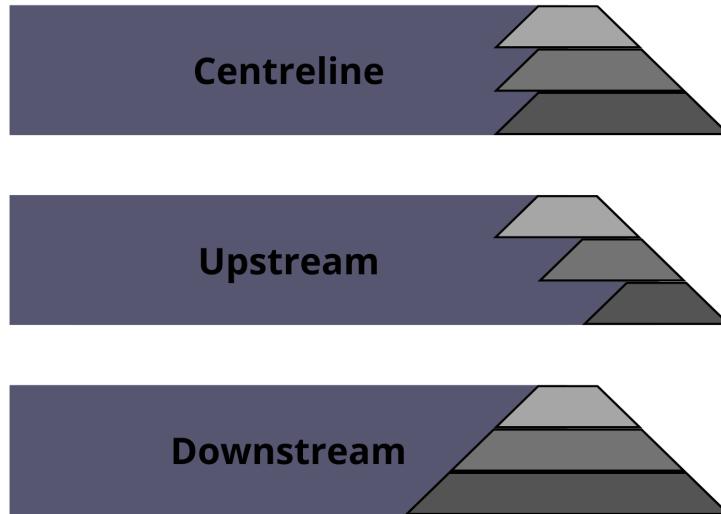


Figure 2: Tailing dams structures.

before the accident, Tractebel was hired to do a separate review of the dam and identified serious structural risks, refusing to attest to the dam's stability. After that, Vale exerted pressure on TÜV Süd to provide the required stability declaration, with the agency ultimately conceding and providing the document despite internal warnings from some of its engineers who recognized the risk<sup>2</sup>.

As a result of the accident, in 2021 Vale closed a deal with the government of the state of Minas Gerais, agreeing to pay R\$37.7 billion in reparations for social, economic and environmental damages. The money has been allocated for several different programs (*e.g.* R\$4.4 billion were allocated to an income transfer program), some executed by the government with Vale's funding and others executed by the company itself.

Another measure adopted was the implementation of the dam de-characterization program, which intended to eliminate all of the company's 30 upstream dams by 2035, having currently completed 60% of the program, with 18 eliminated dams. There were also over 23 thousand compensation deals reached with affected individuals, amounting to R\$2.5 billion<sup>3</sup>, among several other repercussions, such as the acceleration of the transition to mining without tailing dams, which represented 78% of Vale's total production in Brazil in 2024<sup>4</sup>.

The full impact of this tragedy goes beyond scope of this paper. Previous literature indicates that 51% of the devastated 297.28 hectares of land consisted of preserved forest areas (Pereira, Cruz, and Guimarães 2019). Further research indicates that the contamination of the Paraopeba River Basin exceeded the legal threshold for toxic substances,

---

2. See Jasi (2019) for further details.

3. See Rodrigues (2024) for further details.

4. See Vale (2024) for further details.

such as the metals manganese and aluminium (Polignano and Lemos 2020). There were more than 3845 people directly affected (Machado de Freitas et al. 2019), among which both in-house and outsourced Vale workers represented 91% of the deaths of the tragedy, 47% and 44% respectively (Andrade da Silva et al. 2020).

## 1.2 Motivation

This work adds to the disaster literature by empirically evaluating the economic effects of the rupture of the Brumadinho tailings dam. The disaster literature has grown a lot in recent decades, especially due to the effects of climate change. Although Brumadinho's episode is a man made disaster, it is also related to the climate change part of this literature, because both man made and climate change induced catastrophes lead to socio-economic and environmental impacts.

Authors have estimated the impacts of disasters in several different contexts. Coffman and Noy (2012) used the synthetic control method to estimate the effects of a 1992 hurricane on the Hawaiian island of Kauai, showing that the island's population was still 12% smaller 18 years after the disaster than it would have been had the hurricane not occurred. Belasen and Polacheck (2008) use the generalized diff-in-diff technique and data from the US state of Florida to understand how hurricanes affect the labor market, finding a slight increase on the wage of 4% on the first quarter after the incident. In the Brazilian context, Matsunaga (2020) uses the diff-in-diff framework to evaluate the repercussions that the Mariana dam collapse, a similar incident, had on mental health, finding an increase in the number of hospitalizations for mental health disorders in the affected population.

For the case of Brumadinho's incident, there does not seem to be any robust empirical estimation of the economic effects of the disaster. This presents itself as a gap in the literature, which this paper sets out to fill by providing a robust quantitative analysis of the economic impact.

Methodologically, this paper builds on the work of Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010) and Abadie, Diamond, and Hainmueller (2015), using the synthetic control method to construct an estimated counterfactual for Brumadinho, which allows quantification of the effects of the dam rupture on the variables of interest. This work restricts its scope to the economic impact of the disaster, focusing on macroeconomic and labor market variables. This problem presents itself as an interesting object of study, since the effects of the disaster are not obvious. Although it is tempting to assume a negative impact on GDP per capita, wages and employment, the reparations program makes the net effect become uncertain. This becomes clear when we see that the nominal GDP of Brumadinho in 2019 was R\$2.5 billion, but the compensation agreement

alone was of R\$37.7 billion.

The remainder of this paper is organized as follows. Section 2 describes the database, the variables specification and the empirical methodology. Section 3 discusses the results for the interest variables *GDPPerCapita*, *Wage*, *LogWage* and *PropEmployed*. Section 4 analyzes the heterogeneity of the effects on 4 different economic sectors Agriculture, Industry, Services and Public Administration. Section 5 checks the robustness of results. Section 6 concludes.

## 2 Methodology

### 2.1 Overview

To estimate the impact of the disaster on the variables of interest *GDPPerCapita*, *Wage*, *LogWage* and *PropEmployed*, the synthetic control method was utilized, as presented in Abadie and Gardeazabal (2003) and further developed in Abadie, Diamond, and Hainmueller (2010). The donor pool was restricted to municipalities in the state of Minas Gerais (MG), which amounts to a total of 852 units when excluding Brumadinho, the treated unit. Since a donor pool that is too large can lead to overfitting and make synthetic control estimation unfeasible due to computational requirements, a matching technique was employed to select the 30 nearest-neighbor municipalities to Brumadinho according to a Manhattan distance measure, ensuring a donor pool similar to the treated unit.

After the matching, the synthetic control is estimated using pre treatment period data, which spans from 2002-2018 , though some variables have available data only starting from 2007. This provides an estimated counterfactual version of Brumadinho, allowing the calculation of the estimated treatment effects (gaps) for the post treatment period (after 2019) and the usage of placebo tests to perform inference.

Going further, subsection 2.2 explains the origin and composition of the data used, also presenting its variables. After that, subsections 2.3, 2.4 and 2.5 explain the methodology described in the last paragraph in more depth.

### 2.2 Database and Variables

The database used for the model is composed of municipal level panel data. It was constructed using data from four different sources IBGE (2025), Atlas do Desenvolvimento Humano no Brasil (2025) and Base dos Dados (2025). Table 1 displays the variables used, their time periods and data sources. The "Breakdown" column is used so that

related variables can be displayed in a single row, with their different dimensions being described in this column. For example, the database has gross value added data (*GVA*), with variables for the total *GVA* and 4 different sectoral *GVA* breakdowns of this total (Agriculture, Industry, Services and Public Administration), making up 5 different variables. The table gathers these 5 variables into a single row, assigning "Gross Value Added" to the "Measure" column and describing the different 5 different breakdowns of this measure on the "Breakdown" column.

Among the presented variables, 3 of them were built using SQL queries and python to extract them from Base dos Dados (2025) and save them to the database, since this source contains unidentified person level data and municipal level data is needed. *Wage* was built by calculating the mean of the average monthly wage for each given municipality and year using the RAIS Vínculos database. The Number of Companies variable was built using RAIS Estabelecimentos by counting the number of rows when grouping by year and municipality, which corresponds to the number of firms that reported data for each city and year, since in this data source each row contains a unique firm. Lastly, *Employed* was obtained from the RAIS Estabelecimentos database by summing the number of employees of each company for every municipality and year. This measure only shows the formal employment in December of each year, since RAIS only captures formal employment data and is a yearly survey done every month of December.

Other compound variables were created using the original ones. This is important because at times the most relevant variables for the study are not immediately available via the official sources, but can be obtained by creating new variables that alter existing data in a meaningful way. For example, we obtain the per capita *GVA* measures, dividing the *GVA* by the population. This measure is more useful in many applications, such as comparing the relative size of the four sectors between municipalities with very different populations.

Another important manipulation was the calculation of *LogWage*, a logarithmic transformation of the *Wage* variable. This is important since wages tend to present exponential growth time series and this transformation log-linearizes them. Besides that, the treatment effect estimators obtained from this transformation can be interpreted as the approximate percentage effect:

$$\begin{aligned} \delta &= \ln(Wage^I) - \ln(Wage^U) = \ln\left(\frac{Wage^I}{Wage^U}\right) = \ln\left(\frac{Wage^U + \Delta Wage}{Wage^U}\right) = \ln\left(1 + \frac{\Delta Wage}{Wage^U}\right) \Rightarrow \\ \delta &\approx \frac{\Delta Wage}{Wage^U}, \quad \text{since } \Delta Wage = Wage^I - Wage^U \text{ and } \ln(1 + x) \approx x \text{ for small } x. \end{aligned} \tag{1}$$

Where  $\delta$  is the treatment effect estimator,  $Wage^U$  is the  $Wage^I$  variable in the absence of the intervention and  $Wage^I$  is the wage variable in the presence of it. Table 2

Measure	Breakdown	Time Period	Data Source
Gross Value Added ( <i>GVA</i> )	Total and 4 Sectors <sup>1</sup>	2002-2021	(IBGE 2025)
<i>GDP</i>	Total and Per Capita	2002-2021	(IBGE 2025)
Population ( <i>POP</i> )	Total	2002-2021	(IBGE 2025)
Rural Population	Total	2010	(IBGE 2025)
<i>IDHM</i> <sup>2</sup> (Municipal HDI)	Total, Income, Longevity and Education	2010	(Atlas do Desenvolvimento Humano no Brasil 2025)
Transfers from Bolsa Família <sup>3</sup> (BF Transfers)	Per Capita	2013-2017	(Atlas do Desenvolvimento Humano no Brasil 2025)
<i>GINI</i>	Total	2010	(IBGE 2025)
Literacy Rate	Total	2010	(IBGE 2025)
Proportion of Population in Minimum Wage Bracket	Brackets: 0-1; 1-2; 2-5; 5-10; 10+ Minimum Wages	2010	(IBGE 2025)
Average Wage ( <i>Wage</i> )	Total	2002-2024	(Base dos Dados 2025)
Number of Companies	Total	2002-2023	(Base dos Dados 2025)
Total Employment ( <i>Employed</i> )	Total	2002-2024	(Base dos Dados 2025)

<sup>1</sup> Agriculture, Industry, Services and Public Administration.

<sup>2</sup> IDHM is an adaptation of the HDI, for municipalities in Brazil.

<sup>3</sup> Bolsa Família is a conditional cash transfer program in Brazil.

Table 1: Original variables, their time periods and respective data sources.

displays the created variables, specifying how they were obtained.

Measure	Breakdown	Time Period	Formula
$GVA_{PerCapita}$	Total and 4 Sectors <sup>1</sup>	2002-2021	$\frac{GVA}{POP}$
Sector's Share of Total GVA	4 Sectors	2002-2021	$\frac{GVA_{Sector}}{GVA_{Total}}$
Proportion of Employment ( $PropEmployed$ )	Total	2002-2021	$\frac{Employed}{POP}$
Natural Log of Average Wage ( $LogWage$ )	Total	2002-2024	$\ln Wage$

<sup>1</sup> Agriculture, Industry, Services and Public Administration.

Table 2: Compound variables, their time periods and formulas.

## 2.3 Matching

The matching algorithm is used to limit the donor pool, preventing overfitting and making estimation feasible. This process is applied for each of the interest variables independently, creating separate donor pools for them, these different divisions will be referred to as "interest groups" from here. This ensures that the groups from which the synthetic controls will be built are optimized to better recreate the path of the specific interest variable for which they were matched, while preventing overfitting.

The process starts by selecting the variables on which the Manhattan distance measure will be calculated, these variables are specific to each of the interest groups and were selected to improve the pre treatment fit of their synthetic control. Each observed pre treatment year for a variable is considered as its own independent variable, being then used to calculate the distance, we call this new set of variables "matching variables". For example, if  $GDPP_{PerCapita}$  is included in the matching, there would be a variable created for the 2002 GDP named  $GDPP_{PerCapita}2002$ , one for 2003 named  $GDPP_{PerCapita}2003$  and so on for every year until 2018, they would all be treated as separate matching variables for the purpose of calculating the distance.

Each matching variable is then standardized by calculating its z-score, so that scale does not affect the result. Let the sample mean of the matching variable  $x$  be  $\bar{x}$ , the sample standard deviation of  $x$  ( $std_x$ ) is:

$$std_x = \sqrt{\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{N - 1}} \quad (2)$$

For each observation  $x_i$  of  $x$ , its z-score  $z_{x_i}$  can be calculated as:

$$z_{x_i} = \frac{x_i - \bar{x}}{std_x} \quad (3)$$

The z-score of the matching variables substitute the original ones to calculate the distance. From now on, when the matching variables are mentioned, it refers to the calculated z-score.

Suppose there are  $N + 1$  municipalities of which units 2 through  $N + 1$  are untreated and 1 is treated, consider, as well, that each of them has  $P$  matching variables. For the unit  $n$ , its set of variables can be represented by the vector  $MV_n$ , this vector has  $P$  dimensions, one for each matching variable. We can then represent each element of  $MV_n$  as  $MV_{pn}$ , which is matching variable  $p$  of unit  $n$ . The distance between the treated unit and unit  $n$  can then be calculated as:

$$\sum_{p=1}^P \left| \frac{MV_{pn} - MV_{p1}}{MV_{p1}} \right| \quad (4)$$

Where  $\frac{MV_{pn} - MV_{p1}}{MV_{p1}}$  represents the percent difference between the two units.

The distance is calculated for each of the municipalities of Minas Gerais, excluding Brumadinho, for a total of 852 distance measures. The municipalities with the smallest 30 distance measures are selected to compose the donor pool of the interest group for which the matching was made.

## 2.4 Model

After the donor pool is defined by the matching algorithm, the synthetic control method is used to evaluate the impact, as presented in Abadie and Gardeazabal (2003) and further developed in Abadie, Diamond, and Hainmueller (2010). This subsection describes this model, explaining the math and intuition behind its estimation. For a more in depth explanation of synthetic control, read Abadie (2021). To estimate this model with the data, the R package "Synth", presented in Abadie, Diamond, and Hainmueller (2011), is used.

Suppose there are  $N + 1$  municipalities, of which units 2 through  $N + 1$  are untreated and 1 is treated. Additionally, there are  $T$  periods in total, with  $T_0$  periods before treatment ( $t = T_0$  is the last period before the intervention). The outcome of interest for unit  $n$  and period  $t$  is  $Y_{nt}$ , there is also a prediction vector  $X_n$  of  $K$  dimensions for any unit  $n$  ( $K$  predictors), where  $X_n = \{X_{1n}, \dots, X_{Kn}\}$  and  $X_{kn} \quad k \in \{1, \dots, K\}$  is a predictor for the outcome variable for unit  $n$  and could have been observed in any time period  $t$  before the intervention ( $t \leq T_0$ ), but not after. It is important to note that the vector  $X_n$  can, and probably should, contain pre intervention values of the outcome as predictors. Thus,

the matrix  $X_0$  that collects all prediction vectors for the untreated units is constructed as  $X_0 = \begin{pmatrix} X_2 & \dots & X_{N+1} \end{pmatrix}$ .

For unit  $n$ , the potential outcome without intervention for period  $t$  is defined as  $Y_{nt}^U$  and the potential outcome with intervention for the same period is  $Y_{nt}^I$ , they represent the value that the outcome would assume for unit  $n$  in a scenario without and with treatment, respectively, for period  $t$ . For any unit  $n$  such that  $n \in \{2, \dots, N+1\}$  (untreated units), it is true that the realized outcome is equal to the potential outcome without intervention for all periods  $Y_{nt} = Y_{nt}^U \quad \forall t$ , because these units are not treated by assumption. On the other hand, for  $n = 1$  (the treated unit), the realized outcome before treatment is equal to the potential outcome without intervention and after treatment it is equal to the potential outcome with intervention, so that  $Y_{1t} = Y_{1t}^U \quad t \leq T_0$  and  $Y_{1t} = Y_{1t}^I \quad t > T_0$ .

The treatment effect for unit 1 in period  $t$  is then defined as  $\tau_{1t} = Y_{1t}^I - Y_{1t}^U$  and our goal is to estimate it for periods after the intervention (estimate  $\tau_{1t} \quad t > T_0$ ). Since  $Y_{1t} = Y_{1t}^I \quad t > T_0$ , the equation can be rewritten for post intervention periods as:

$$\tau_{1t} = Y_{1t} - Y_{1t}^U \quad t > T_0 \quad (5)$$

Now, the problem boils down to finding an estimate for  $Y_{1t}^U$ , which is the outcome value for unit 1 and period  $t$  if it had not been treated, that is, the counterfactual outcome. The way that the synthetic control tackles this is by creating a weighted average of the untreated units from the donor pool, this is the so called synthetic control, because it is an artificial control unit built from the donor pool of untreated units. The synthetic control estimator for the counterfactual outcome is then:

$$\hat{Y}_{1t}^U = \sum_{n=2}^{N+1} w_n Y_{nt} \quad (6)$$

Where  $w_n$  represents the weight attributed to unit  $n$ . The weights are restricted to be nonnegative and sum to one, which is a reasonable assumption and important for the interpretability of the estimator. The treatment effect estimator becomes:

$$\hat{\tau}_{1t} = Y_{1t} - \hat{Y}_{1t}^U \quad t > T_0 \quad (7)$$

Thus, the treatment effect estimator is the difference between the outcome of the treated unit and that of the synthetic control, this difference will then be called the gap between the two trends. This name will be used in the section 3 when presenting the results. The gaps will also be calculated for the pre intervention period, being expected to be close to 0.

The problem that arises now is the choice of the weights, the goal is to choose the optimal vector of weights  $W^* = \begin{pmatrix} w_2 & \dots & w_{N+1} \end{pmatrix}^T$  so that the best treatment effect

estimator is obtained. Abadie and Gardeazabal (2003) propose solving the following minimization problem for choosing  $W^*$ :

$$\begin{aligned} W^* &= \arg \min_W \|X_1 - X_0 W\| = \arg \min_W \left( \sum_{k=1}^K v_k (X_{k1} - w_2 X_{k2} - \dots - w_{N+1} X_{kN+1})^2 \right)^{1/2} \\ \text{s.t. } & \sum_{n=2}^{N+1} w_n = 1; \quad w_n \geq 0 \quad \forall n \end{aligned} \tag{8}$$

$\|X_1 - X_0 W\|$  is a weighted measure distance between the predictors for the treated unit and the synthetic control defined by  $W$ . Thus, solving this problem means finding the set of weights that renders the smallest distance measure between the treated unit and the synthetic control, in other words, it is simply finding the most similar synthetic control according to the defined distance measure. The measure distance weights are contained in the vector  $V = \begin{pmatrix} v_1 & \dots & v_K \end{pmatrix}^T$ , and define how important each predictor is for the estimation, so that the higher  $v_k$  is, the more important it is to minimize the difference  $(X_{k1} - w_2 X_{k2} - \dots - w_{N+1} X_{kN+1})^2$  for the predictor  $k$ , relative to this difference for the other predictors.

The problem of defining the optimal vector of measure weights  $V^*$  remains, since every choice of  $V^*$  will render a specific solution  $W(V^*)$ . Abadie and Gardeazabal (2003) propose solving the following problem to choose  $V^*$ :

$$\begin{aligned} V^* &= \arg \min_V \left( \sum_{t=1}^{T_0} (Y_{1t} - w_2(V) Y_{2t} - \dots - w_{N+1}(V) Y_{N+1t})^2 \right)^{1/2} \\ \text{s.t. } & v_k \geq 0 \quad \forall k \end{aligned} \tag{9}$$

Solving this means selecting  $V^*$  so that the associated set of weights  $W(V^*)$  generates the smallest difference between the outcomes of the treated unit and the synthetic control.

The entirety of this process is important to find a synthetic control that is not overfit, but actually meaningful. If the weights for our counterfactual estimator were found simply by minimizing the difference between outcomes and not of all the other predictors as in equation 8, then the risk of overfitting would be high. In that case it would be much more likely to obtain a set of weights that fit very well the pre treatment path of the outcome for the treated unit, but that generate a synthetic control that is very different from the treated unit on its characteristics, possibly generating a post treatment path for the synthetic control that does not approximate the counterfactual outcome of the treated unit  $\hat{Y}_{1t}^U \quad t > T_0$ . Minimizing the difference on the other predictors helps prevent this, because it generates a set of weights that approximate not only the outcome path, but also the observable characteristics (the predictors), generating an estimate that is

more robust to the overfitting problem.

Another important thing to note is that having a good synthetic control entails observing a good pre treatment fit on the outcome path, because if the pre treatment fit is poor, chances are that the post treatment outcome will not estimate the counterfactual  $\hat{Y}_{1t}^U$   $t > T_0$  well. For this reason, it is essential to have a long pre intervention time span, meaning that  $T_0$  should be as large as possible.

## 2.5 Inference

The goal of the inference here is to define a way of testing the null hypothesis of no treatment effect, against the alternative hypothesis of existence of treatment effect. As proposed by Abadie, Diamond, and Hainmueller (2010), a placebo distribution for the absence of treatment can be calculated iteratively using the donor pool units. This means calculating a synthetic control estimator for each of the units in the donor pool, which is called a placebo test.

This process will iterate through the donor pool, selecting one unit at a time, say unit  $n$ . It will then construct a placebo donor pool for this unit, which will be the original donor pool excluding unit  $n$  itself, notice that the treated unit is not included in this group. Then, the synthetic control estimator described in subsection 2.4 will be obtained for unit  $n$  using its placebo donor pool.

Once the iteration is done, there will be  $N$  placebo test estimations, which can be used as a comparison to the results for the treated unit. The idea is to see if the result obtained from the synthetic control of the treated unit is significantly more extreme than that of the placebo tests. A way to do that is to use the MSPE as a measure of fit, so that a more extreme result would render a higher MSPE, a worse fit.

The problem with this is that since the donor pool was designed to have units similar to the treated one, it does not necessarily contain similar enough units to each of the untreated ones, so that when the placebo tests are calculated it is common to see many estimations with very poor pre treatment fit. Synthetic controls that have poor pre treatment fit are much more likely to present poor post treatment fit, even if there is no intervention, which would make our results for the treated unit seem less extreme, even though they are not.

Abadie, Diamond, and Hainmueller (2010) propose using a post/pre MSPE ratio (*RatioMSPE*) that helps correct this issue. It does that by comparing how extreme the post treatment MSPE is in relation to the pre treatment MSPE, instead of just comparing post treatment MSPE alone. The  $\text{RatioMSPE}_n$  is calculated for unit  $n$  through:

$$RatioMSPE_n = \left( \frac{T_0}{T - T_0} \frac{\sum_{t=T_0+1}^T (Y_{nt} - \hat{Y}_{nt}^U)^2}{\sum_{t=1}^{T_0} (Y_{nt} - \hat{Y}_{nt}^U)^2} \right)^{1/2} \quad (10)$$

Another adjustment suggested by Abadie, Diamond, and Hainmueller (2010) to prevent the problem is to discard placebo tests that have a pre treatment MSPE  $MSPE_{t \leq T_0}$  higher than a threshold of the treated unit's pre treatment MSPE. The threshold used in this paper if of 5 times, so that any placebos  $n$  such that  $MSPE_{n,t \leq T_0} > 5MSPE_{1,t \leq T_0}$  are discarded from the placebo test group. Say that there were  $P$  units that satisfied this condition, this means that the final number of placebo units used will be  $N - P$ . We can then construct a set containing the indexes of the remaining units in the placebo test group:

$$\mathcal{P} = \{n : n \in \{2, \dots, N + 1\} \text{ and } MSPE_{n,t \leq T_0} \leq 5MSPE_{1,t \leq T_0}\} \quad (11)$$

When the placebo tests that pass the threshold are discarded and  $RatioMSPE$  is calculated for all of the remaining, a pseudo distribution for the absence of treatment against which to compare  $RatioMSPE_1$  is naturally obtained. A pseudo p-value is then calculated by:

$$p = \frac{1}{N - P} \sum_{n \in \mathcal{P}} I(RatioMSPE_1 \geq RatioMSPE_n) \quad (12)$$

Where  $I(\cdot)$  is an indicator function that is equal to 1 if the condition is satisfied and 0 otherwise. The pseudo p-value ( $p$ ) measure indicates the proportion of placebo units that had a  $RatioMSPE$  at least as extreme as the one obtained for the treated unit.

The standard threshold for  $p$  is assumed, so that if  $p < 0.05$  the null hypothesis of no treatment effect is rejected. It is important to note that in this last step, where the null hypothesis is tested, the evaluation consists of a joint significance test for all of the post treatment treatment effects (gaps) together.

## 3 Results

### 3.1 Overview

In this section I present and interpret the results of the model for all interest variables, each on their own subsections. In the subsections, I present the synthetic control results with a graph containing the time series for Brumadinho and its synthetic control, and later another graph showing the gaps between the two. Lastly, I display the graphs for

the placebo distribution and the inference results.

The interest variables are  $GDPPerCapita$ ,  $Wage$ ,  $LogWage$  and  $PropEmployed$ .  $Wage$  and  $LogWage$  are presented to analyze the difference in both specifications, since it is standard practice in economic literature to use a logarithmic transformation in wage variables<sup>5</sup>. Additionally, I make a transformation on the  $GDPPerCapita$  results by multiplying it by the yearly population of Brumadinho to obtain the effects for the Total  $GDP$ .

## 3.2 GDP Per Capita

### 3.2.1 Overview

This subsection presents the estimation results for the  $GDPPerCapita$  interest variable, later expanding the analysis by obtaining the Total  $GDP$  transformation, which provides a better measure of the total disaster effect than the per capita variable.

### 3.2.2 Per Capita Effects

This subsection analyzes the synthetic control results for  $GDPPerCapita$ . Figure 3 displays the variable's time series for both Brumadinho and the synthetic control. There is a very good fit before treatment, as seen by how the trend of the synthetic unit for the variable of interest mimics almost perfectly that of Brumadinho. In the post treatment period, a gap opens between the two, with Brumadinho's GDP per capita being much lower than that of its control. It is noticeable that the difference between the two trends increases with time, reaching a much higher level in the last observed treatment year (2021).

The evolution of the gaps can be observed in Figure 4. Here it is clearly seen how the  $GDPPerCapita$  was impacted by the disaster. In the first year the gap is not very large, being close in dimension to that of 2010, before the treatment, but in 2021, the last treatment year, the gap is greater than R\$100,000.00, which represents a significant impact on the variable.

Even though the results look good visually, it is important to use inference to have a way of analyzing the results objectively. To achieve that, a placebo distribution is generated by calculating a synthetic control for each unit in the donor pool, and keeping the ones with a good enough fit<sup>6</sup>.

In Figure 5 the resulting gaps of the placebo tests, after the ones with poor fit are removed, are shown, along with the gaps for Brumadinho. From a visual inspection, the

---

5. See subsection 2.2 for further information.

6. See subsection 2.5 for further information.

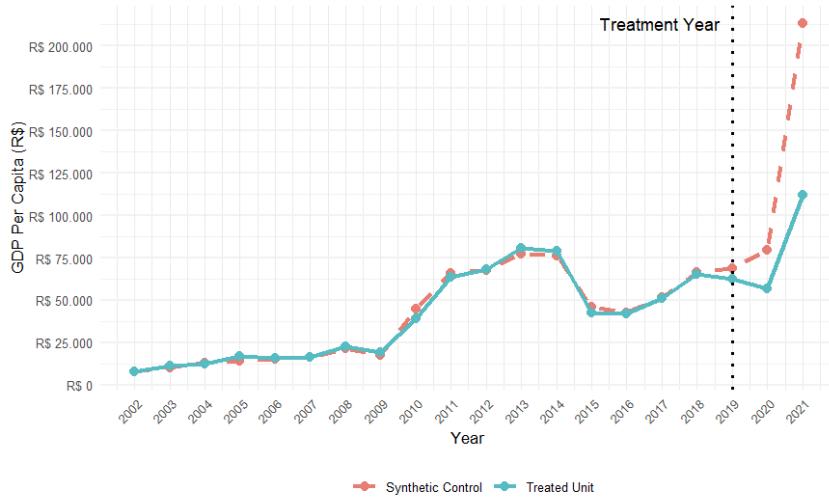


Figure 3: GDP Per Capita treated and synthetic trends.

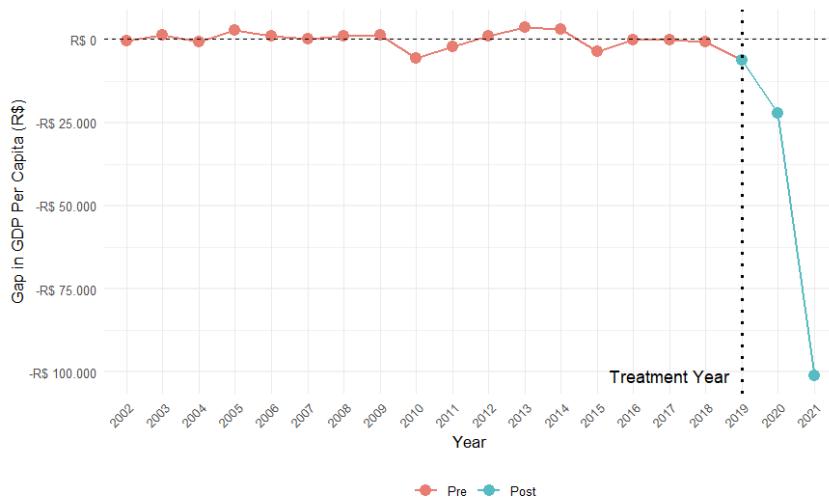


Figure 4: GDP Per Capita gaps.

post treatment gaps for the treated unit seem more extreme than those of most of the placebo tests.

The visual inspection of the gaps for the placebo test alongside that of Brumadinho give a better comparison measure than looking at Brumadinho's gaps alone. However, it is important to consider that Brumadinho's synthetic control has a better pre treatment fit than most of the other placebo units and that units with poor pre treatment fit are more likely to present extreme post treatment gaps, even without treatment.

To correct for the difference in fit, the post/pre MSPE ratio (*RatioMSPE*) is used<sup>7</sup>, as proposed by Abadie, Diamond, and Hainmueller (2010). This variable indicates how big the post treatment fit is relative to the pre treatment fit, with its use being more appropriate to analyze how extreme Brumadinho's gaps are compared to the placebo

7. See subsection 2.5 for further information.

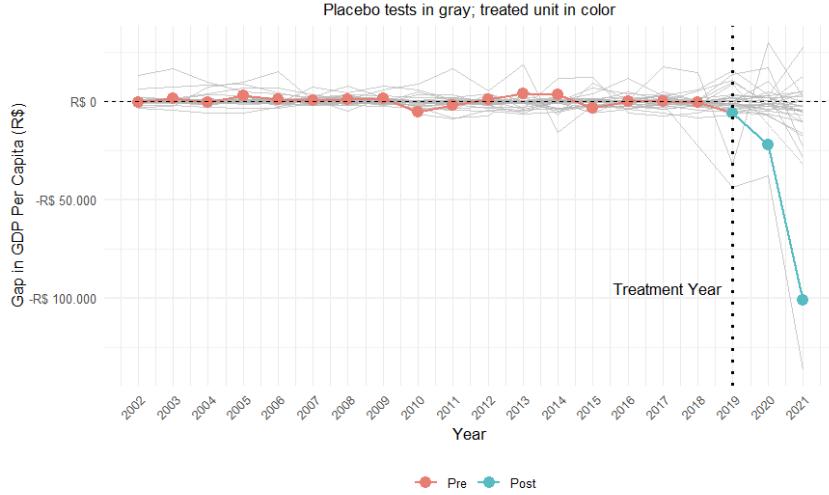


Figure 5: GDP Per Capita placebo gaps.

tests.

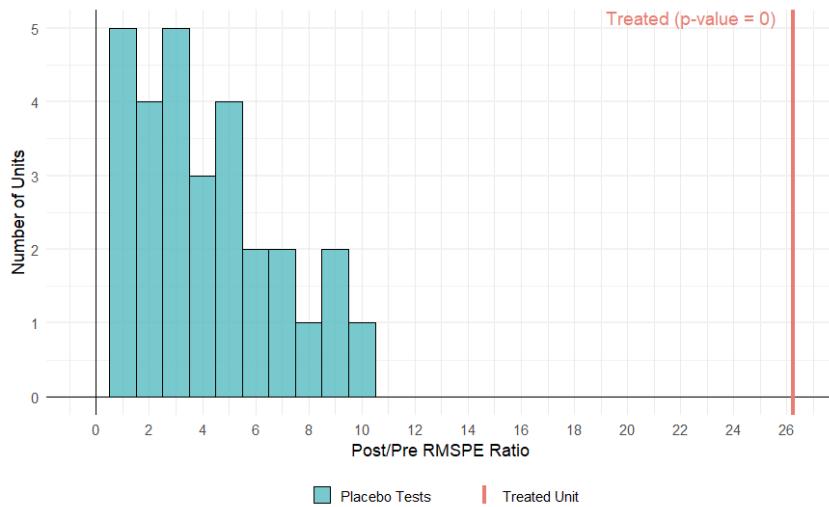


Figure 6: GDP Per Capita placebo distribution.

A placebo distribution for the donor pool’s *RatioMSPE* and a pseudo p-value are presented in Figure 6. We can see that Brumadinho’s *RatioMSPE* is over two times more extreme than that of any unit from the donor pool, which renders a p-value of 0. Thus, it is concluded that the effect of the dam rupture on *GDPPerCapita* was negative and statistically significant.

### 3.2.3 Total GDP transformation

a

### 3.3 Wage

In this subsection the model's results for *Wage* are analyzed. Figure 7 displays the time series of this variable for Brumadinho and its synthetic control.

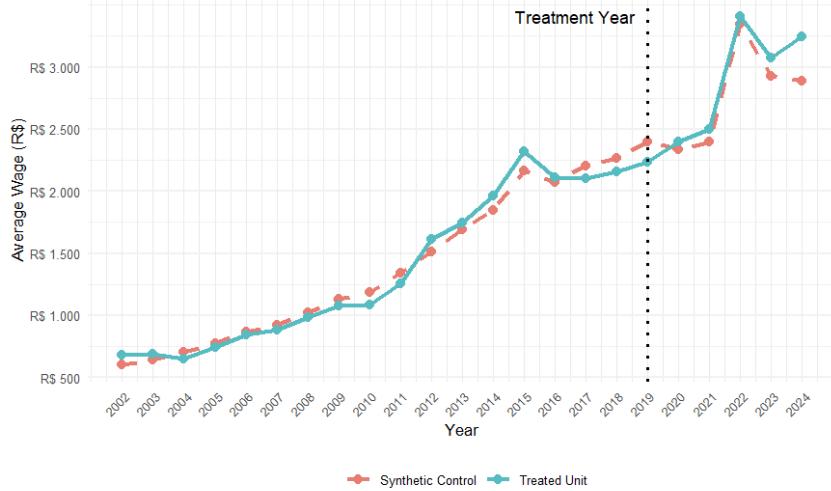


Figure 7: Wage treated and synthetic trends.

Before treatment, the path for the synthetic control seems to follow that of Brumadinho, although it is not as close as for the *GDPPerCapita* variable. This is likely due to wage variables being much more volatile than extremely aggregate measures such as gdp per capita.

Nevertheless, there is a good pre treatment fit. In the post intervention period, the path of the control also closely matches that of the treated unit, with visual inspection suggesting that the only relatively large gap is obtained in the last observed treatment year (2024).

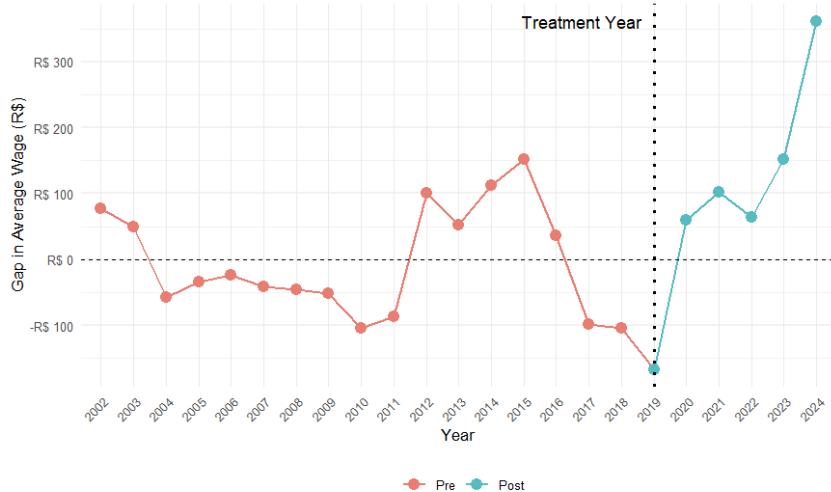


Figure 8: Wage gaps.

In Figure 8 the gaps between the time series are plotted and provide a different view of the results. It seems that the gaps sequence for the post treatment period is not very different from that of the pre treatment period, except for the last treatment year, where a gap slightly larger than the previous ones is obtained. However, an objective test is still necessary to have a conclusive result.

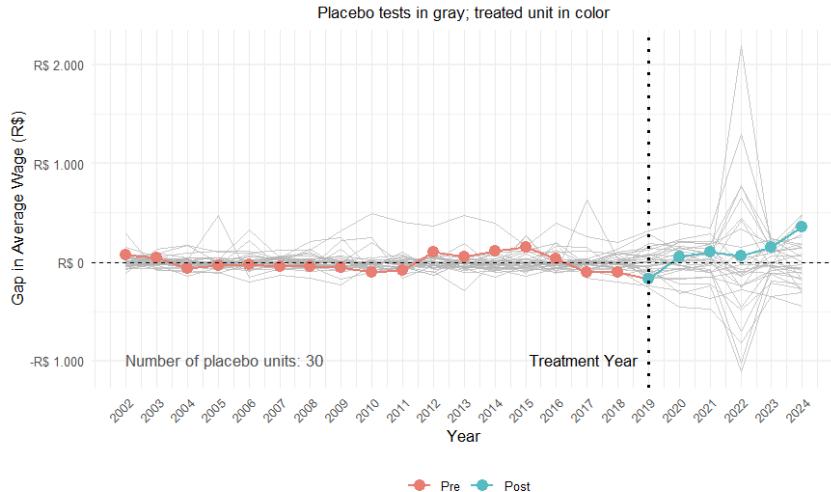


Figure 9: Wage placebo gaps.

Figure 9 provides the resulting gaps for the placebo tests alongside those of Brumadinho. The image shows that the treated unit's gaps are not as extreme as those of most of the placebo tests for most post intervention periods.

This insight becomes clearer when observing Figure 10. In this image, the placebo distribution for the MSPE ratio variable (*RatioMSPE*) is plotted alongside Brumadinho's obtained ratio value. This view shows that the treated unit's ratio is much less extreme than that of most placebo tests.

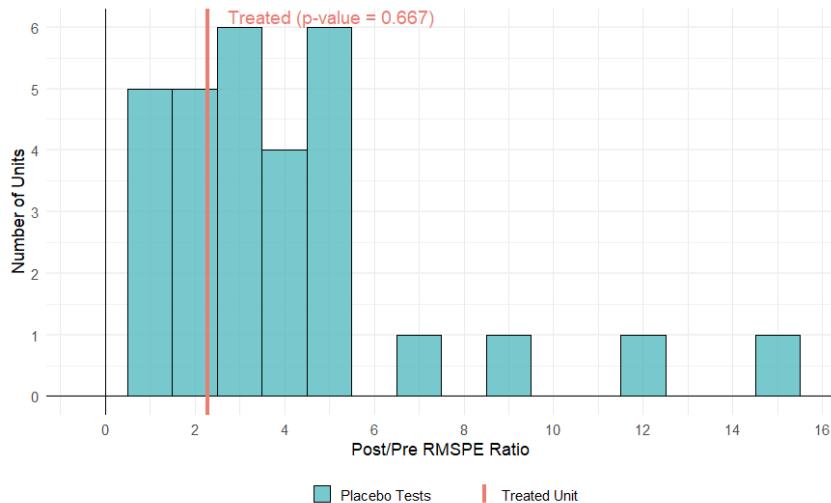


Figure 10: Wage placebo distribution.

Finally, the conclusion that Brumadinho's ratio is not extreme is confirmed by a high pseudo p-value of 0.667, which means that 66.7% of the placebo tests had a ratio at least as extreme as that of the treated unit. Thus, it is concluded that there is not enough evidence to reject the null hypothesis of no effect from the dam rupture on *Wage*.

### 3.4 Log Wage

In this subsection, the results of the synthetic control model for *LogWage* are analyzed. This is a different specification of the model for the wage variable, using a logarithmic transformation, which is a standard procedure in economic literature as introduced by Mincer (1974).

This kind of variable manipulation is useful because it provides gaps (treatment effect estimators) that can be interpreted as the approximate percentage rate return<sup>8</sup>. Another important property of this transformation is that it log-linearizes wage time trends, since this variable tends to have exponential growth over time. Both of these properties can help achieve a better synthetic control estimator and more interpretable results, the percentage changes.

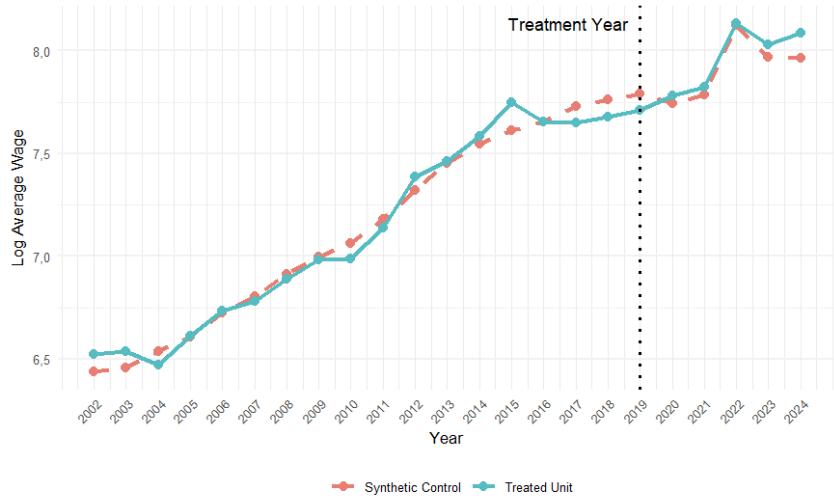


Figure 11: Log Wage treated and synthetic trends.

Figure 11 plots the *LogWage* variable time trends for the treated and synthetic control units for comparison. Once again, the pre intervention *LogWage* control time series seems to fit Brumadinho's series well, although it is not as good the obtained fit for the *GDPPerCapita* variable. In the post intervention period, the synthetic control also seems to closely match the Brumadinho's path for the variable of interest. This indicates that there might not have been any treatment effect on the variable, since the built control does not deviate much from the treated unit.

8. See subsection 2.2 for further information.

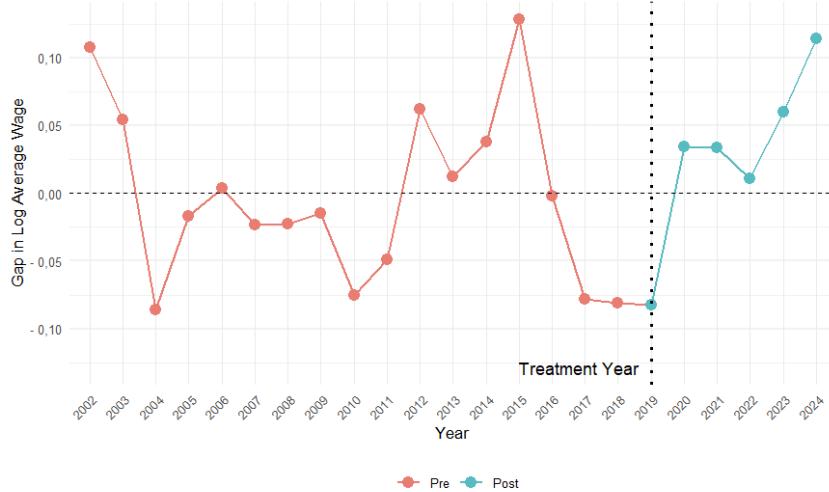


Figure 12: Log Wage gaps.

Figure 12 provides a better view of the difference between the time trends of both variables. The post intervention gaps appear to not deviate much from the pre intervention gaps, pointing towards the absence of impact on this variable. To further analyze this possibility, placebo tests were performed for statistical inference.

In Figure 13, the gaps for Brumadinho and the placebo tests are presented. It seems that the gaps for the treated unit are not as extreme as most of those for the placebo tests, which further indicates that there was no treatment effect on the *LogWage* variable.

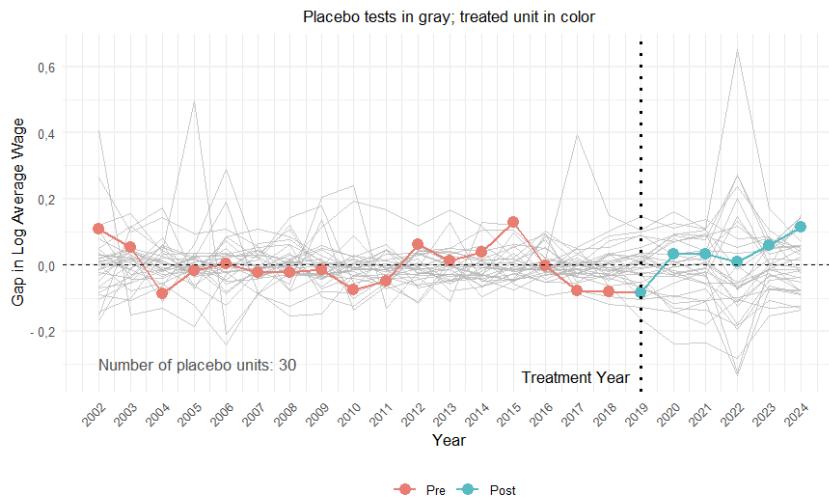


Figure 13: Log Wage placebo gaps.

To objectively assess the existence of intervention effect, Figure 14 presents the placebo distribution of the MSPE ratio (*RatioMSPE*), alongside the ratio of Brumadinho. It becomes clear that the treated unit's result is not more extreme than most of the placebo distribution, with a calculated pseudo p-value of 0.8, indicating that 80% of the placebo tests have a ratio at least as extreme as that of the treated unit.

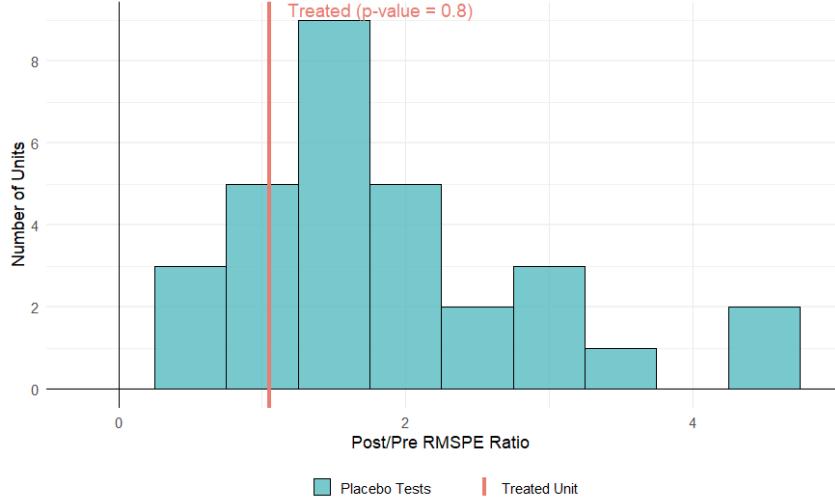


Figure 14: Log Wage placebo distribution.

Thus, it is concluded that there is not enough evidence to reject the null hypothesis that the dam rupture did not impact *LogWage*.

### 3.5 Prop Employed

This subsection analyzes the model's results for *PropEmployed*. This variable is chosen instead of *Employed* for the analysis of the effect on employment, because it represents total employment as a proportion of the population, a measure that is not affected by municipality size as the raw employment is. This makes it more appropriate for the synthetic control analysis, since an untreated unit of different size than the treated one, but very similar in other relevant characteristics will likely not present a comparable *Employed* trend, but will probably have a *PropEmployed* trend very close to that of the treated one.

Figure 15 presents the time trends of *PropEmployed* for both Brumadinho and the synthetic control. Despite the existence of more noise than in the other variables' analysis, the trends appear to follow a relatively close path in the pre treatment period.

The noise probably stems from different reasons, such as the variable representing the employment data of the formal sector only for December of each year, not containing the information of the other months. If the monthly information was available, it would be possible to calculate a yearly average proportion of employment, which could help reduce noise. It is also likely that the employment data is more volatile than the other variables, making it harder to find a good pre treatment fit.

After the intervention, a gap is seen between both units, indicating that there might have been a slight positive impact on employment in 2019, and a negative impact in 2020 and 2021. However, the post treatment differences does not look much more extreme than

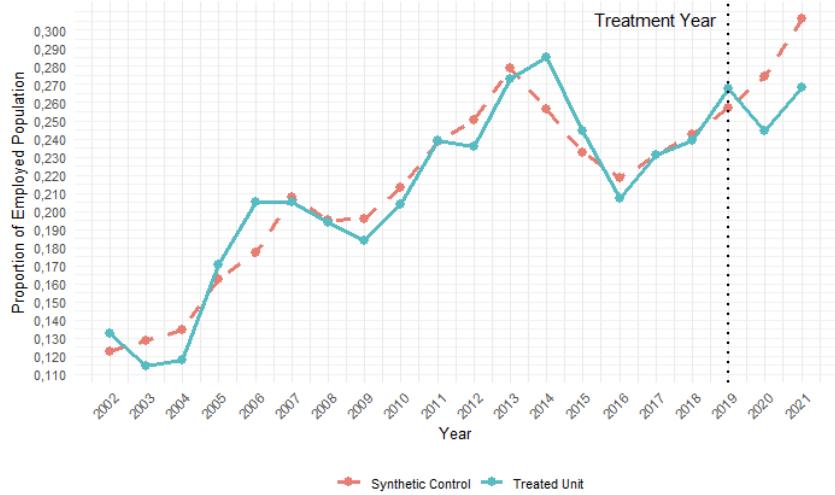


Figure 15: Proportion of Employment treated and synthetic trends.

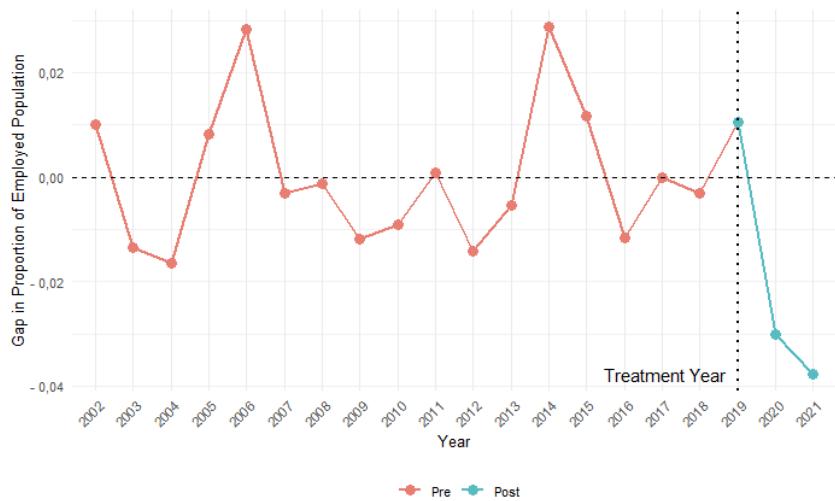


Figure 16: Proportion of Employment gaps.

some of the gaps observed before treatment, indicating that they might not be significant, so that further evaluation is necessary.

The gaps displayed in Figure 16 provide new insight. Although the gap in 2021 is the biggest, the image shows that the post treatment gaps are not much greater than some of the ones before the intervention. This means that there might not be enough evidence to reject the null hypothesis of no treatment effect. The placebo test inference is used to obtain an objective assessment on the significance of these results.

Figure 17 shows the gaps for Brumadinho and the placebo tests. The post treatment gaps for the treated unit do not appear more extreme than those of the placebo tests, which indicates that the treatment effects found are likely not statistically significant. The placebo distribution is used, to provide a conclusive analysis.

In Figure 18 the placebo distribution for *RatioMSPE* is plotted alongside the ratio variable of Brumadinho. The pseudo p-value is 0.483, meaning that 48.3% of the placebo

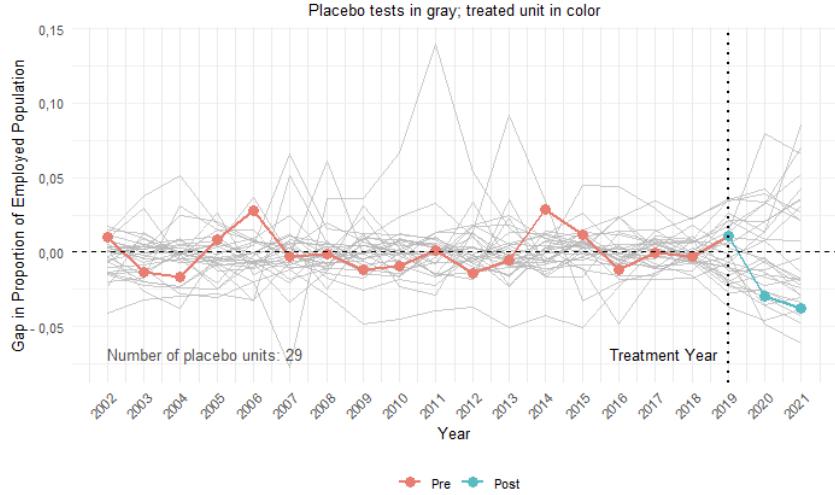


Figure 17: Proportion of Employment placebo gaps.

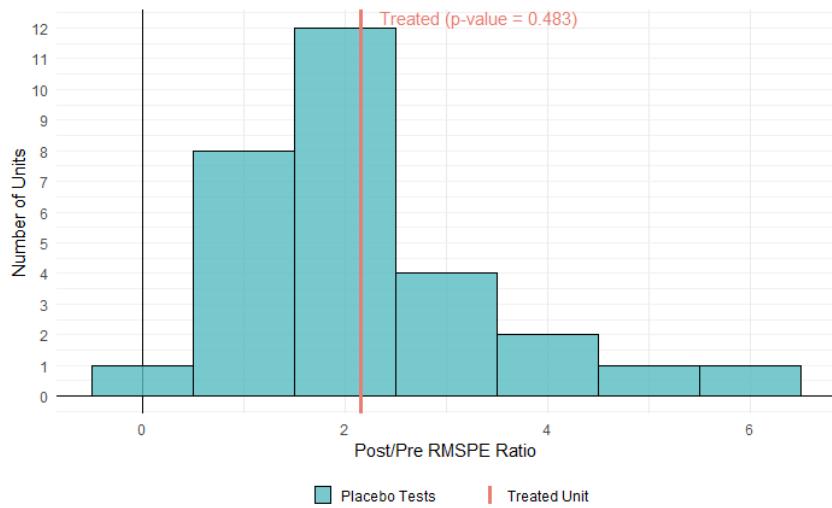


Figure 18: Proportion of Employment placebo distribution.

tests have a calculated *RatioMSPE* at least as extreme as that of the treated unit. This leads to the conclusion that there is not enough evidence to reject the null hypothesis that the dam rupture did not impact *PropEmployed*.

## 4 Heterogeneity

### 4.1 GVA Per Capita

In this subsection, we analyze the impact of the disaster on the 4 sectors' gross value added per capita variable. These *GVAPerCapita* measures represent a breakdown of *GDPPerCapita*, previously evaluated in subsection 3.2, into its components of the 4 sectors Agriculture, Industry, Services and Public Administration. By doing this break-

down, it is possible to understand which areas of Brumadinho's economy were impacted and to what extent that happened.

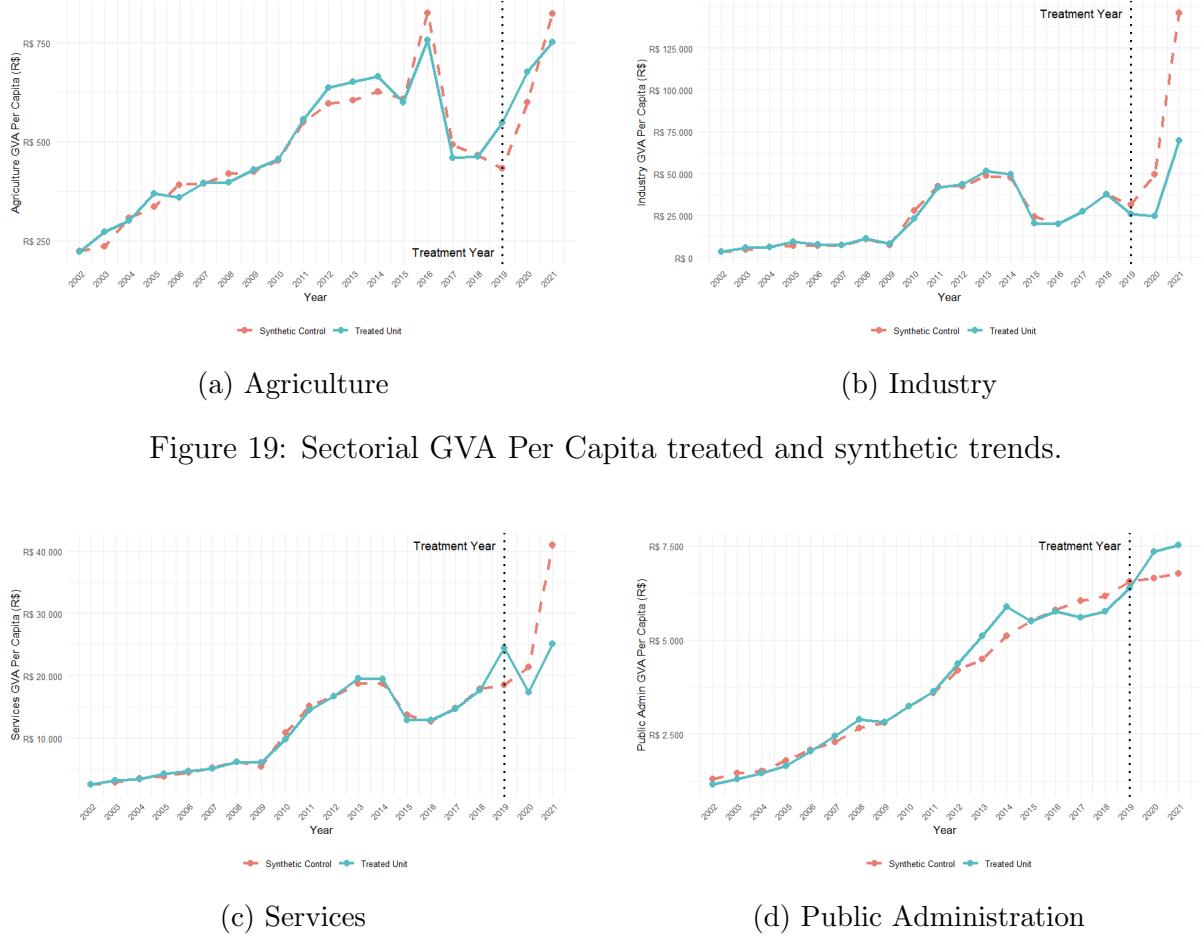


Figure 19: Sectorial GVA Per Capita treated and synthetic trends.

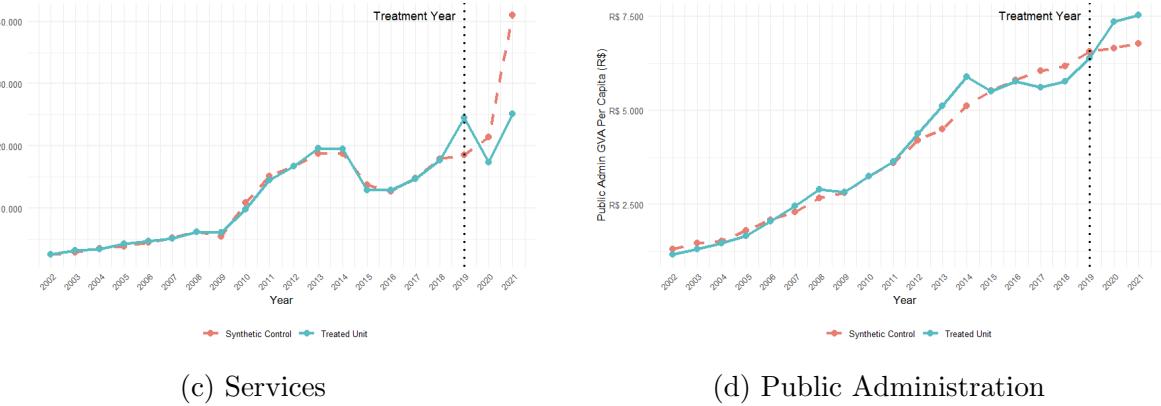


Figure 19: Sectorial GVA Per Capita treated and synthetic trends. (Continued)

Figure 19 presents Brumadinho's and each of the synthetic controls' time series for each of the sectoral *GVAPerCapita* side by side. The graphs show that the series for the Industry and Services sectors display similar trajectories to that of the *GDPPerCapita* shown in Figure 3, with positive gaps for 2020 and 2021. However, the Industry sector presents a very small negative gap for 2019 and the Services sector show a large positive gap for that year. For the Agriculture sector there is a positive gap for the first two treatment years and a negative one for the last, but the results seem noisier than that of the Industry and Services sector. For the Public Administration sector, there is a slight negative gap in 2019, followed by two bigger positive gaps in 2020 and 2021, but the results are also noisier.

In Figure 20 the gaps for each sector are plotted. For the Industry and Services sectors, it becomes clear that the post treatment gaps are much more extreme than the pre treatment gaps obtained, being supporting evidence for the existence of the previously described treatment effects. The gaps series for the Agriculture and Public Administration sectors appear much noisier and it is unclear whether there was an intervention

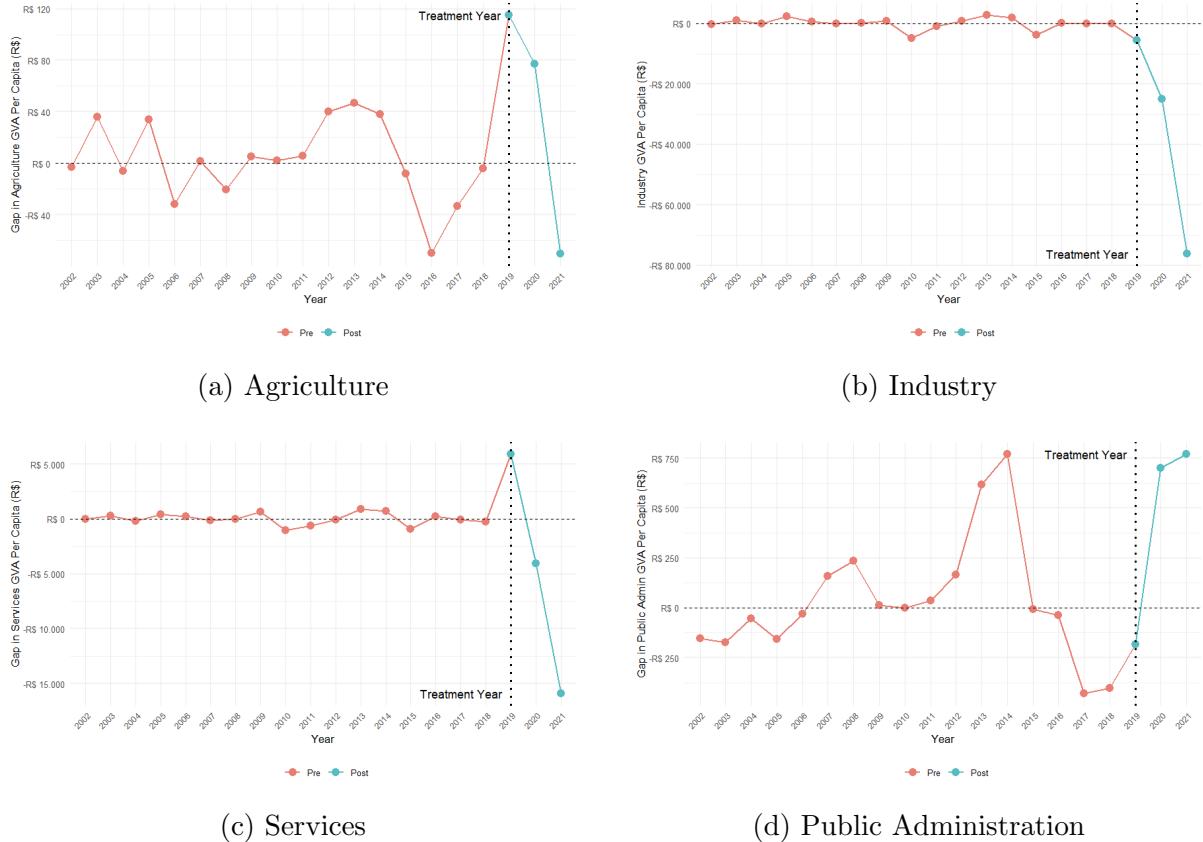
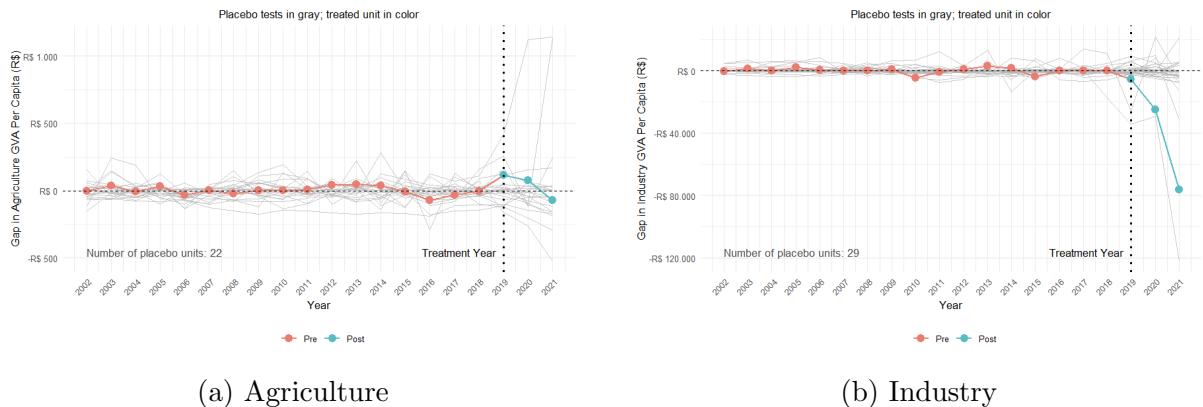
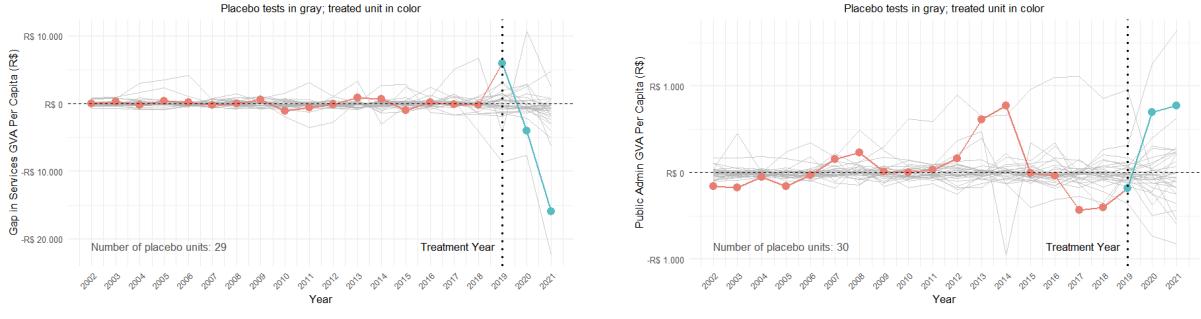


Figure 20: Sectorial GVA Per Capita gaps.

impact or not, although the Agriculture has post treatment gaps that seem more extreme compared to the pre treatment ones than those of the Public Administration sector.



In Figure 5 it is possible to see the gabs of the placebo tests and Brumadinho for the 4 sectors. Visual inspection of the graphs show that the post treatment gaps for the Industry and the Services sectors continue to appear extreme, even among the placebo tests. The Agriculture post intervention gaps do not seem to approach extreme levels compared to the placebo tests. Lastly, for the Public Administration sector the post treatment gaps appear to be among the most extreme, however, it is important to notice that the pre treatment gaps are also among the ones with the worst fit, which will correct

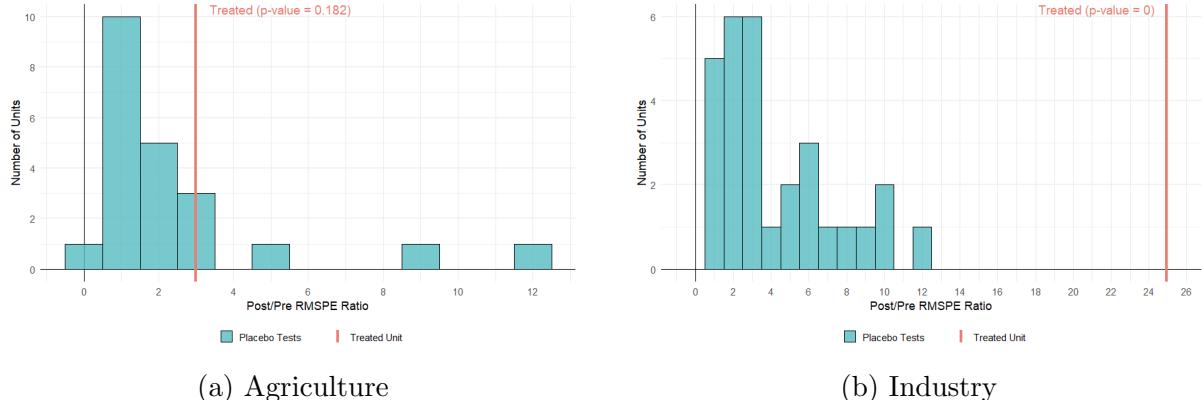


(c) Services

(d) Public Administration

Figure 21: Sectorial GVA Per Capita placebo gaps.

the large gaps when calculating  $RatioMSPE$ .



(a) Agriculture

(b) Industry

(c) Services

(d) Public Administration

Figure 22: Sectorial GVA Per Capita placebo distributions.

Figure 22 shows the calculated placebo distributions of  $RatioMSPE$  alongside Brumadinho's ratio for each of the 4 sectors. This display helps conclude that the Industry and Services sector had statistically significant impacts from the dam rupture, with both presenting a pseudo p-value of 0. The Public Administration of the treated unit had a  $RatioMSPE$  which is not extreme, with a pseudo p-value of 0.5, meaning that 50% of the placebo tests had a ratio at least as extreme as that of Brumadinho, leading to the

conclusion that there is not enough evidence to reject the null hypothesis of no impact from the dam rupture on the Public Administration's  $GVAPerCapita$ . The Agriculture sector also presented a ratio that was not extreme enough, rendering a pseudo p-value of 0.182, which also leads to the conclusion that there is not enough evidence to reject the null hypothesis of no dam rupture effect on the sector's  $GVAPerCapita$ .

## 5 Downstream Robustness

In this section, the robustness of the results for all the interest variables ( $GDPPerCapita$ ,  $Wage$ ,  $LogWage$  and  $PropEmployed$ ) is evaluated by constraining the donor pool. This is important because the mud from the disaster reached the Paraopeba River and, therefore, the municipalities downstream of the dam, although there was no physical destruction outside of Brumadinho. Thus, there might still have been effects from the mud on other downstream municipalities, as well as spillover effects from the impact. This check is performed by taking the downstream municipalities out of the matching pool, so that the 30 selected units are not from this group.

Figure 23 shows the maps of the donor pools generated by matching for  $GDPPerCapita$  with and without restricting the inclusion of downstream municipalities.

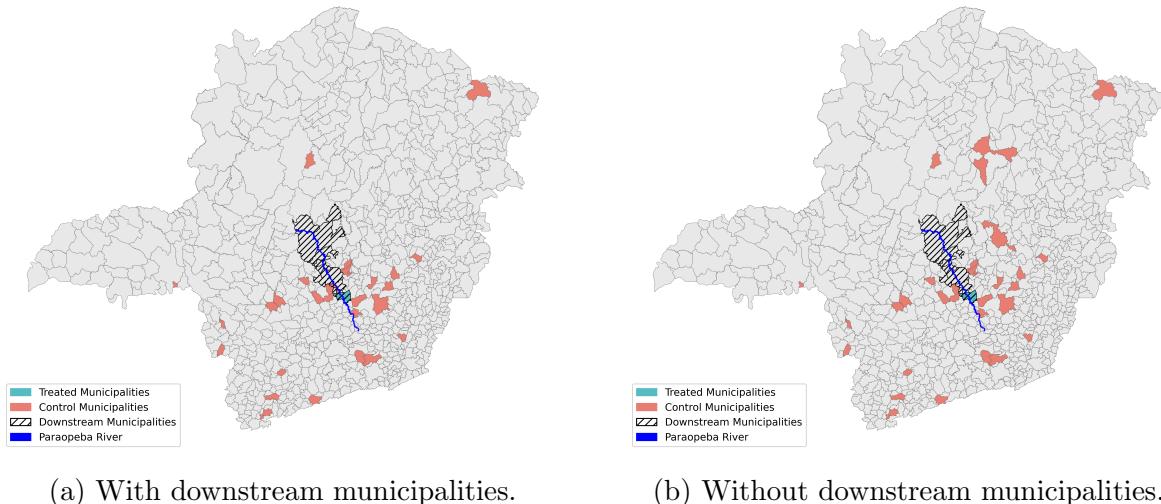
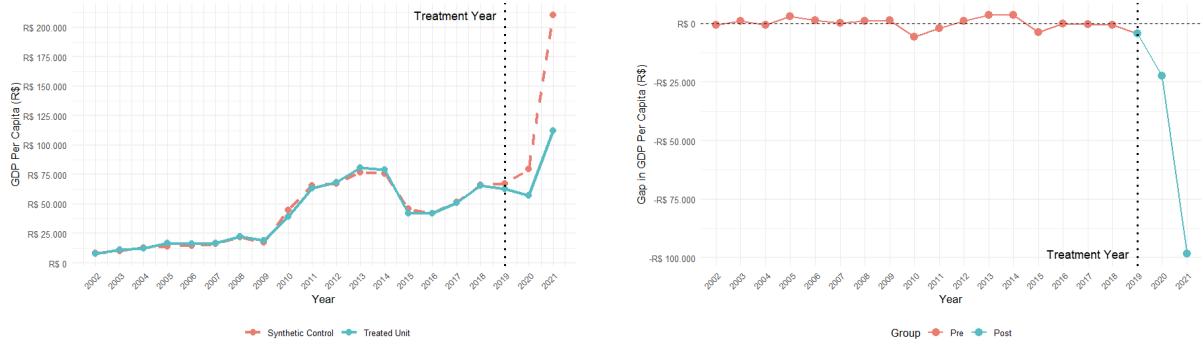


Figure 23: GDP Per Capita matched donor pool maps.

Figure 24a shows the time series for Brumadinho and the synthetic control created without downstream municipalities. Clearly, this graph is almost identical to Figure 3 from subsection 3.2, indicating that the results are robust to this restriction. Figure 24b plots the series of gaps between the synthetic control and the treated unit for this check. This graph further strengthens the hypothesis of robust results, once again being nearly identical to its predecessor in Figure 4.



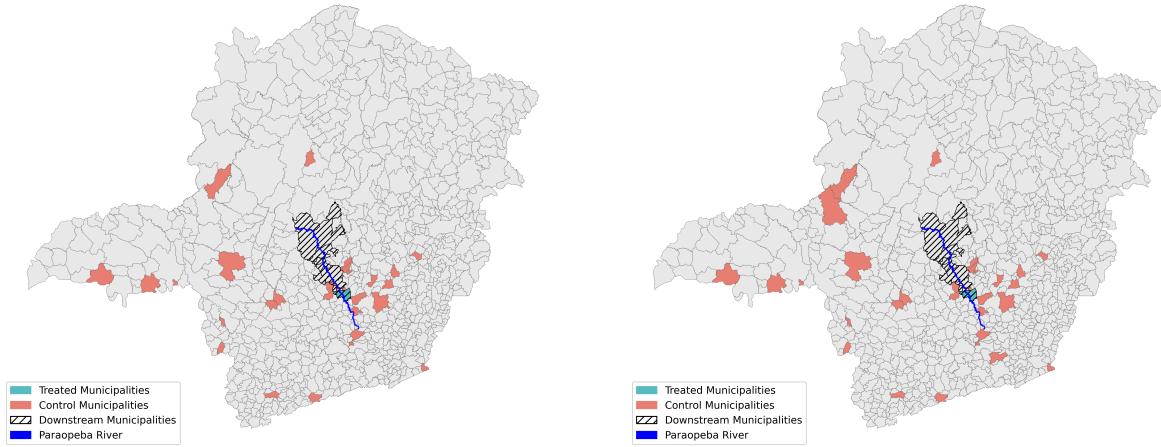
(a) GDP Per Capita trends.

(b) GDP Per Capita gaps.

Figure 24: GDP Per Capita trends and gaps no downstream.

These results indicate that the estimation from subsection 3.2 is robust to eliminating municipalities downstream of the Paraopeba River from the donor pool.

Figure 25 presents the map of the donor pools for  $Wage$  generated by the matching algorithm with and without the downstream restriction.



(a) With downstream municipalities.

(b) Without downstream municipalities.

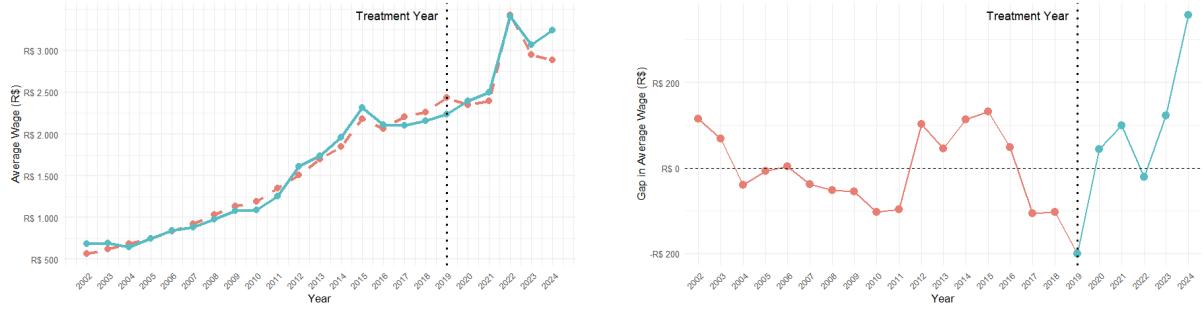
Figure 25: Wage matched donor pool maps.

Figure 26a displays the  $Wage$  trends for Brumadinho and the synthetic control estimated with the restriction on the downstream municipalities. There is very little difference between this graph and the original in Figure 7, which indicates robustness. Figure 26b shows the gap series with the restriction. The graph is very similar to the one in Figure 8, further displaying the robustness of these results.

These results lead to the conclusion that the estimation from subsection 3.3 is robust to the restriction on downstream municipalities.

Figure 27 has the donor pool maps created by the matching for the  $\log Wage$  with and without the restriction.

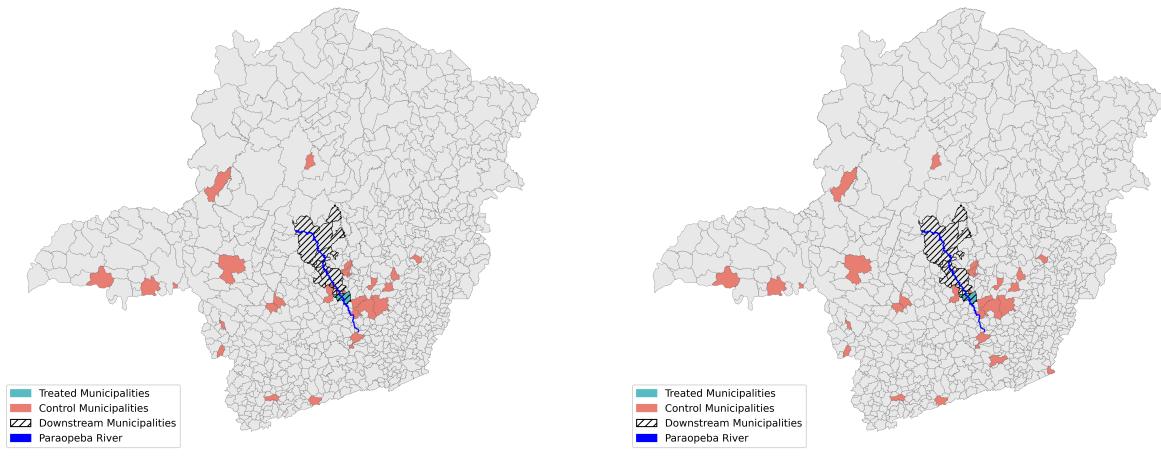
Figure 28a plots the trends for the treated unit and its synthetic control generated



(a) Wage trends.

(b) Wage gaps.

Figure 26: Wage trends and gaps no downstream.

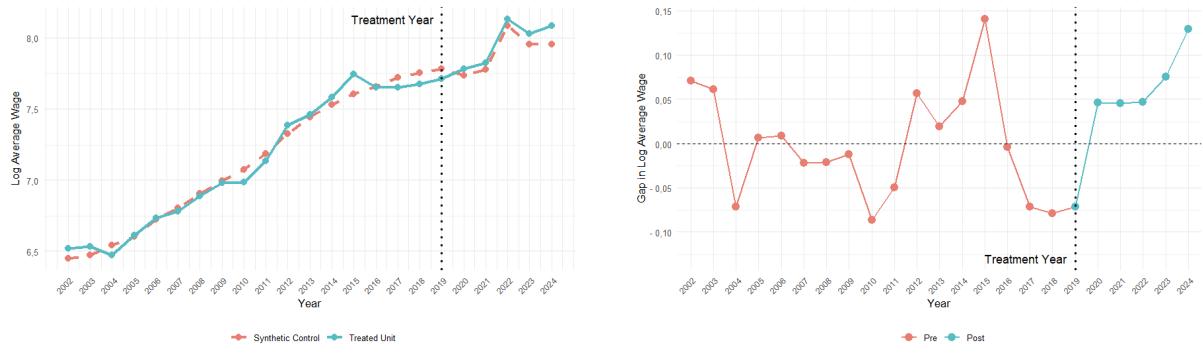


(a) With downstream municipalities.

(b) Without downstream municipalities.

Figure 27: Log Wage matched donor pool maps.

using the restricted donor pool. The graph is nearly identical to the one in Figure 11, which shows robustness of the estimation. Figure 28b presents the gaps with the restriction. When comparing this graph to the one in Figure 12 it is seen that they are very similar, which further indicates robustness.



(a) Log Wage trends.

(b) Log Wage gaps.

Figure 28: Log Wage trends and gaps no downstream.

This leads to the conclusion that the results from subsection 3.4 are robust to the downstream restriction.

Figure 29 contains the maps of the donor pools created by the matching algorithm, both with and without the use of the restriction on the inclusion of downstream municipalities.

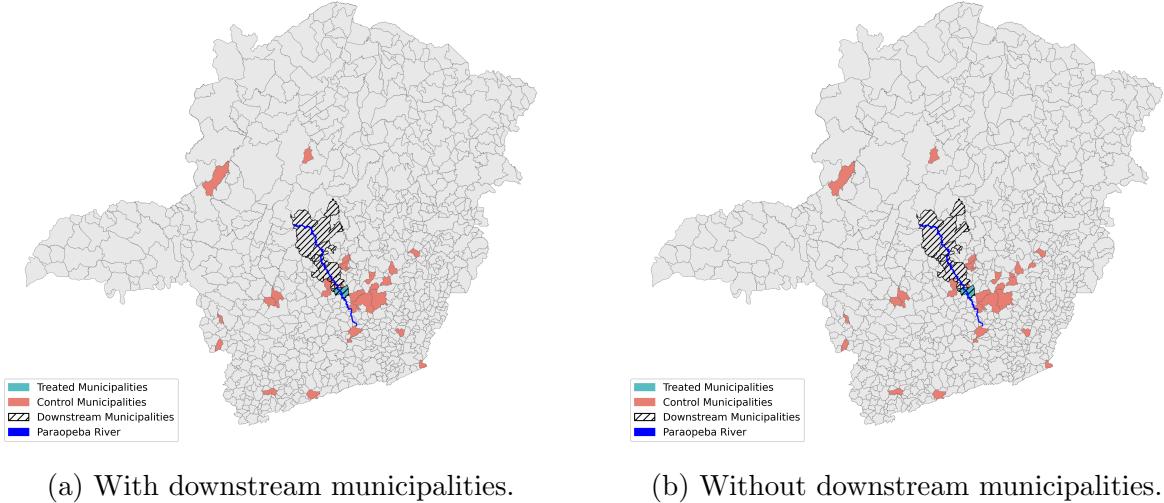


Figure 29: Proportion of Employment matched donor pool maps.

Figure 30a displays the trends for both Brumadinho and the restricted synthetic control side by side. The graph looks identical to the original one in Figure 15, with no visible difference between the two, which indicates that the results are robust. Figure 30b graphs the gaps between the treated unit and the synthetic control. This plot also looks identical to its predecessor, in Figure 16, providing more evidence that the results are robust to this check.

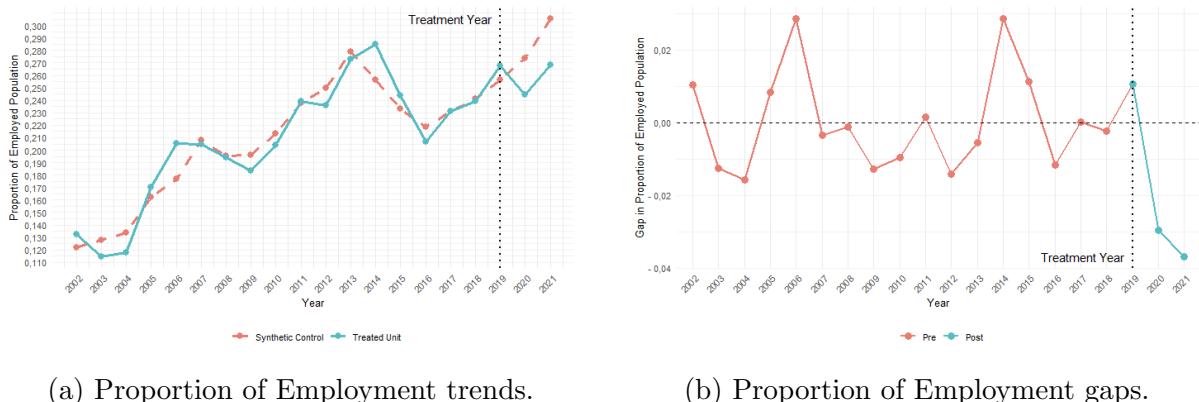


Figure 30: Proportion of Employment trends and gaps no downstream.

This leads to the conclusion that the results from subsection 3.5 are robust to the downstream restriction.

## 6 Conclusion

This paper analyzed the effects of the Brumadinho dam rupture disaster across 4 main interest variables  $GDPPerCapita$ ,  $Wage$ ,  $LogWage$  and  $PropEmployed$ , using a matching algorithm along with the synthetic control method and placebo tests for inference. It also evaluated heterogeneity of the effect on  $GDPPerCapita$  by performing the same analysis on a 4 sector breakdown of  $GVAPerCapita$ , Agriculture, Industry, Services and Public Administration.

The results show that there is a negative and statistically significant impact on the  $GDPPerCapita$  variable, with the effect being small in the first treatment year (2019) and increasing to more than R\$100.000,00 in the last year (2021). This result is in line with what was expected, since it is natural to imagine that a disaster of this magnitude would disrupt the local economy.

For the wage variables, two specifications were presented, one with the regular  $Wage$  and another using  $LogWage$ , a logarithmic transformation of the previous. For both, the resulting gaps are not statistically significant, meaning that there is not enough evidence to reject the null hypothesis of no treatment effect. Additionally, the post treatment trends of Brumadinho and its synthetic control for both of these variables follow a very similar path, indicating that this result is not an issue of low test power<sup>9</sup>, but that there actually was no effect.

For  $PropEmployed$  the fit was noisier than for the other variables, but the synthetic control still seemed to follow the trend of Brumadinho well. The post treatment calculated gaps were statistically insignificant, so that there was not enough evidence to reject the null hypothesis of no treatment effect from the disaster.

In the heterogeneity analysis, the impact on  $GDPPerCapita$  is broken down into the  $GVAPerCapita$  impact across 4 different economic sectors, Agriculture, Industry, Services and Public Administration. The analysis concluded that the Agriculture and Public Administration sectors did not present statistically significant gaps, with not enough evidence to reject the null hypothesis of no effect from the disaster on these sectors'  $GVAPerCapita$ .

The Industry and Services sectors presented statistically significant gaps, which show strong evidence that the disaster impacted them. The Industry was negatively affected throughout all treatment years, which is according to expectation since the dam rupture heavily impacted all mining activities which had an important share of the municipality's GDP. The Services gaps were negative in 2020 and 2021, but here was a large and positive estimated impact for the year of 2019. This last result is interesting and may provide insight into why the estimated negative gap on  $GDPPerCapita$  is small in 2019. A plausible theory for this is that the disaster generated the need for services, such as

---

<sup>9</sup>. Test power refers to the ability of a given test to reject the null hypothesis when it is false.

services related to the restoration of the area, which created a temporary increase in the sector's *GVA*. This theory could also explain why the impact on the Industry was low in 2019 compared to the other periods, because this sector could also have been stimulated by the need to rebuild after the disaster, with the construction sub-sector being heavily employed on this task.

In conclusion, *GDPPerCapita* appears to be the only interest variable affected, having had a negative impact from the accident, while the other variables *Wage*, *LogWage* and *PropEmployed* do not present significant effects. Additionally, the main economic sectors affected are Industry and Services, while Agriculture and Public Administration do not present statistically significant effects.

Despite the results being evaluated by statistical tests and robustness checks, it is important to recognize that these results rely on the standard assumption of validity of the synthetic control model, and the violation of these would result in invalid estimates. However, the only assumption that seems plausible to have been violated is that of common support, especially for the *PropEmployed* variable, but even for this interest variable the synthetic control appears to have a reasonable fit and the rendered results do not provide evidence of treatment effect, which is better than providing a miss identified impact.

## References

- Abadie, A. 2021. “Using synthetic controls: Feasibility, data requirements, and methodological aspects.” *Journal of economic literature* 59 (2): 391–425.
- Abadie, A., A. Diamond, and J. Hainmueller. 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): 493–505.
- . 2011. “Synth: An R package for synthetic control methods in comparative case studies.” *Journal of Statistical Software* 42:1–17.
- . 2015. “Comparative politics and the synthetic control method.” *American Journal of Political Science* 59 (2): 495–510.
- Abadie, A., and J. Gardeazabal. 2003. “The economic costs of conflict: A case study of the Basque Country.” *American economic review* 93 (1): 113–132.
- Andrade da Silva, M., C. Machado de Freitas, D. R. Xavier, and A. Rocha Romão. 2020. “Sobreposição de riscos e impactos no desastre da Vale em Brumadinho.” *Ciência e Cultura* 72 (2): 21–28. <https://doi.org/https://doi.org/10.1016/j.engfailanal.2021.105934>.
- Atlas do Desenvolvimento Humano no Brasil. 2025. *Municipal Data*. <http://www.atlasbrasil.org.br/consulta/planilha>. Accessed: 2025-08-17.
- Base dos Dados. 2025. *Data Portal*. Online database. Compiled data from RAIS and CAGED. Accessed October 6, 2025. <https://basedosdados.org>.
- Belasen, Ariel R, and Solomon W Polacheck. 2008. “How hurricanes affect wages and employment in local labor markets.” *American Economic Review* 98 (2): 49–53.
- Coffman, Makenna, and Ilan Noy. 2012. “Hurricane Iniki: measuring the long-term economic impact of a natural disaster using synthetic control.” *Environment and Development Economics* 17 (2): 187–205.
- IBGE. 2025. *Municipal Data*. <https://sidra.ibge.gov.br/home/>. Accessed: 2025-08-16.
- Jasi, A. 2019. “Brazil dam collapse prosecutors investigate collusion.” Accessed October 24, 2025. <https://www.thechemicalengineer.com/news/brazil-dam-collapse-prosecutors-investigate-collusion/>.

- Machado de Freitas, C., C. Barcellos, C. I. Rodrigues Fróes Asmus, M. Andrade da Silva, and D. R. Xavier. 2019. “Da Samarco em Mariana à Vale em Brumadinho: desastres em barragens de mineração e Saúde Coletiva.” *Cadernos de saúde pública* 35 (5): 21–28.
- Matsunaga, Liz. 2020. “Disasters and mental health: evidence from the Fundao tailing dam breach in Mariana, Brazil.” Master’s thesis, Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- Pereira, L. F., G. B. Cruz, and R. M. F. Guimarães. 2019. “Impactos do rompimento da barragem de rejeitos de Brumadinho, Brasil: uma análise baseada nas mudanças de cobertura da terra.” *Journal of Environmental Analysis and Progress* 4 (2): 122–129.
- Polignano, M. V., and R. S. Lemos. 2020. “Rompimento da barragem da Vale em Brumadinho: impactos socioambientais na Bacia do Rio Paraopeba.” *Ciência e Cultura* 72 (2): 37–43.
- Rodrigues, L. 2024. “Brumadinho dam collapse: 23 thousand compensation deals inked.” Accessed October 24, 2025. <https://agenciabrasil.ebc.com.br/en/geral/noticia/2024-01/brumadinho-dam-collapse-over-23-thousand-compensation-deals-forged>.
- Senra, R. 2019. “Brumadinho, a história de uma tragédia que poderia ter sido evitada.” Accessed October 24, 2025. <https://www.bbc.com/portuguese/brasil-47399659>.
- Vale. 2024. “Cerca de 80% da produção de minério de ferro da Vale no Brasil não utiliza barragens de rejeitos.” Accessed October 24, 2025. <https://vale.com/pt/w/cerca-de-80-porcento-da-producao-de-minerio-de-ferro-da-vale-no-brasil-nao-utiliza-barragens-de-rejeitos>.

## Appendix

### A Mean of Predictors for Brumadinho and Synthetic Control

Predictor	Brumadinho	Synthetic Control	Sample Mean
<i>GDPPerCapita</i>	38482.48	38428.46	22172.31
<i>POP</i>	34290.765	32610.965	29025.400
Agriculture's Share of GVA	0.018	0.031	0.044
Industry's Share of GVA	0.573	0.509	0.437
Services' Share of GVA	0.294	0.345	0.347
Public Admin.'s Share of GVA	0.116	0.115	0.172
Income <i>IDHM</i>	0.761	0.703	0.691
Long. <i>IDHM</i>	0.857	0.857	0.846
Educ. <i>IDHM</i>	0.639	0.634	0.621
Per Capita BF Transfers	53.338	55.786	69.061
<i>GINI</i>	0.572	0.444	0.469
Literacy Rate	94.100	94.126	92.757

Table A.1: GDP Per Capita estimation mean of predictors.

Predictor	Brumadinho	Synthetic Control	Sample Mean
<i>WAGE</i>	1346.42	1350.186	1260.059
<i>GDPPerCapita</i>	38482.48	33379.17	23898.43
<i>POP</i>	34290.76	34020.12	26351.83
Agriculture's <i>GVAPerCapita</i>	470.291	229.938	1095.627
Industry's <i>GVAPerCapita</i>	22131.02	18007.28	10614.8
Services' <i>GVAPerCapita</i>	10210.21	8979.386	6812.572
Public Admin.'s <i>GVAPerCapita</i>	3572.898	3157.424	2717.646
Agriculture's Share of GVA	0.018	0.011	0.063
Industry's Share of GVA	0.573	0.542	0.447
Services' Share of GVA	0.294	0.313	0.336
Public Admin.'s Share of GVA	0.116	0.135	0.155

Table A.2: Wage estimation mean of predictors (Part 1).

Predictor	Brumadinho	Synthetic Control	Sample Mean
Income <i>IDHM</i>	0.761	0.703	0.698
Long. <i>IDHM</i>	0.857	0.856	0.85
Educ. <i>IDHM</i>	0.639	0.65	0.624
Per Capita BF Transfers	53.338	67.58	67.766
<i>GINI</i>	0.572	0.486	0.474
Literacy Rate	94.1	94.189	93.38
Prop. of POP less 1 MW	0.269	0.295	0.298
Prop. of POP 1-2 MW	0.198	0.203	0.206
Prop. of POP 2 -5 MW	0.115	0.114	0.11
Prop. of POP 5-10 MW	0.034	0.022	0.021
Prop. of POP more 10 MW	0.027	0.007	0.006

Table A.3: Wage estimation mean of predictors (Part 2).

Predictor	Brumadinho	Synthetic Control	Sample Mean
<i>LogWAGE</i>	7.106824	7.109412	7.045529
<i>GDPPerCapita</i>	38482.48	34313.87	24563.46
<i>POP</i>	34290.76	55452.97	28355.51
Agriculture's <i>GVAPerCapita</i>	470.291	255.666	1090.744
Industry's <i>GVAPerCapita</i>	22131.02	19089.9	11190.85
Services' <i>GVAPerCapita</i>	10210.21	9598.019	6975.629
Public Admin.'s <i>GVAPerCapita</i>	3572.898	2918.932	2733.255
Agriculture's Share of GVA	0.018	0.013	0.063
Industry's Share of GVA	0.573	0.534	0.449
Services' Share of GVA	0.294	0.335	0.335
Public Admin.'s Share of GVA	0.116	0.118	0.153

Table A.4: Log Wage estimation mean of predictors (Part 1).

Predictor	Brumadinho	Synthetic Control	Sample Mean
Income <i>IDHM</i>	0.761	0.719	0.699
Long. <i>IDHM</i>	0.857	0.854	0.85
Educ. <i>IDHM</i>	0.639	0.67	0.626
Per Capita BF Transfers	53.338	62.213	67.384
<i>GINI</i>	0.572	0.494	0.475
Literacy Rate	94.1	94.529	93.493
Prop. of POP less 1 MW	0.269	0.287	0.295
Prop. of POP 1-2 MW	0.198	0.205	0.205
Prop. of POP 2 -5 MW	0.115	0.122	0.111
Prop. of POP 5-10 MW	0.034	0.029	0.022
Prop. of POP more 10 MW	0.027	0.01	0.006

Table A.5: Log Wage estimation mean of predictors (Part 2).

Predictor	Brumadinho	Synthetic Control	Sample Mean
<i>PropEmployed</i>	0.2047647	0.2047647	0.203
<i>GDPPerCapita</i>	38482.48	23630.81	25263.87
Agriculture's Share of GVA	0.018	0.028	0.04
Industry's Share of GVA	0.573	0.493	0.475
Services' Share of GVA	0.294	0.306	0.331
Public Admin.'s Share of GVA	0.116	0.173	0.154
Income <i>IDHM</i>	0.761	0.713	0.694
Long. <i>IDHM</i>	0.857	0.859	0.847
Educ. <i>IDHM</i>	0.639	0.634	0.631
Per Capita BF Transfers	53.338	41.859	65.568
<i>GINI</i>	0.572	0.469	0.468
Literacy Rate	94.1	94.524	93.55

Table A.6: Proportion of Employment estimation mean of predictors.

## B Synthetic Control weights

Municipality Name	Municipality IBGE Code <sup>1</sup>	Weight
Pirapora	3151206	0.499
Santos Dumont	3160702	0.230
Mateus Leme	3140704	0.117
Delta	3121258	0.111
Itatiaiuçu	3133709	0.036
Arcos	3104205	0.002
Juatuba	3136652	0.001
Nova Serrana	3145208	0.001
Pains	3146503	0.001
São José da Lapa	3162955	0.001

<sup>1</sup> Municipality identifier code used by IBGE (Brazilian Institute of Geography and Statistics).

Table B.1: GDP Per Capita weights (Only municipalities with at least 0.1% weight)

Municipality Name	Municipality IBGE Code <sup>1</sup>	Weight
Conceição das Alagoas	3117306	0.282
Rio Piracicaba	3155702	0.250
Delta	3121258	0.205
Pains	3146503	0.143
Carandaí	3113206	0.099
Itamonte	3133006	0.021

<sup>1</sup> Municipality identifier code used by IBGE (Brazilian Institute of Geography and Statistics).

Table B.2: Wage weights (Only municipalities with at least 0.1% weight)

Municipality Name	Municipality IBGE Code <sup>1</sup>	Weight
Conceição das Alagoas	3117306	0.343
São José da Lapa	3162955	0.336
Pedro Leopoldo	3149309	0.112
Carandaí	3113206	0.106
Sarzedo	3165537	0.102
Vazante	3171006	0.001

<sup>1</sup> Municipality identifier code used by IBGE (Brazilian Institute of Geography and Statistics).

Table B.3: Log Wage weights (Only municipalities with at least 0.1% weight)

Municipality Name	Municipality IBGE Code <sup>1</sup>	Weight
Visconde do Rio Branco	3172004	0.51
Itatiaiuçu	3133709	0.230
São Joaquim de Bicas	3162922	0.201
Pedro Leopoldo	3149309	0.058

<sup>1</sup> Municipality identifier code used by IBGE (Brazilian Institute of Geography and Statistics).

Table B.4: Proportion of Employment weights (Only municipalities with at least 0.1% weight)