Wage Shocks from the Brumadinho Tragedy

João Marrey Mendonça

2024

1 Introduction

On the 25th of January 2019, one of the biggest socioenvironmental catastrophes of Brazil happened: the collapse of the dam of the Córrego Feijão iron mine at the city of Brumadinho, Minas Gerais state (MG). The accident led to 272 deaths, a number approximately 18 times bigger than the Mariana disaster in 2015, a similar incident. More than 11.7 million m^3 of toxic mud raged through the region and contaminated the Paraopeba River.

Among the reasons that lead to the collapse is the structure of the dam. The Brumadinho dam was operated by the Brazilian mining company Vale S.A. since 1976, it was built as an upstream tailings dam, which is frailer if compared to centreline and downstream tailings dam design (Figure 1).

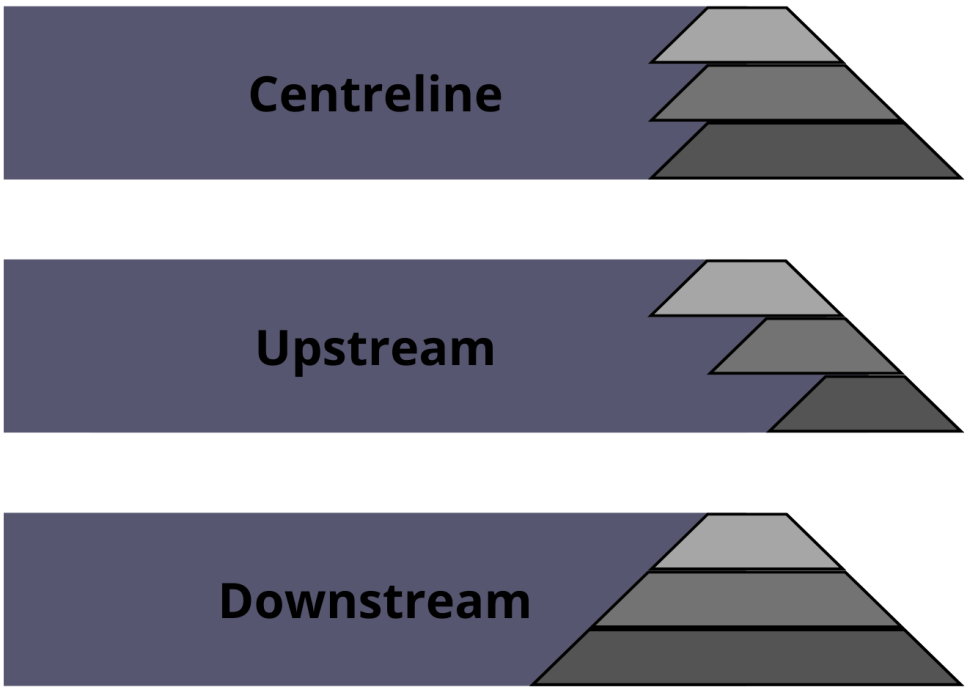

Figure 1: Tailing dams structures.

Besides, the BBC Brasil stated that since 2016 the dam had been without inspection by the ANM (Brazilian National Mining Agency). Furthermore, in the same year, there was a change in the internal inspection from the Tractebel to the Tüv Süd agency done by Vale after the former pointing out the risk of collapse, a different assessment than that of the new agency.

Due to the accident, and to make reparations, the company has compensated 15.4 thousand families and eliminated 13 other upstream dams which remain inactive and represent 40% of the dams which are supposed to be closed until the end of 2035 in Vale's De-characterization program, also adopting a better residual filtration system, decreasing its build up to up to 80%. Another measure was compromising to allocate 38 billion reais (BRL) in a judicial agreement for socioeconomic and environmental reparations in the region.

This work intends to analyze, through an econometric approach, the impact that the collapse had on the region's labor market of Brumadinho and other affected municipalities, focusing on the wage level. This approach is necessary because the net effect of the disaster on the wages is not obvious, despite it being expected that the accident impacted negatively on the wages at first, with the financial reparations and incentives given to families by Vale, the net effect is uncertain.

2 Literature Review

The impacts of this tragedy go much further beyond the scope of this paper. Previous literature indicates that 51% of the 297.28 hectares of environmental devastation used to be preserved forest area [Pereira et al., 2019]. In addition, the contamination of the Paraopeba River Basin exceeded by many times the threshold for toxic substances, such as the heavy metals Manganese and Cadmium [Polignano and Lemos, 2020]. There were more than 3845 people affected directly [Freitas et al., 2019], among which the very workers of Vale, which represented 91% of the deaths of the tragedy [Silva, 2020].

Works with a more empirical approach find relevant negative impacts for the affected population. General equilibrium models using simulations to try and predict the impact of the tragedy over the economy, found significant negative impacts on the short term [Da Silva et al., 2021].

This paper also takes big inspiration in the work of [Matsunaga, 2020], which tries to explore the effect of the Mariana dam collapse over mental health and finds an increase in the number of hospitalizations by mental health disorders in the affected population using a differences-in-differences framework. We use similar empirical models, but analyze a different disaster and the labor market instead of mental health disorders.

3 Data Base and Matching

We defined three different groups of affected municipalities, treated groups, against which we will test the results. This division is made so that we can consider the heterogeneous treatment effects between these groups, since the mud did not affect every location in the same manner and intensity, e.g. cities closer to the dam would receive stronger mud inflow. The groups are:

- **Group 1:** Composed only by the city of Brumadinho, which is the only place where the mud caused physical damage, besides the river contamination.
- **Group 2:** Cities downstream of the Paraopeba River which were closer to Brumadinho. Contains the municipalities of Betim, São Joaquim de Bicas, Igarapé and Mário Campos.
- **Group 3:** Cities downstream of the Paraopeba River which were further away from Brumadinho. Contains the municipalities of Juatuba, Florestal, Pará de Minas, Esmeraldas, São José da Varginha, Pequi, Fortuna de Minas, Maravilhas, Papagaios, Paraopeba, Pompéu and Curvelo.

To perform the matching between the municipalities and select the control group, cities with similar characteristics but that were not affect by the disaster, we used data from IBGE (Brazilian Geography and Statistics Institute) and DATASUS (a database for the unified Brazilian health care system). We consider variables such as population, GINI index and HDI, as well as economic variables such as GDP per capita, percentage of poor and extremely poor population and composition of the added value of the economy by sector (Figure 2).

Variables
Population (2014)
GDP per Capita (2014)
GINI Index (2010)
IDHM (2010)
IDHM Education (2010)
IDHM Longevity (2010)
IDHM Income (2010)
% Extremely Poor (2010)
% Poor (2010)
% Agriculture Added Value
% Industry Added Value
% Services Added Value
% Public Adm. Added Value

Table 1: Variables used in the matching process for cities similar to the affected ones.

Our matching approach was to create a model that searches for the nearest neighbors to each treated municipality (affected by the dam) by minimizing the differences in these variables between non treated municipalities in the same state. We opted out of using Propensity Score Matching (PSM) since one of the defined treatment groups contains only a single treated unity, which leads to statistical limitations in the use of a probabilistic regression.

Analyzing the resulting control group generated by the matching for Brumadinho, we find that most of the control municipalities are also mining cities, with the company Vale operating in them. That is a good indication that the matching process is valid, because it has returned cities with similar socioeconomic characteristics to the treated.

In this work, we have applied the matching process separately for each treated city, choosing the 5 most similar municipalities to the treated one, according to the matching algorithm, as its control group. In case two or more municipalities of the same group share a common city in their control groups, this city is only used once as a counterfactual, not being repeated on the database.

4 Model

For each group, we use differences-in-differences models to evaluate the impact of the disaster on our interest variable, represented by the log of the average monthly wage for the month of December of each year. We use data from the RAIS, a public database containing extremely detailed socioeconomic data for the entirety of Brazil, for the period between 2015 and 2022 for our analysis. Four specifications are used:

1. **Homogeneous Effect:** Treatment effect constant in time

$$Y_{it} = \beta + \alpha_i + \alpha_t + \phi + e_{it} \tag{1}$$

2. **Post Treatment Heterogeneous Effect:** Treatment effect varying in time 2019-2022

$$Y_{it} = \beta + \alpha_i + \alpha_t + \phi_{2019} + \phi_{2020} + \phi_{2021} + \phi_{2022} + e_{it} \tag{2}$$

3. **Pre and Post Treatment Heterogeneous Effect:** Treatment effect varying in time 2019-2022 and dummies for placebo test 2015-2017

$$Y_{it} = \beta + \alpha_i + \alpha_t + \phi_{2015} + \phi_{2016} + \phi_{2017} + \phi_{2019} + \phi_{2020} + \phi_{2021} + \phi_{2022} + e_{it} \tag{3}$$

4. **Callaway Sant'Anna:** Here we make an additional robustness test using the [Callaway and Sant'Anna, 2021] estimator. Similarly to the third specification, the treatment coefficients for 2019 through to 2022 are reported, as well as the placebo tests. However, the test coefficients span from 2016 through to 2018, with the base year, i.e. year with the treated dummy omitted to avoid the "dummy trap", being 2015, while on the third specification it was 2018.

$$Y_{it} = \beta + \alpha_i + \alpha_t + \phi_{2016} + \phi_{2017} + \phi_{2018} + \phi_{2019} + \phi_{2020} + \phi_{2021} + \phi_{2022} + e_{it} \tag{4}$$

- i - is the municipality index.
- t - is the time index.
- Y_{it} - represents the interest variable, which is the log of the monthly average wage at each year for the month of December.
- β - represents the intercept of the regression.
- α_i - represents the municipality fixed effect.
- α_t - represents the time fixed effect.
- ϕ - represents the treatment effect, i.e. being affected by the accident.
- ϕ_{2015} - represents the treatment effect in 2015.
- ϕ_{2016} - represents the treatment effect in 2016.
- ϕ_{2017} - represents the treatment effect in 2017.

- ϕ_{2018} - represents the treatment effect in 2018.
- ϕ_{2019} - represents the treatment effect in 2019.
- ϕ_{2020} - represents the treatment effect in 2020.
- ϕ_{2021} - represents the treatment effect in 2021.
- ϕ_{2022} - represents the treatment effect in 2022.
- e_{it} - represents the error term.

5 Results

We display the results for every group and for each specification of the differences-in-differences model.

5.1 Group 1

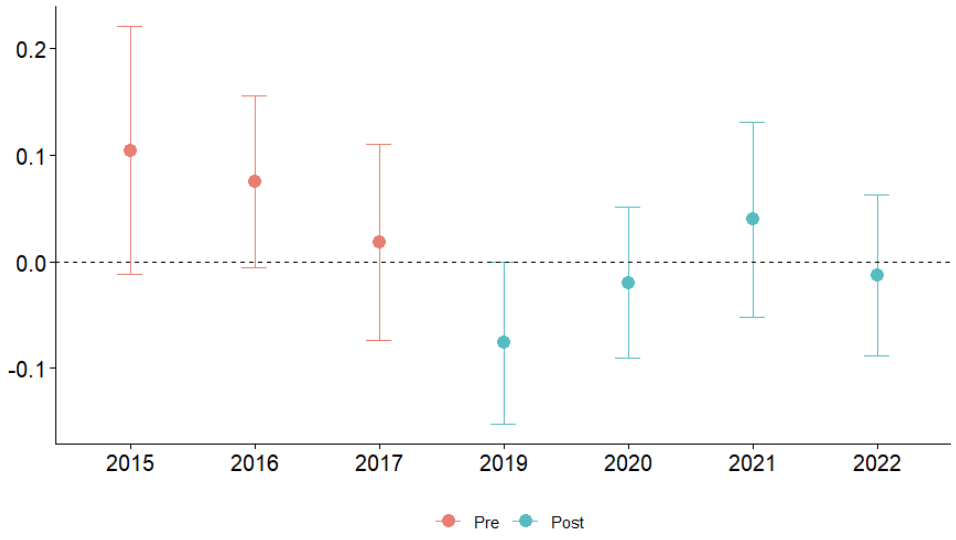

Figure 2: Estimated coefficients for Group 1 & Specification 3

The first group contains only the city of Brumadinho, the most affected by the accident.

In the first regression specification, we see a negative effect over the wages, but it is insignificant at 5%. It is Worth noting that this is not the best specification, because it does not allow for the treatment effect to be heterogeneous through time.

In the second specification, the impact is still negative, remaining statistically significant for the years of 2019 and 2020, but turning insignificant in 2021 and 2022, with the peak of the effect in 2019. These results are in agreement with what was expected, since the impact is much greater in the year of 2019 and seem to decay throughout time. At

Variable	**(1) Homogeneous Effect**	**(2) Heterogeneous Post-Disaster**	**(3) Heterogeneous Pre and Post-Disaster**	**(4) Callaway Sant'Anna**
ϕ	-0.067027* (0.034632)	-	-	-
ϕ_{2015}	-	-	0.104564* (0.059531)	-
ϕ_{2016}	-	-	0.075317* (0.041148)	-0.0292 (0.0572)
ϕ_{2017}	-	-	0.018609 (0.046944)	-0.0567 (0.0392)
ϕ_{2018}	-	-	-	-0.0186 (0.0306)
ϕ_{2019}	-	-0.126005*** (0.032256)	-0.076383** (0.038872)	-0.0764*** (0.0259)
ϕ_{2020}	-	-0.069234** (0.029177)	-0.019612 (0.036078)	-0.0196 (0.0243)
ϕ_{2021}	-	-0.010196 (0.040665)	0.039426 (0.046796)	-0.0394 (0.0543)
ϕ_{2022}	-	-0.062672* (0.032005)	-0.013049 (0.038641)	-0.0130 (0.0410)
Observations	48	48	48	48
R^2	0.948524	0.951168	0.953957	-
Adjusted R^2	0.928842	0.925964	0.922714	-

Table 2: Results for Group 1 (Brumadinho). Homogeneous and Heterogeneous Effects.

the appendix, in Figure A.1, it is possible to visualize a graph displaying the coefficients together with a 95% confidence interval.

In the third specification, we see that 2019 remains a statistically significant negative impact, but the estimated effect is smaller than on the second specification. Besides, the treatment effect stops being significant in 2020, possibly due to the loss of degrees of freedom, due to a greater number of coefficients. The main result for this model, however, is that the coefficients before 2019, the year of the accident, are all insignificant at 5% levels, indicating that the hypothesis of parallel tendencies is being respected. We can visualize the evolution of the coefficients in Figure 2.

Lastly, we report the Callaway Sant'Anna specification as a robustness test for the analysis. We can see that it gets to very similar results to those of the third specification, with the effect in 2019 being even more significant than before, and the placebo test indicating the existence of parallel tendencies. In the appendix, it is possible to see a graph with the results on Figure A.2, indicating that the only statistically significant coefficient is the one from 2019.

It is important to note that the preferred specification is the second one, because the parallel tendencies hypothesis appears to be respected, as shown by the insignificance of the pretreatment dummy coefficients in the third and fourth specifications. Thus, we can take the placebo test out of the regression, gaining degrees of freedom while maintaining our trust on the non-bias of the estimators.

5.2 Group 2

Variable	**(1) Homogeneous Effect**	**(2) Heterogeneous Post-Disaster**	**(3) Heterogeneous Pre and Post-Disaster**	**(4) Callaway Sant'Anna**
ϕ	-0.030624 (0.036593)	-	-	-
ϕ_{2015}	-	-	0.002515 (0.042928)	-
ϕ_{2016}	-	-	-0.038329 (0.036845)	-0.0408 (0.0233)
ϕ_{2017}	-	-	-0.014041 (0.028470)	0.0243 (0.0216)
ϕ_{2018}	-	-	-	0.0140 (0.0141)
ϕ_{2019}	-	-0.009051 (0.037644)	-0.021515 (0.040596)	-0.0215 (0.0296)
ϕ_{2020}	-	-0.013784 (0.030011)	-0.026248 (0.033466)	-0.0262 (0.0246)
ϕ_{2021}	-	-0.005298 (0.116450)	-0.017762 (0.118598)	-0.0178 (0.1382)
ϕ_{2022}	-	-0.094362* (0.049009)	-0.106825** (0.051527)	-0.1068** (0.0410)
Observations	168	168	168	168
R^2	0.916781	0.918906	0.919316	-
Adjusted R^2	0.900017	0.900422	0.898690	-

Table 3: Results for Group 2 (Betim, São Joaquim de Bicas, Igarapé, Mário Campos).

The second group contains cities downstream of the river that were closer to the place of the accident.

We can see that the coefficients of the first and second specifications are not statistically significant, indicating that there was no impact on the log of the wages post disaster. In the appendix Figure A.3 we are able to visualize the coefficients for the second specification, seeing that the 0 value axis is inside of the confidence interval for all of them.

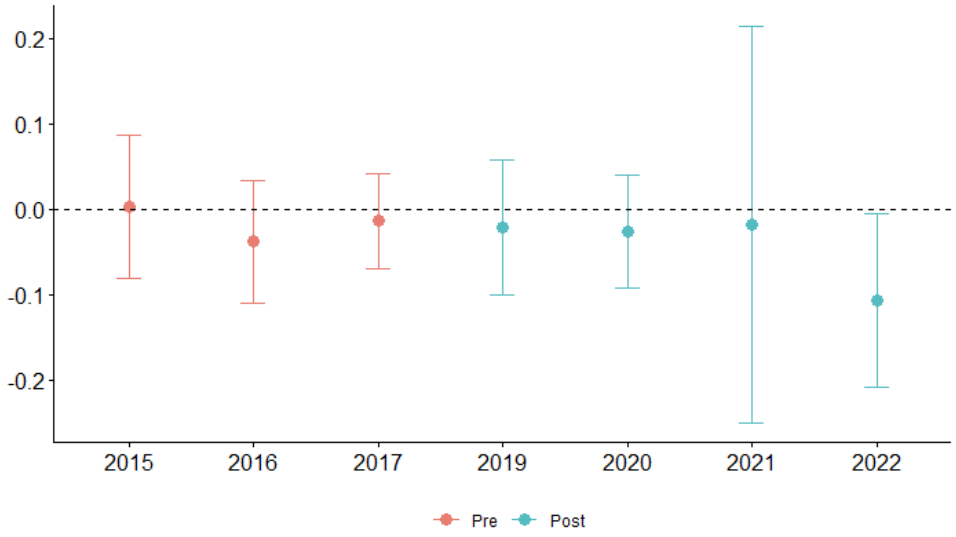

Figure 3: Estimated coefficients for Group 2 & Specification 3

In the third specification, we observed that the parallel tendencies also seem to be valid. The only significant coefficient is the one for 2022, that is unexpected, since there seems to be an effect for this year and not for 2019, however it is probably related to the greater instability of the estimators due to the loss of degrees of freedom if compared to the second specification. In Figure 3 we can analyze the coefficients and that only 2022 has a significant effect at 5%.

In the fourth specification we use the Callaway Sant'Anna estimator to verify the robustness of the results. We obtain almost the exact same conclusions as the third specification: the parallel tendencies seem to be respected and only the coefficient for 2022 is significant at 5%. In the appendix Figure A.4 we have a graph of the coefficients with their confidence intervals. Notice that it is very similar to the one of the third specification.

As the parallel tendencies seem to be respected, the preferred specification becomes again the second one. With that, we conclude that there were no statistically significant effects over the second group, related to the collapse of the dam.

5.3 Group 3

The third group contains cities downstream of the river that were far to the place of the accident.

In the model of the first specification, we see that the estimated coefficient is statistically insignificant at 5%. In the second specification, we relax the hypothesis of treatment constant in time, however, all of the coefficients are not only statistically insignificant, but also the ones from 2019 through to 2021 are also positive, suggesting that they are likely null in reality. In the appendix Figure A.5 it is possible to see a graph of the coefficients of the second specification, notice how all are very close to zero.

In the third specification the parallel tendencies hypothesis seems to be valid and all coefficients continue to be statistically insignificant at 5%. It is possible to visualize the coefficients in Figure 4.

Variable	**(1) Homogeneous Effect**	**(2) Heterogeneous Post-Disaster**	**(3) Heterogeneous Pre and Post-Disaster**	**(4) Callaway Sant'Anna**
ϕ	-0.001845 (0.017020)	-	-	-
ϕ_{2015}	-	-	-0.035515 (0.032130)	-
ϕ_{2016}	-	-	-0.026895 (0.026551)	0.0086 (0.0318)
ϕ_{2017}	-	-	-0.019906 (0.022496)	0.0070 (0.0182)
ϕ_{2018}	-	-	-	0.0199 (0.0257)
ϕ_{2019}	-	0.012084 (0.020090)	-0.008495 (0.022879)	-0.0085 (0.0136)
ϕ_{2020}	-	0.000136 (0.020843)	-0.020443 (0.023547)	-0.0204 (0.0207)
ϕ_{2021}	-	0.009776 (0.040179)	-0.010803 (0.041738)	-0.0108 (0.0510)
ϕ_{2022}	-	-0.029376 (0.027895)	-0.049955* (0.030007)	-0.0500* (0.0285)
Observations	544	544	544	544
R^2	0.917183	0.917569	0.917812	-
Adjusted R^2	0.903911	0.903742	0.903402	-

Table 4: Results for Group 3 (Juatuba, Florestal, Pará de Minas, etc.).

In the fourth specification we use the Callaway Sant'Anna estimator once again and obtain similar results to the third specification, with parallel tendencies and all of the coefficients statistically insignificant. In the appendix Figure A.6 we can observe the graph of these coefficients.

As in the other groups, the preferred specification is the second, since the parallel tendencies hypothesis seems to be respected. Here the effect appears to be null, with most of the periods even reporting a positive estimated effect, however insignificant.

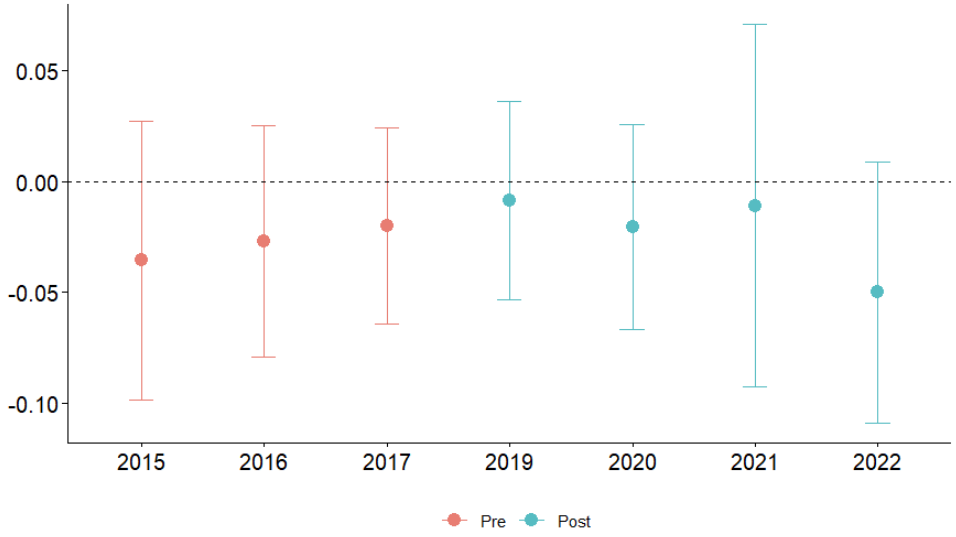

Figure 4: Estimated coefficients for Group 3 & Specification 3

6 Conclusion

From the four different specifications and three groups, we can have a clearer view over the impacts of the Brumadinho disaster in the labor market of the affected region. We find a negative effect from the disaster over the log of the average wages in the city of Brumadinho, it indicates a fall of 12.6% of the average monthly wage in 2019, with the gradual dissipation of this effect in the following years.

For the other treatment groups, which include other affected municipalities, but in a lower degree, we do not find any significant effect at 5% confidence level in the second specification. Thus, we are unable to identify any significant impact on the log of the average wages of these municipalities.

It is also important to note that further robustness tests performed, such as the placebo tests and the use of the Callaway Sant'Anna estimator, seem to indicate that the parallel tendencies are being respected and the results are robust.

7 References

References

[Callaway and Sant'Anna, 2021] Callaway, B. and Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

[Da Silva et al., 2021] Da Silva, C. E. S. F., Wolf, R., Oliveira, T. D., de Mello Valério, V. E., and Perobelli, F. (2021). The brumadinho dam rupture disaster in minas gerais (brazil) and the productivity: regional economic impacts using a computable general equilibrium approach.

[Freitas et al., 2019] Freitas, C. et al. (2019). Da samarco em mariana à vale em brumadinho: desastres em barragens de mineração e saúde coletiva. *Cienc. Cult.*

[Matsunaga, 2020] Matsunaga, L. (2020). Disasters and mental health: evidence from the fundao tailing dam breach in mariana, brazil. Master's thesis, Faculdade de Economia, Administração e Contabilidade, University of São Paulo, São Paulo.

[Pereira et al., 2019] Pereira, L., de B. Cruz, G., and Guimarães, R. F. (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 and Lemos, 2020] Polignano, M. V. and Lemos, R. S. (2020). Rompimento da barragem da vale em brumadinho: impactos socioambientais na bacia do rio paraopeba. *Cienc. Cult.*, 72(2):37–43.

[Silva, 2020] Silva, M. (2020). Sobreposição de riscos e impactos no desastre da vale em brumadinho.

A Appendix

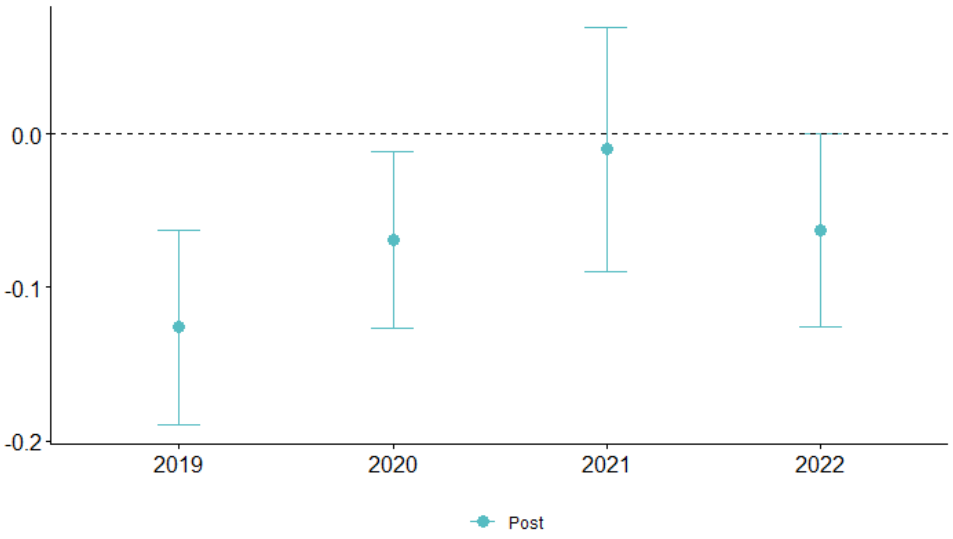

Figure A.1: Estimated coefficients for Group 1 & Specification 2

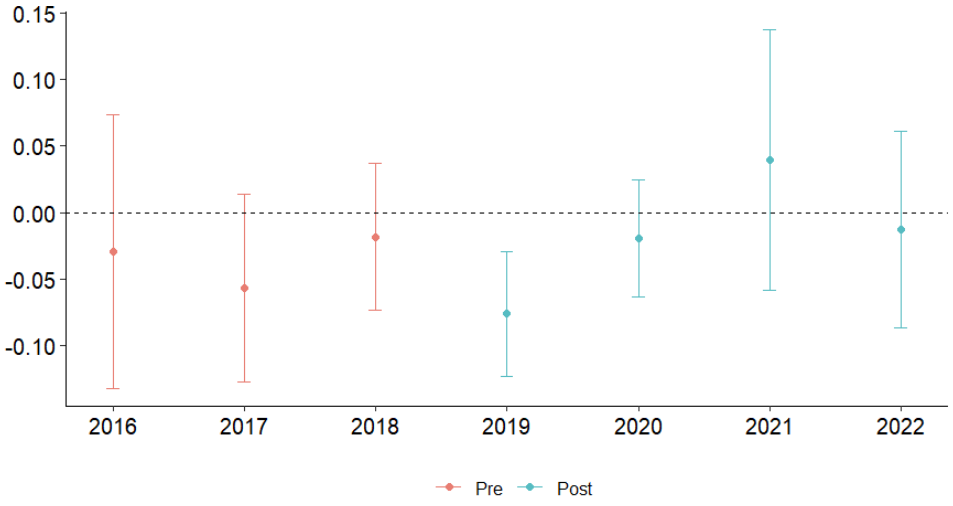

Figure A.2: Estimated coefficients for Group 1 & Specification 4

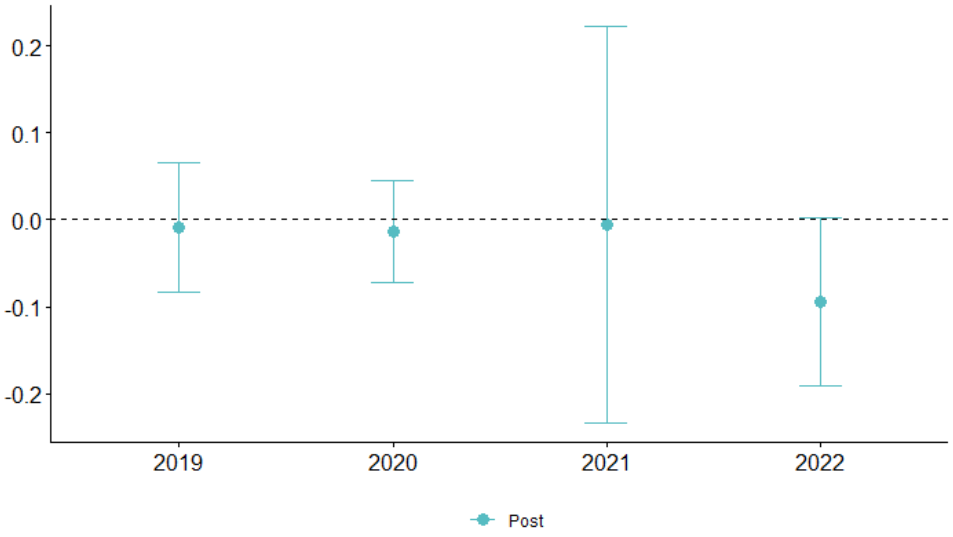

Figure A.3: Estimated coefficients for Group 2 & Specification 2

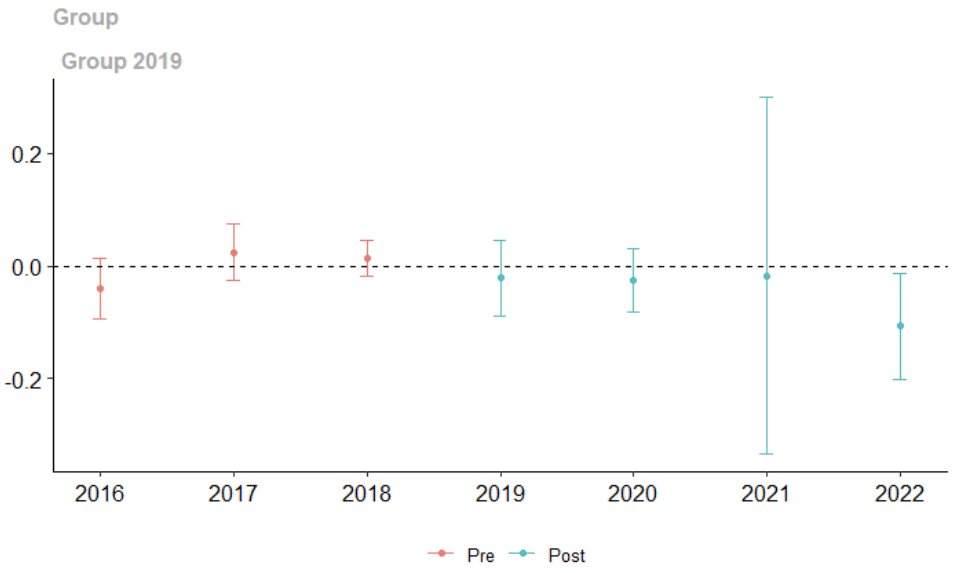

Figure A.4: Estimated coefficients for Group 2 & Specification 4

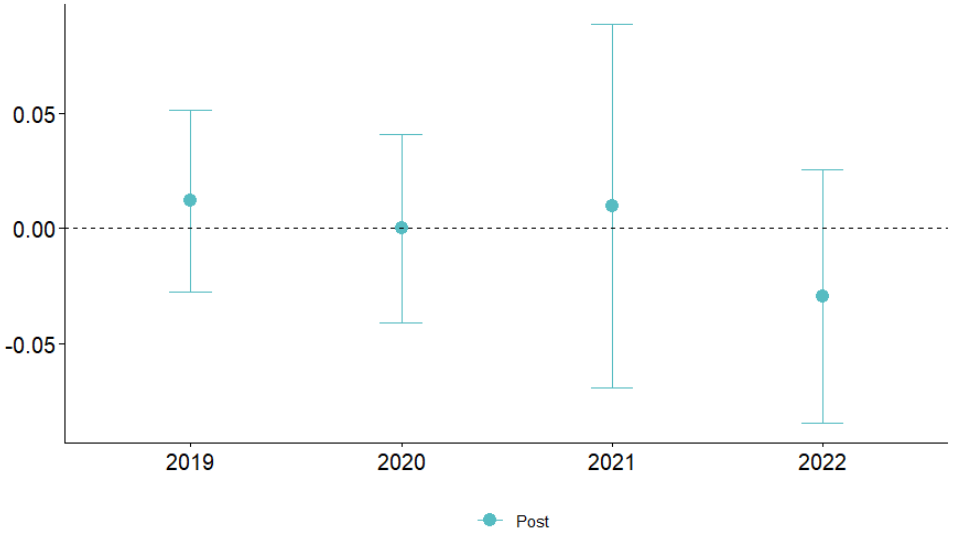

Figure A.5: Estimated coefficients for Group 3 & Specification 2

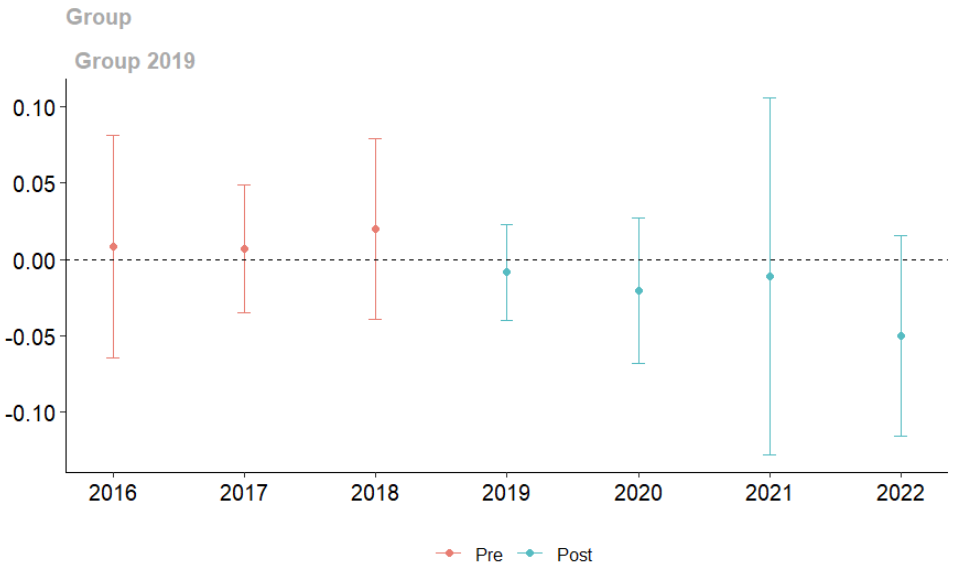

Figure A.6: Estimated coefficients for Group 3 & Specification 4