Universitat
Pompeu Fabra
Barcelona

**Faculty of Economics
and Business**

Academic year: **2024–2025**

Code: **SWE06**

Final Year Project

Market Access and Regional Inequality: Evidence from Brazil's Capital Relocation

Joan Costa Rodriguez (Degree in Economics)

Tutored by Professor Bruno Conte Leite

Acknowledgments

I would like to thank my tutor, Dr. Bruno Conte, for his guidance and support throughout the thesis, for responding to my questions, and for motivating me to continue my studies in economics at a graduate level. I would also like to thank both Dr. Melanie Morten from Stanford University and Dr. Jacqueline Oliveira from Rhodes University for their generosity in providing and sharing the data on travel time across municipalities for the computation of trade costs.

Abstract

This paper examines the impact of market access on regional inequality, using the relocation of Brazil's capital to Brasília as a natural experiment. The capital's relocation was followed by a substantial expansion of infrastructure, particularly highways, which reduced trade costs and enhanced market access throughout all the regions in the country. This study measures the effects of the increased market access across all Brazilian states, with particular attention to differences between rural and industrial regions. Market access and inequality indices are constructed using official data sources. The findings indicate that increases in market access are associated with an increase in regional inequality in industrial regions, while the effect is significantly attenuated in rural areas.

Keywords: Income Inequality, Market Access, Development, Regional Inequality

Contents

List of Figures

List of Tables

1 Introduction

The use of infrastructure construction as a tool for regional development has been a long-standing argument among economists throughout history. This thesis does not dispute the role of infrastructure on growth but seeks to examine more deeply how such investments affect different municipalities where the infrastructure is implemented. The primary objective is to determine whether the reduction of trade costs through highway construction decreases or increases inequality across regions within a state.

To address this question, the thesis takes advantage of the relocation of Brazil's capital city to Brasília in 1960 as a natural experiment. This event serves as an ideal setting to analyse how regions that previously had limited access to goods from other areas experienced significant changes following the major expansion of infrastructure and the adoption of a radial transportation network that was carried out from the 1960's onwards. The natural experiment framework facilitates the identification of causal effects, resembling a difference-in-differences estimation, where changes in market access across states capture the true impact of infrastructure construction.

The concept of market access has gained prominence in recent years, particularly following the influential work of Donaldson and Hornbeck (2016), who developed the variable as a means to measure the effect of infrastructure and other economic factors on trade costs and within-country trade. Other authors, such as Morten and Oliveira (2024) and Astorga-Rojas (2024), have further explored this concept in the context of Brazil and have significantly caused influence in the framework of this study. However, the aim of this project differs from previous research by focusing and contributing to previous literature specifically on the relationship between market access and regional inequality.

To ensure the robustness of the results, various inequality indices are employed, all constructed using census data from official sources. The market access variable is developed based on the trade costs obtained from the lowest travel time routes between state capitals, as it is done in Morten and Oliveira (2024). The analysis proceeds by estimating a range of models: starting with baseline specifications, then adding control variables, instrumenting market access, and finally introducing an interaction term to distinguish between rural and industrial regions in Brazil, in order to assess whether the effects differ across these areas.

The study is organized as follows. Section 2 reviews the relevant literature and discusses the various approaches that have been taken to address similar cases. Section 3 describes the data sources and details the construction of the primary variables used to measure market access and the inequality indices. Section 4 outlines the methodology employed for constructing the main variables and presents the different regression models estimated. Section 5 develops and discusses the results of the models and provides key insights. Finally, Section 6 concludes the thesis.

2 Conceptual Framework

The main goal of the following theoretical subsection will be to explore the previous literature on market access and income inequality between regions and to look at both the results and the approximations that different authors have achieved to look for causalities and similarities of the variables with other factors.

2.1 Literature Review

Previous literature related to this paper can be divided into two distinct parts, each reflecting different scopes of what this article aims to conclude. First, those that study market access as a variable at different levels, starting at a municipal level with only inter-country trade and ending at national level studies that also consider external trade.

The market access analysed in this paper is constructed at the state level, offering a different perspective from Donaldson and Hornbeck (2016), who compute market access at the county level in the United States, a methodology similar to that of Astorga-Rojas (2024), who applies it at the MCA (Minimum Comparable Area) level in Brazil. By working at a more disaggregated level, the number of observations increases and the resulting estimates may be more consistent, as bilateral trade costs are captured more comprehensively, incorporating richer information on the costs of trade across all MCAs. However, their approaches differ in important ways: Donaldson and Hornbeck focus on the impact of market access on agricultural land values and population distribution, computing market access weighted by population, whereas Astorga-Rojas examines how market access contributes to agricultural productivity and the adoption of key technologies, constructing the variable using the GDP share of each geospatial unit.

This thesis adopts both definitions of market access established in both studies to obtain robust results while working at a state level (constructing one variable with population and another with the share of GDP). Both papers find that market access positively affects agricultural productivity and land value. To strengthen their analyses, both apply different instrumental variable strategies, but reach different conclusions after instrumenting market access. Astorga-Rojas (2024) instruments market access to emphasize the natural experiment at the core of his analysis, constructing a market access variable that assumes no roads existed prior to the infrastructure built around Brasília. The strength and statistical relevance of the instrument increases over time, resulting in a higher estimated elasticity of the outcome variable with respect to market access. In contrast, Donaldson and Hornbeck (2016) instrument market access using waterways to create a water-based market access measure. While this approach appears reasonable, the considerably bigger magnitudes they obtain may indicate a possible violation of the IV identification assumptions. The assumption made in their paper is that counties farther from natural waterways may have experienced greater increases in agricultural land value due to factors unrelated to market access. Although the IV estimates do not reject a strong effect of market access on agricultural land value, a greater emphasis is placed on empirical strategies that control for local shocks to land values. In this paper, a similar approach to the one shown by Donaldson and Hornbeck is achieved using the kilometers of railways for each state across time as an instrument.

On the causality of market access and infrastructure construction on growth and income, Donaldson (2018) examines the effects of railroad construction in India and makes significant contributions to the field. His study demonstrates that the expansion of the railroad network led to a reduction in trading costs, a decrease in inter-regional price gaps, and an increase in trade volumes. When the railroad network reached the average district, real agricultural income in that district rose by approximately 16%. This evidence shows that railroads contributed to increased real income, a factor relevant to the results of this work. For further details, see also Donaldson and Burgess (2012).

A similar approach is seen in Morten and Oliveira (2024), from which this paper draws significant inspiration, especially in the computation of trade costs. Morten and Oliveira compute travel costs through an estimation of the time it takes to travel from one place to another via the lowest freight route. This mechanism is used in this paper, not considering, as in de Sousa et al. (2012), any border-related costs between states. Sousa et al., in their study, compute trade costs slightly differently by considering border state costs and construct their bilateral trade cost through: $\tau_{ij} \equiv d_{ij}^{\delta}(1 + \mathrm{brc}_{ij})u_{ij}$ where brc_{ij} are the border-related costs and u_{ij} the unobserved determinants. The different methods of computing trade costs and market access are important considerations when comparing papers in this field of study, as there is no objective or determined way to compute certain predictors, which may cause slight differences when comparing the results obtained by different methodologies.

The second strand of literature central to this thesis concerns the use of inequality indices to measure disparities across regions. While most of the inequality measures employed in this thesis are consistent with those used by Magalhães and Alves (2022), there is a much broader literature underpinning these indices. In their study, Magalhaes and Alves compute inequality indices both at the national level across all MCAs and at the macro-regional level. Their findings reveal a marked increase in regional inequality at both the national and macro-regional scales during the 1960's, with all indices peaking in the early 1970's, a period characterized by unsustainable regional disparities. This was followed by a significant decline in inequality throughout the 1980's. Although the present study focuses on regional inequality at the state level, similar patterns can be observed, reflecting the broader trends documented in the literature.

The implementation of regional development policies, such as the creation of the Superintendency for the Development of the Northeast in 1959 and the Superintendency for the

Development of the Amazon in 1966—whose effects are examined in greater depth by Acker (2021)—played a crucial role in this process. Alongside these targeted policies, broader federal initiatives, the expansion of national transportation and telecommunications infrastructure, the development of capital markets, and the relocation of the federal capital to Brasília all contributed to the emergence of new economic activities across the country. Collectively, these developments fostered regional convergence and a reduction in inequality indices during the 1980's—a trend observed not only at the national and macro-regional levels, as documented by Magalhães and Alves (2022), but also at the state level, as demonstrated in this paper.

Other authors, such as Azzoni (2001), have also examined regional inequality patterns in Brazil. In his analysis spanning 1940–2000, Azzoni investigates the relationship between regional disparities and economic growth using two key metrics: the Theil index and Coefficient of variation to measure per capita income dispersion. Rather than focusing only on national-level data, he aggregates Brazil into five macro-regions (North, Northeast, Southeast, South, and Center-West) while also maintaining the state-level analysis. The study distinguishes between conditional and absolute convergence, revealing that periods of rapid GDP per capita growth correlate strongly with rising regional inequality. Crucially, Azzoni demonstrates that the rate of inequality change depends on national income growth speed, faster economic expansion accelerates either regional divergence. His findings suggest evidence of regional convergence dynamics, though, these are modulated by national growth patterns. His study shows to be very useful for a comparison with the state inequality levels achieved.

3 Data

Historical data available from IPEA (Instituto de Pesquisa Econômica Aplicada) and IBGE (Instituto Brasileiro de Geografia e Estatística) on population per municipality and GDP per municipality is used to compute GDP per capita and all the inequality indices. The raw GDP data is available for the years 1939, 1950, 1960, 1970, 1975, 1980, 1985, and 1996, while the population data is available for 1940, 1950, 1960, 1970, 1976, 1980, 1985, and 1992. These data points are interpolated to obtain continuous observations, allowing for a proper analysis.

All the GDP data is in 2010 Brazilian reals (Reais) to facilitate the comparison and analysis of the data.

As Brazil experienced a major increase in the number of municipalities during its 20th century development, the unit of analysis for the computation of the inequality indices are the minimum comparable areas (MCAs). The MCAs used are those developed in Ehrl (2017). Minimum comparable areas are geospatial units that are time-consistent, created to enable comparison between any arbitrary sub-period between two census years. They were first introduced in Reis (2011). These geospatial units can be analysed as municipalities with constant borders throughout the studied period.

An issue encountered while processing the GDP per capita data was the mismatch of information between the population and GDP data per municipality. Since there is no existing database on GDP per capita per municipality or for the MCAs, there are some minor municipalities in all the years analysed that do not have a match-either from the population per municipality database with the GDP one, or vice versa. This can be due to various factors, such as the merger of a municipality with another, a name change without official record, the dissolution of a municipality, or the nonexistence of data for that period. These mismatches account for less than 1% of the total database and have been excluded to ensure a more accurate and correct analysis.

Some typographical errors in the GDP data for some municipalities were also encountered in the official data from IPEA and IBGE. To solve these issues, the data used on the GDP per capita is filtered using the percentiles from 0 until the 99.75 percentile, as they resulted in significant bias in the obtained results. The errors included extremely large and incorrect GDP numbers for specific years in the dataset that translated into the MCAs and consequently into the inequality indices and the results.

All the digitalized highway data in the format of shapefiles and the bilateral travel time and distances across municipalities are obtained from the database of Morten and Oliveira (2024) and can be seen in Figure 2 in the Appendix. The source for the shapefiles developed in the paper from 1960 onwards is from the Ministério dos Transportes, and for the highway's information from 1940 and 1950 it was manually created from the 1960's map deleting the roads constructed in 1940 and 1960 for the former and 1950 and 1960 for the first. To achieve this the data from Brazilian National Transport plans (Ministerio dos Transportes, 1974) and Rodrigues (2008) were used.

The trade costs for the construction of the market access variable are computed with the bilateral traveling cost in the same way as suggested in Astorga-Rojas (2024) where:

$$\tau_{ijt} = e^{c(i,j)t} \tag{1}$$

The exponential form of Equation 1 has the interpretation that the instantaneous trade cost are of iceberg form, a usual assumption in many different models of trade. The bilateral lowest travel travel time between the states will be used as $c(i,j)_t$. The travel time reference for each state is the one of its capital with the others, this is done with the goal of simplifying and facilitating the computation process and obtaining more significant results. Thus, generating a $26 * 26$ matrix with each state having 25 different values of trade costs.

The results are then normalized as shown in Arkolakis and Allen (2014), to facilitate its interpretation and the one of the market access variable. The trade costs of the diagonal of the matrix were all established with a value of 1, making the assumption that there is no cost to trade between a state and itself.

The data for the control variables and the instrument were also obtained from IPEA and IBGE. The Agricultural GDP data contained certain typographical errors, which were corrected by the author. The control variables were constructed using the same data employed to compute GDP per capita, aggregating values across all MCAs within each state. Specifically, the denominator was calculated by summing the relevant variables for all MCAs that comprise a state, while the numerator was sourced from official state-level data. This approach was used to construct both control variables.

4 Methodology

4.1 Inequality Indices

As stated by Magalhães and Alves (2022) regional inequalities can be measured by a wide variety of indices. In this thesis, the four nonparametric inequality indices shown in their paper are considered to add more robustness to the analysis and three will be used for the results to check if they are consistent to scale changes. The regional inequality indices will be formed with the MCAs and computed at a state level with the per capita GDP obtained. There are two states that will not be considered for the computation of the inequality indices as they are formed by only one MCA which makes the computation of the inequality indices not possible, the states that are not considered and will not take part in the results are Rondonia (RO) and Roaraima (RR).

4.1.1 Gini Coefficient Across Municipalities

$$G = \left(\frac{1}{2\mu}\right) \frac{1}{n(n-1)} \sum_{i}^{n} \sum_{j}^{n} |y_i - y_j| \tag{2}$$

The Gini coefficient across regions computed by Kakwani (1980) is considered one of the most popular and used indices to measure inequality. It offers a comprehensive view of inequality by considering the entire distribution of a population. This complete information measure allows for direct comparisons between populations, irrespective of their size. As a consequence, in this study, we can directly compare the results obtained from the data at different scales.

Equation 2 is formed by y_i and y_j which are the GDP per capita of regions i and j, n is the number of MCAs that form a state and μ is the arithmetic mean of the per capita GDP of the regions that form a state. The Gini coefficient goes from 0 (perfect equality between all the MCAs in a state) or 1 (perfect inequality across all MCA in a state) or what it is the same, all the GDP is concentrated in one MCA. Thus, the Gini coefficient can be viewed as the arithmetic mean of the difference of the per capita GDP between all the MCAs that make up a state divided by the largest possible value of this average (2μ) as it is shown in Shankar and Shah (2003).

4.1.2 Theil Index

$$T = \sum_i x_i \log \left(\frac{x_i}{q_i} \right) \tag{3}$$

Theil index is, as shown in Siddique and Khan (2021), one of the most extensively used approaches for analysing regional inequality. Its interpretation differs from the Gini coefficient as the values are not comparable between different units or nations, so the index limits will be specific of this individual analysis.

In Equation 3, the same approach as in Theil (1967) is followed, where x is the share of GDP of the MCA and q_i is the population of the MCA. If the income distribution across the geospatial units studied is perfectly equally distributed, the Theil index is equal to 0, which means that there is equality between the GDP of all the MCAs that form a state, or, in other words, regional GDP mirrors its population size.

4.1.3 Coefficient of Variation

$$CV = \frac{\sqrt{\frac{1}{T-1} \sum_i (y_i - \mu)^2}}{\mu} \tag{4}$$

The Coefficient of variation (CV) is also a widely used measure of regional inequality, as it quantifies the dispersion of values around the mean as shown in Equation 4. Specifically, the CV of regional per capita GDP (y) is calculated as the ratio of its standard deviation to its arithmetic mean (μ).

A CV value of 0 indicates perfect equality, meaning all regions have identical per capita GDP. As the CV increases, it reflects greater inequality across the regions, with values further from zero indicating a wider spread of per capita GDP among the MCAs that make up the distribution. Thus, a higher CV means more pronounced disparities in regional economic performance, as demonstrated in Allison (2016).

4.1.4 Standard Deviation of per capita GDP

$$\sigma = \sqrt{\frac{1}{T-1}\sum_i (y_i - \mu)^2} \tag{5}$$

Where T is the number of MCAs that form a state, y_i is the per capita GDP of region i, and μ is the arithmetic average of regional per capita GDP. This index is computed by first converting the data to a logarithmic scale. This helps to reveal potential proportional variations that wouldn't be as clear on a natural scale. Consequently, this transformation effectively shows information on variations from both the tails and the middle of the distributions. This transformation is discussed in Sala-i-Martin and Barro (1991). A lower standard deviation indicates a more egalitarian distribution of wealth.

To see the results of the computations all the inequality indices in the 24 studied states across different time periods see Table 11 in the Appendix. See also Figure 7, Figure 8, Figure 9 and Figure 10 to see the evolution of each index across the years studied.

4.2 Market Access

The market access variable is computed in two different ways to add robustness to the results and it is interpolated to facilitate a uniform yearly study. Firstly, the empirical approach from Donaldson and Hornbeck (2016) where:

$$MA_i \approx \sum_d \tau_{id}^{-\theta} N_d \tag{6}$$

With τ being the trade costs between state o and another state, d being the different number of states and N_d, the population of the state d and θ being the trade elasticity, which measures how sensitive the exports are to an increase in the trade cost.

The second approach is the one suggested by Astorga-Rojas (2024) used for the specific scenario of Brazil, where:

$$MA_{it} = \sum_{j \neq i} \tau_{ijt}^{-\theta} \text{shareGDP}_{jt} \tag{7}$$

In which, instead of using the population of the state, the share of Brazil's GDP in that state is used. The trade elasticity values used to compute the market access variable in the two different approaches differ depending on the method and paper selected. We can take as a reference those obtained by Astorga-Rojas (2024) and Morten and Oliveira (2024), which each can be examined with the different methods used in Table 1.

Table 1: Trade elasticity estimates

Study	**Method**	**Elasticity (θ)**
Astorga-Rojas	2SLS	3.39
Astorga-Rojas	OLS	2.80
Morten and Oliveira	OLS	2.96
Morten and Oliveira	IV	3.26

Source: Author's own elaboration.

Figure 3 and Figure 4 in the Appendix show the growth and evolution of the market access variable using the Morten and Oliveira IV trade elasticity and Equation 6 and Equation 7. It can be seen that states farther from the coast and in the interior have experienced more significant growth in market access since the 1940's, as their initial conditions were much worse than those of the more populated areas in the eastern part of the country as it can be noted by looking at Figure 2.

This same pattern can also be seen in Figure 5 and Figure 6, where market access is computed using Equation 6 and with the same estimated θ coefficient. The growth values for market access in this computation are slightly higher than those obtained when using GDP, possibly due to migration trends.

4.3 Empirical Estimations

With market access and the inequality indices constructed different econometric regression are constructed in order to obtain more robust and precise results:

$$ln(y_{it}) = \alpha_i + \gamma_t + \beta ln(MA_{it}) + \epsilon_{it} \tag{8}$$

Equation 8 shows the simplest regression with only one regressor. It consists in a difference-in-differences panel data estimation where MA_{it} by itself is used as the tool to measure the treatment effect of the natural experiment, α_i and γ_t act as the fixed effects and the time fixed effects respectively and β as the elasticity that measures the effect of the increase of market access on the inequality indices y_{it}.

4.3.1 Differences Across Regions

The addition of an interaction term to the model to examine differentiated results across regions states can also be considered a factor of interest. The creation of a dummy variable to distinguish southern states from the interior and north of Brazil is significant to the development of regions as coastal and southern states are more industry and urban-led, while the interior of Brazil is economically driven by agriculture and rural factors as per IBGE data. The division of the states has been done according to the macro-regions division where the North, Center-West and North-East macro-regions states are considered the treated variable in the dummy created.

$$ln(y_{it}) = \alpha_i + \gamma_t + \beta ln(MA_{it}) + \delta ln(MA_{it})(DummyState = 1) + \epsilon_{it} \tag{9}$$

Equation 9 incorporates the interaction term to assess the heterogeneous effects of market access on inequality across coastal and interior Brazilian states within the difference-in-differences regression shown in Equation 8. This approach treats coastal and southern states as a "control group", which are more industrially led, while interior states serve as "treatment group" characterized by rural-driven economies. The coefficient β represents the baseline elasticity of MA's impact on inequality for coastal states. The interaction term coefficient δ quantifies the differential effect of MA in interior states modifying the total market access elasticity to $\beta + \delta$.

4.3.2 Addition of Control Variables

To avoid issues of omitted variable bias and to obtain better results, equation 10 is constructed, including two control variables. The first one being the share of GDP of each state dedicated to agriculture to control for statal changes in the means of the economy. As shown in Azzoni (2001), economic expansion leads to regional divergence and those states that depend mostly on agricultural GDP are the ones where that expansion is less noticeable as shown in Arias (2016). The other control variable is the ratio of urban population on each state to control for any migration shock or change in the population shape that may have caused a big variation on the structure of the state population and as a consequence on the regression. The final regression takes the following form:

$$ln(y_{it}) = \alpha_i + \gamma_t + \beta ln(MA_{it}) + \delta ln(MA_{it})(DummyState = 1) + \psi AS_{it} + \lambda UR_{it} + \epsilon_{it} \quad (10)$$

With ASs_{it} being the share that agriculture has in the total GDP of the state across time and UR_{it} being the ratio between urban and total population of each state across time.

4.3.3 Instrumented Market Access

Equation 8 cannot be fully interpreted as the causal effect of market access on regional inequality, as there may be unobservable endogenous factors that could bias the results. For example, the presence of railways as an alternative to highways for transporting goods could facilitate trade and reduce trade costs, independently affecting both market access and inequality. To address this potential endogeneity, I use the kilometers of railroads in each state across time as an instrument to construct an instrumented version of the market access variable constructing Equation 11.

$$ln(y_{it}) = \alpha_i + \gamma_t + \beta ln(MA\_IV_{it}) + \epsilon_{it} \quad (11)$$

Railroad kilometers can be considered a plausible instrument because they influence inequality indices through their impact on market access and can reasonably be considered exogenous. Employing this instrumental variable approach could add robustness and improve results.

5 Results

All the results presented are computed using market access as defined by Equation 7, which follows the approach suggested by Astorga-Rojas (2024) for the specific context of Brazil. Results using Equation 6 for the computation of market access are also provided in the Appendix. Since the latter measure is more sensitive to migration and population shocks, the interpretation and main computations in this analysis rely on the market access measure based on the share of GDP. While the coefficients obtained across all regressions do not differ substantially, there are some minor variations in the significance of certain estimates that may cause that the final interpretation differ across different computations.

The trade elasticity θ used for the computation of market access and in the subsequent estimation of the various models is taken from Morten and Oliveira (2024). In their study, $\theta = 3.26$, an estimate obtained using instrumental variables and based on the natural experiment of the construction of Brasília, as detailed in their paper.

5.1 Baseline Average Effect of MA

Table 2: OLS Estimates with One Regressor

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access	0.006	0.028***	−0.053***
	(0.008)	(0.010)	(0.019)
Observations	1,224	1,224	1,224
R^2	0.001	0.007	0.007
Adjusted R^2	−0.064	−0.057	−0.057
F Statistic (df = 1; 1149)	0.598	8.113***	7.858***

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

Table 2 shows the correlation between the logarithm of market access and the different inequality indices at a baseline level, without any additional predictors. The results indicate different effects depending on the inequality indices used. The log Gini coefficient presents a highly non-significant estimate, whereas the log Coefficient of variation and log Theil index display significant but contrasting impacts. For the Coefficient of variation, there is a positive effect: a 10% increase in market access translates into a 0.28% increase in the Coefficient of variation, indicating greater inequality and a more unequal distribution of income across states. In contrast, for the Theil index, a 10% increase in market access is associated with a 0.53% reduction in the index, suggesting a decrease in inequality across the regions within a state.

The common factor among the three dependent variables and models analyzed is that both the R^2 and Adjusted R^2 values are close to zero. This indicates that the model, which includes only the logarithm of market access as an explanatory variable together with fixed and time effects, does not explain the variance in the dependent variable. This lack of explanatory power may be due to omitted variables that are relevant for income inequality but are not controlled for in this specification. Additionally, the statistical significance observed should be interpreted with caution before examining the behaviour of the residuals as their distribution may be causing a bias in the significance and interpretation of the estimates. To address this, the residual vs fitted values plots (see Figure 11a and Figure 11b in the Appendix) are examined. From these, it is evident that the Theil index exhibits a high degree of heteroskedasticity and non-random patterns, which introduce bias into the estimates and consequently affect their interpretation. To address this issue, the inclusion of additional control variables could be beneficial.

On the other hand, Figure 11b shows a more randomly distributed pattern of residuals, making the estimate for the Coefficient of variation more reliable and interpretable. Nevertheless, the absence of interactions or instrumental variables in this model may still influence the estimated effects and its interpretation. From this, and only as an introductory intuition about the effect of market access on income inequality, we could suggest that market access increases inequality across regions, as it may contribute to the concentration of production factors and income on average on all states as, at this point, the only coefficient that can be used to interpret this baseline effect is the Coefficient of variation, since it is the only estimate that shows both statistical significance and a proper residual distribution. When market access is computed using population as the weighting factor, the statistical significance of the results changes slightly; however, the estimated values remain very similar. As a result, the overall interpretation of the findings does not change. For detailed estimates, see Table 8 in the Appendix.

For a more robust analysis of the baseline model, the addition of control variables should be considered to determine whether there is any improvement or change in the estimates obtained. Table 7 in the Appendix presents these updated estimates. The control variables added are those described in Section 4.3.2, namely the share of urban population in the state and the share of agricultural GDP in the total GDP of the state. Both control variables show strong significance across the three inequality indices used as dependent variables, and a substantial improvement in the R^2 and Adjusted R^2 is observed, with values increasing from nearly zero to 0.197 and 0.180, respectively, for the adjusted measure. The log Gini coefficient remains statistically insignificant, as seen in Table 2, and the contrast between the estimates for the Coefficient of variation and the log Theil index persists.

To assess whether the significance and estimates remain unbiased, the residuals versus fitted values plots are checked again for both significant coefficients (see Figure 11c and Figure 11d in the Appendix). The results are very similar to those from the baseline model, with noticeable heteroskedasticity in the Theil index plot, which may create bias in the p-value and thus affect the final interpretation. In contrast, the residuals for the Coefficient of variation remain randomly distributed.

5.1.1 First Stage

To study the causality between market access and income inequality at a baseline level in greater depth, I propose the use of an instrumental variable, as explained in Section

4.3.3, where market access is instrumented with the kilometers of railroad in each state over time, leading to Equation 11. The instrument proves to be both strong and significant ($p - value < 0.01$), as reflected in the first-stage regression results shown in Table 3.

Table 3: First Stage: Dependent Variable log Market Access

Dependent Variable:	log Market Access
Variable	
Railway_km	0.0007***
	(0.0002)
Fixed-effects	
state	Yes
year	Yes
Fit statistics	
Observations	1,224
R^2	0.94201
Within R^2	0.14543

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Source: Author's own elaboration.

The estimate obtained is small but highly significant, indicating that for every one-unit increase in the kilometers of railway per state, there is an estimated increase in log market access of 0.0007. Although this effect may appear modest, the Within R^2—which measures the explanatory power of railway kilometers on log market access after accounting for state and time fixed effects—shows that 14.5% of the variation in log market access can be attributed to the kilometers of railway. This finding further supports the relevance of railway kilometers as an instrument. This result is consistent with expectations, as railways serve as an alternative to highways in within-country trade and are an important factor influencing market access.

5.1.2 2SLS Results

The estimates from the second stage are presented in Table 4. Notable changes in significance are observed: when market access is instrumented, the logarithm of the Gini coefficient becomes significant, while the Coefficient of variation loses its significance. The sign of the coefficient turns negative for all inequality indices when railroad kilometers are used as an instrument. Both the R^2 and Adjusted R^2 remain close to zero, with estimates very similar to those obtained from the baseline specification in Equation 8, as shown in Table 2.

Table 4: Estimates with Instrumented Market Access

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access with IV	-0.085^{***}	-0.042	-0.154^{***}
	(0.022)	(0.026)	(0.050)
Observations	1,224	1,224	1,224
R^2	0.001	0.007	0.007
Adjusted R^2	-0.064	-0.057	-0.057
F Statistic	14.574^{***}	2.597	9.350^{***}

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

The coefficients obtained indicate that a 10% increase in market access leads to a 0.85% decrease in the Gini coefficient and a 1.54% decrease in the Theil index. These results suggest that, when using instrumented market access, an increase in market access—or, equivalently, a reduction in trade costs—reduces regional inequality. This finding contrasts with the initial intuition that greater market access would lead to a concentration of production factors and income. The substantial change in the estimated effects, observed only when considering the impact of railroads on market access, raises questions and potential concerns, as the direction of the relationship has shifted considerably compared to the baseline regression. One possible explanation is that the greater presence of railroads in more industrialized states may be influencing the coefficients.

To assess the robustness of these results, I examined the residuals versus fitted values plots (see Figures 11f and 11e in the Appendix). These plots reveal noticeable heteroskedasticity and non-linear patterns, which may introduce bias into the significance of the covariates and limits the interpretation of those coefficients and its significance, doubting the effect of the instrument as it also happens in Donaldson and Hornbeck (2016).

5.2 Effect of Market Access on Different Regions

To tackle the differing results obtained in the instrumental variable analysis in Table 4 and the baseline one, I explore the possibility that market access has varying effects depending on the type of state, constructing a dummy variable as described in Section 4.1.1. This variable distinguishes the more rural macro-regions in Brazil (North, Center-West, and North-East) from the more industrialized ones (South and South-East). Table 5 presents the estimates and results from including the interaction term of this dummy variable in the baseline regression, thereby constructing Equation 9.

Table 5: OLS Estimates with Region Dummy

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access	0.026**	0.054***	−0.004
	(0.013)	(0.016)	(0.030)
log Market Access × DummyState	−0.018*	−0.024**	−0.045**
	(0.009)	(0.011)	(0.022)
Observations	1,224	1,224	1,224
R^2	0.004	0.011	0.010
Adjusted R^2	−0.061	−0.054	−0.054
F Statistic (df = 2; 1148)	2.201	6.432***	6.091***

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

The results confirm the possibility of heterogeneous effects depending on the Brazilian macro-region. The coefficients change substantially compared to both the IV regression and the baseline model. Log market access now shows strong significance with the log of the Gini coefficient as the dependent variable, as well as with the log of the Coefficient of variation. Additionally, the interaction between log market access and the newly created regional dummy is significant for all three inequality indices studied. Log market access exhibits a positive effect at the baseline level in the industrial regions: a 10% increase in market access leads to a 0.26% increase in the Gini coefficient and a 0.54% increase in the Coefficient of variation. In contrast, in the rural and less developed macro-regions, the effect of market access—when accounting for the interaction—is more than halved, with an effect of $0.026 - 0.018 = 0.008$ on the Gini coefficient and $0.054 - 0.024 = 0.03$ on the Coefficient of variation. Very similar results are obtained when market access is computed using population as the weighting factor, which adds further robustness to the analysis and the estimates. For a more detailed view, see Table 9 in the Appendix.

These findings suggest a significant differentiated effect depending on the region where the increase in market access or reduction in trade costs occurs. In industrial and more urban areas, there is a clear and significant increase in regional inequality, possibly due to the income concentration that improved infrastructure, such as road construction, can cause. However, this effect is substantially reduced in more rural and agriculturally driven regions, likely due to already high trade costs stemming from poor infrastructure. Improving infrastructure in these regions may have more diffuse effects, potentially benefiting a broader set of areas rather than just the region where the infrastructure is directly implemented.

As in previous sections, I examine the residual vs. fitted values plots (Figures 12b and 12a) to assess potential heteroskedasticity or non-linear patterns that could bias the results. The plots reveal randomly distributed residuals centered around zero for the two significant dependent variables. This indicates homoskedasticity and the absence of non-linear relationships, suggesting that there is no systematic bias affecting the significance of the estimates. Consequently, we can interpret the estimates with reasonable confidence.

The R^2 and Adjusted R^2 values show results very similar to those from the baseline regression in Equation 8 (Table 2) and the instrumented market access model in Equation 11 (Table 4). In all cases, both metrics remain near zero across the inequality indices analyzed. This indicates that the model—whether using only market access, the instrumented version or including the regional dummy interaction—does not explain the variance in inequality indices on its own possibly to omitted-variable-bias.

This limitation is addressed in Section 5.1 and it is solved partially through the inclusion of control variables, as demonstrated in previously in Table 7 and with the Dummy differentiating states in Table 6.

Table 6: OLS Estimates with Region Dummy and Control Variables

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access	0.040***	0.070***	0.032
	(0.011)	(0.014)	(0.027)
Urban Ratio	0.217***	0.237***	0.587***
	(0.057)	(0.070)	(0.135)
Agricultural Share	−1.359***	−1.583***	−3.059***
	(0.072)	(0.088)	(0.171)
log Market Access × DummyState	−0.035***	−0.045***	−0.084***
	(0.008)	(0.010)	(0.019)
Observations	1,224	1,224	1,224
R^2	0.260	0.245	0.249
Adjusted R^2	0.210	0.194	0.198
F Statistic (df = 4; 1146)	100.675***	92.837***	94.938***

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

The results obtained confirm this to be the best model so far, while maintaining the intuition established with the baseline model plus the interaction term. The differentiated effects across regions persist: the baseline estimate indicates that log market access increases inequality, whereas the effect is reduced by more than half when specifically analysing rural regions. This adds further robustness to the intuition previously developed. Significance remains strong for all predictors, except for log market access in the Theil index, which mirrors the findings from the model with only the interaction term.

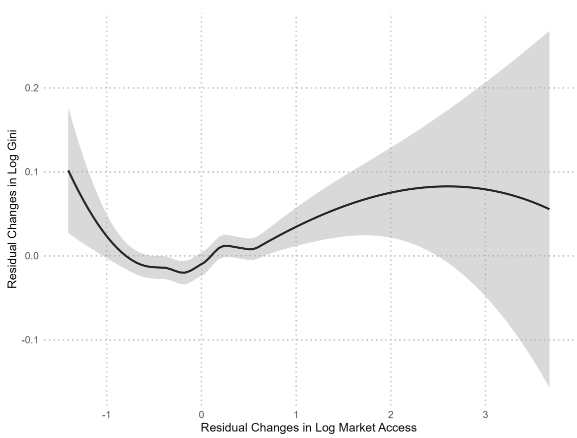
The addition of the control variables: urban ratio and agricultural share, as previously shown in Table 7 for the model without the interaction, yields very similar results here. The explanatory power of the model is greatly improved, with the Adjusted R^2 and R^2 increasing to around 0.25 and 0.2, respectively. This represents a substantial improvement over the previous best model, which did not include the interaction term and lacked the differentiation across the macro-regions. With market access computed using population as the weighting factor, the results exhibit a very similar pattern, which adds further robustness to the analysis. For detailed estimates, see Table 10 in the Appendix.

The analysis of the residual versus fitted values plots is also conducted to assess the consistency of the estimates. Examining Figures 12c and 12d, it can be observed that, similar to the baseline model with interactions, the residuals are randomly distributed around zero. This suggests that the significance of the coefficients obtained is meaningful and that there is a degree of homoskedasticity in the residuals distribution, which supports the underlying assumption of the model and gives certainty on the final interpretation of the model.

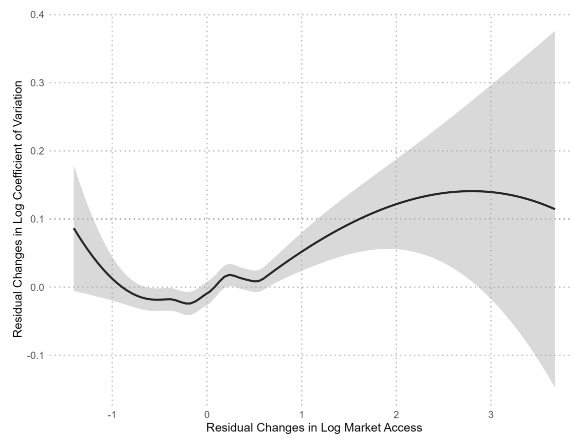
To better examine the relationship between the two significant dependent variables and log market access, I constructed plots to visualize the local polynomial relationship between changes in these variables. The two panels shown in Figure 1 illustrate this relationship. These plots are designed to display the conditional association between changes in market access and changes in economic outcomes, such as inequality, after removing the effects of confounding factors like fixed effects or geographic controls. Specifically, the figure presents the local polynomial relationship between residual changes in the log of the significant inequality indices and residual changes in log market access, with the shaded region representing the 95% confidence interval.

To construct these plots, both the dependent and independent variables are first "residualized"—that is, each variable is regressed on controls such as state fixed effects and time effects. The resulting residuals capture the variation in each variable not explained by these controls. A local polynomial (nonparametric) regression curve is then fitted to these points, allowing for a flexible and potentially non-linear depiction of the relationship.

Figure 1: Local Polynomial Relationships between Inequality Measures and Market Access



(a) Local Polynomial Relationship between Changes in Log Gini Coefficient and Log Market Access



(b) Local Polynomial Relationship between Changes in Log Coefficient of Variation and Log Market Access

Source: Author's own elaboration.

Panel (a) of Figure 1 presents the local polynomial relationship between residual changes in the log Gini coefficient and residual changes in log market access. The estimated curve reveals a non-linear association: at lower levels of improvement in market access, the Gini coefficient tends to decline, suggesting that modest improvements in market access may have contributed to reducing inequality. However, beyond a certain threshold, further increases in market access are linked to rising inequality. This pattern suggests that while the initial expansion of highways may have provided broadly shared benefits, the subsequent, more substantial gains likely accrued disproportionately to wealthier regions, thereby increasing inequality. This finding aligns with the earlier analysis including the dummy variable for more rural regions.

A similar dynamic is observed in Panel (b), which explores the relationship between market access and the Coefficient of variation—another indicator of inequality that captures the relative dispersion of income or wealth. Here as well, inequality initially decreases as market access improves, but then rises again at higher levels of access. The consistency of these results across both inequality measures reinforces the interpretation that the distributional effects of improved market access were not uniform across the population and regions.

6 Conclusion

This thesis develops several models to estimate the effects of market access on regional inequalities, employing different inequality indices to enhance the robustness of the analysis. The empirical strategy centers on the natural experiment created by the relocation of Brazil's capital, using this event as a basis for identifying causal effects. The baseline estimates indicate that increases in market access are associated with higher regional inequality, suggesting a concentration of income or productive factors in regions that receive infrastructure investments—particularly highways aimed at reducing trade costs. The interpretable and statistically significant results show that a 10% increase in market access leads to a 0.28% increase in the Coefficient of variation, indicating a modest rise in inequality as market access expands.

The analysis further investigates whether the effect of market access varies across different macro-regions, specifically contrasting more rural, less developed areas, where wealth tends to be more evenly distributed and infrastructure is less prevalent, with more industrialized regions. The results reveal important heterogeneity: the increase in inequality associated with greater market access is substantially reduced—by more than half—when focusing on rural regions. This suggests that infrastructure investments in these areas tend to generate a much smaller increase in income inequality compared to the average effect.

These findings are consistent with the patterns observed in the graphical analysis of inequality and market access evolution. The relationship between market access and regional income inequality is not linear. In regions characterized by lower initial inequality and a predominantly rural economy, the growth in inequality resulting from increased market access is less pronounced. However, as these regions develop and undergo economic and social transformation, the impact of market access on inequality intensifies.

It is important to acknowledge certain limitations of this thesis. Some influential factors in the measurement of trade costs—such as border effects between states or the potential for waterway trade—are not considered in order to simplify the model. Additionally, some data had to be interpolated to ensure a uniform analysis, which may introduce bias if infrastructure or inequality shocks are given in certain periods, as these may not be fully captured in the interpolated data.

Overall, this thesis contributes to the literature by providing evidence on the effect of market access on regional inequality, and it strengthens the validity of its findings by employing multiple measures of both market access and inequality.

References

Acker, A. (2021). Amazon development. *Oxford: Oxford University Press.*

Allison, P. D. (2016). Measures of inequality. *American Sociological Review*, (6):865–880.

Arias, J. (2016). The contribution of agriculture to development. *Inter-American Institute for Cooperation on Agriculture.*

Arkolakis, C. and Allen, T. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Astorga-Rojas, D. (2024). Access to markets and technology adoption in the agricultural sector: Evidence from brazil. *EconStor-ZBW - Leibniz Information Centre for Economics*, (289868).

Azzoni, C. R. (2001). Economic growth and regional income inequality in brazil. *The Annals of Regional Science*, 35(1):133–152.

de Sousa et al., J. (2012). Market access in global and regional trade. *Regional Science and Urban Economics*, 42(6):1037–1052.

Donaldson and Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Burgess, R. (2012). Railroads and the demise of famine in colonial india. Available at LSE repository.

Ehrl, P. (2017). Minimum comparable areas for the period 1872-2010: an aggregation of brazilian municipalities. *Estudos Econômicos (Saô Paulo)*, 47(1).

Kakwani, N. C. (1980). Income inequality and poverty : methods of estimation and policy applications. *New York: World Bank*, 1(10092).

Magalhães, J. C. R. and Alves, P. J. H. (2022). A century of regional inequalities in brazil. *Institute for Applied Economic Research (IPEA)*, (Discussion Paper No. 271).

Morten, M. and Oliveira, J. (2024). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, 16(2):389–421.

Reis, E. J. (2011). Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. *IPEA - Instituto de Pesquisa Econômica Aplicada.*

Rodrigues, P. R. A. (2008). Introdução aos sistemas de transporte no brasil e à logística internacional. *Edições Aduaneiras.*

Sala-i-Martin, X. and Barro, R. J. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution*, 22(1):107–182.

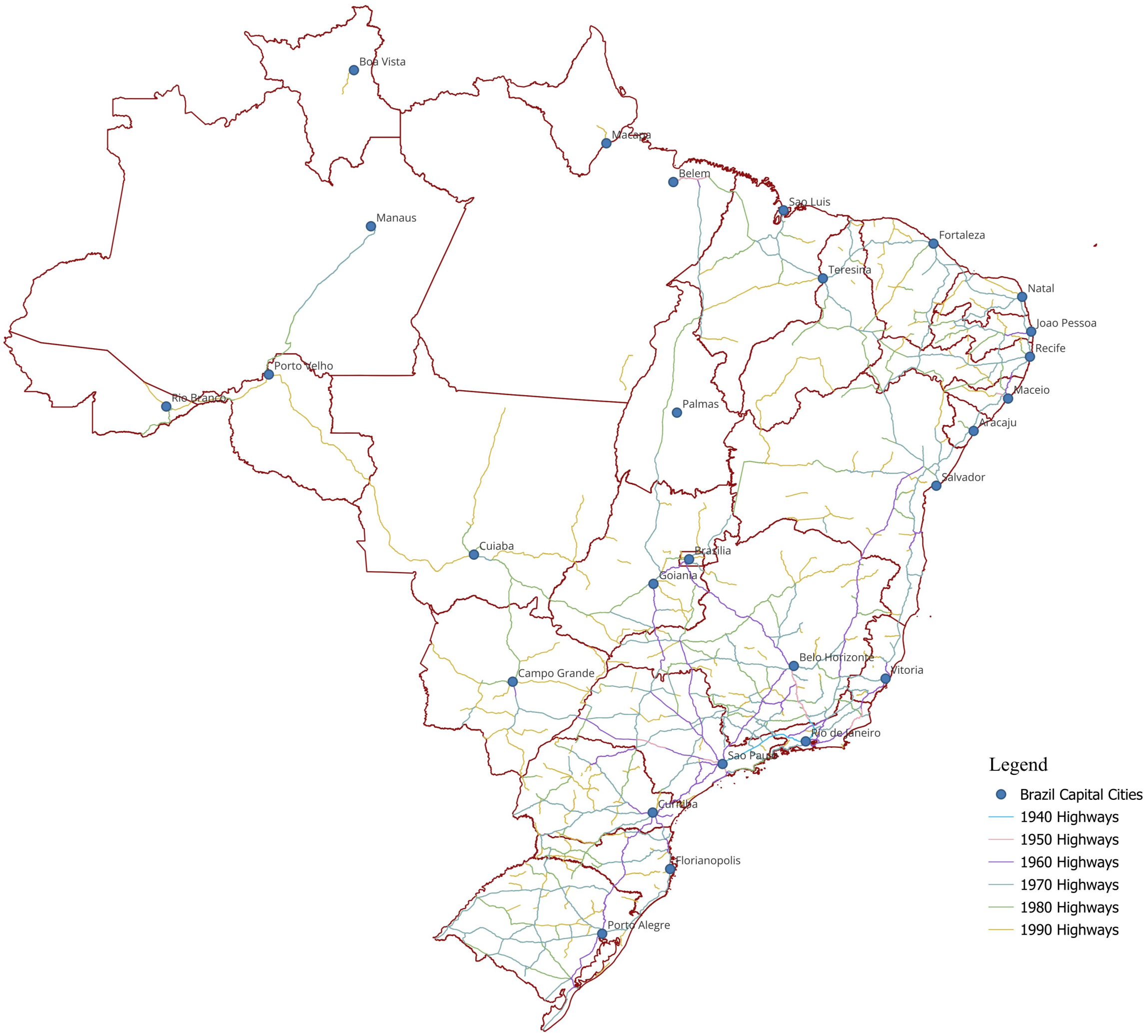
Shankar, R. and Shah, A. (2003). Bridging the economic divide within countries: a scorecard on the performance of regional policies in reducing regional income disparities. *World Development*, 31(8):1421–1441.

Siddique, A. B. and Khan, M. S. (2021). Spatial analysis of regional and income inequality in the united states. *Schar School of Policy and Government, George Mason University.*

Theil, H. (1967). Economics and information theory. *Studies in Mathematical and Managerial Economics*, 7.

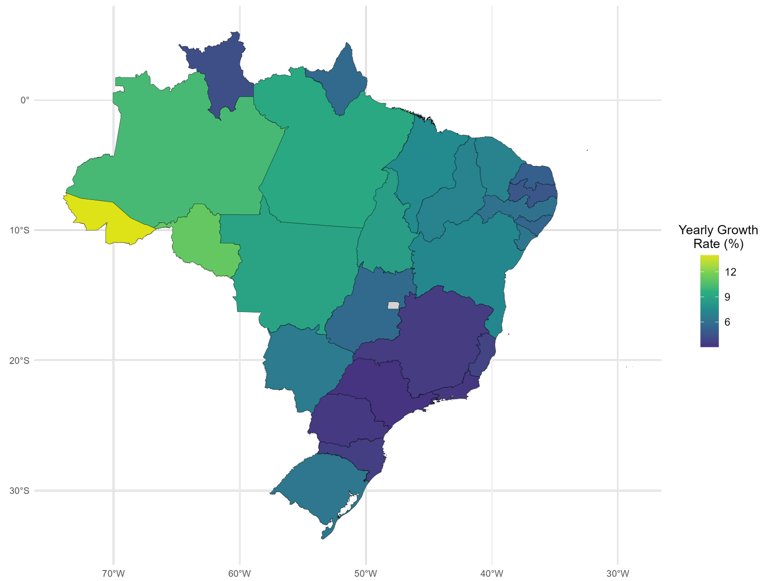
Appendix A: Additional Figures

Figure 2: Highways Evolution Across the Years.



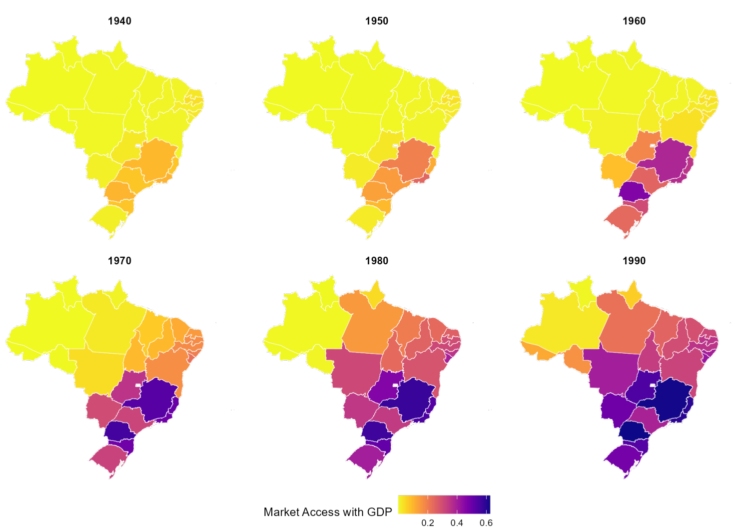
Source: Author's own elaboration.

Figure 3: Yearly GDP Market Access Growth Rate per State: 1940-1990



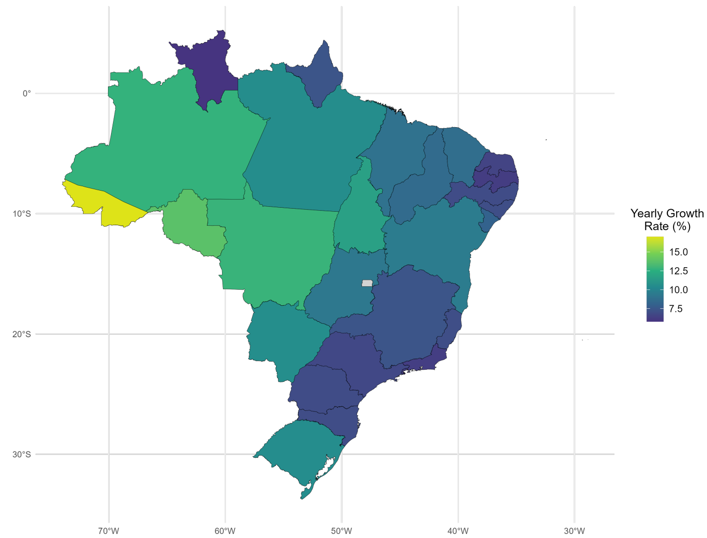
Source: Author's own elaboration.

Figure 4: Market Access Evolution per State: 1940-1990



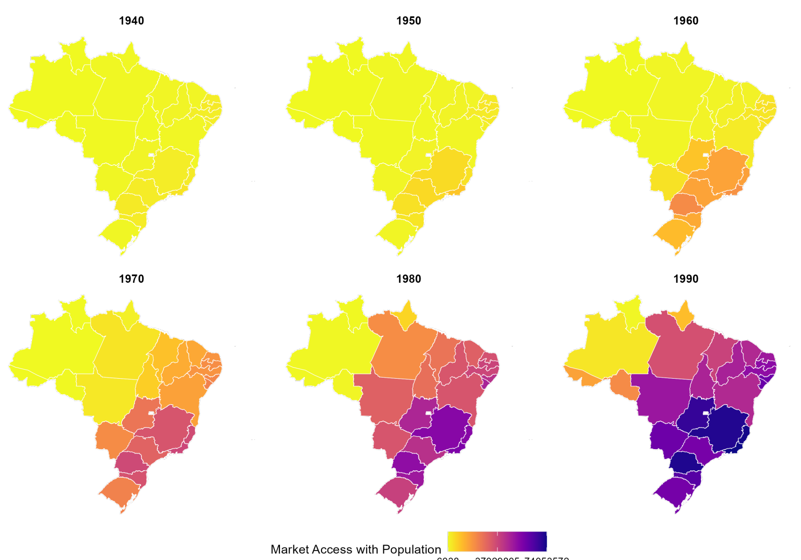
Source: Author's own elaboration.

Figure 5: Market Access Evolution per State computed with Population: 1940-1990



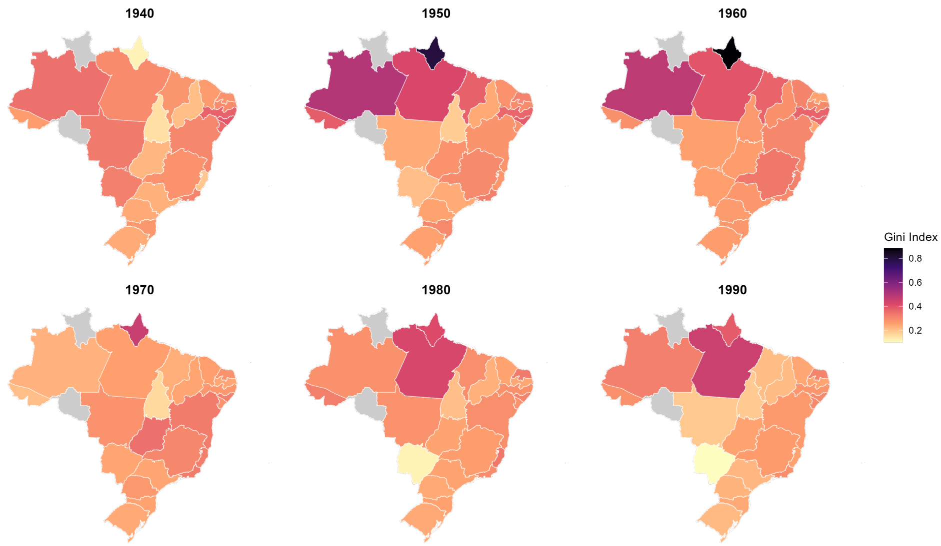
Source: Author's own elaboration.

Figure 6: Market Access Evolution per State computed with Population: 1940-1990



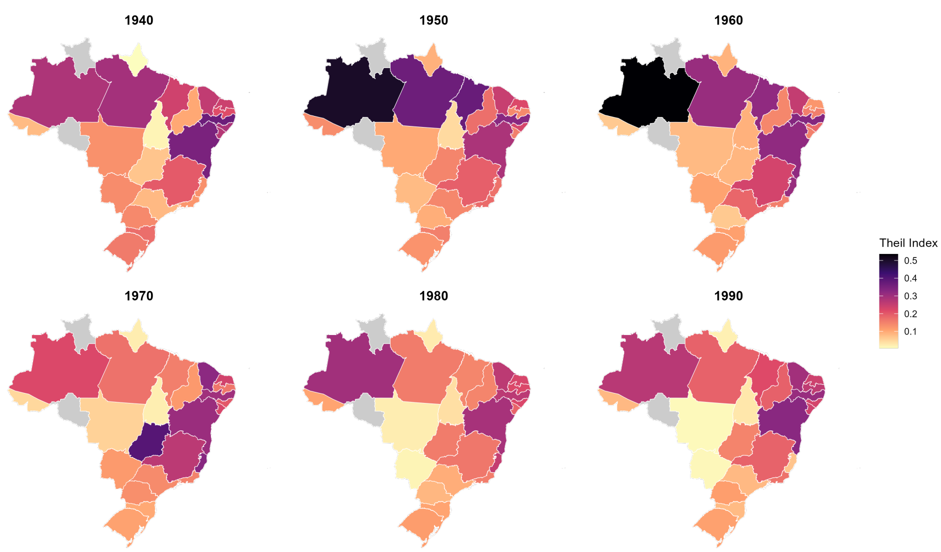
Source: Author's own elaboration.

Figure 7: Gini Evolution by State: 1940-1990



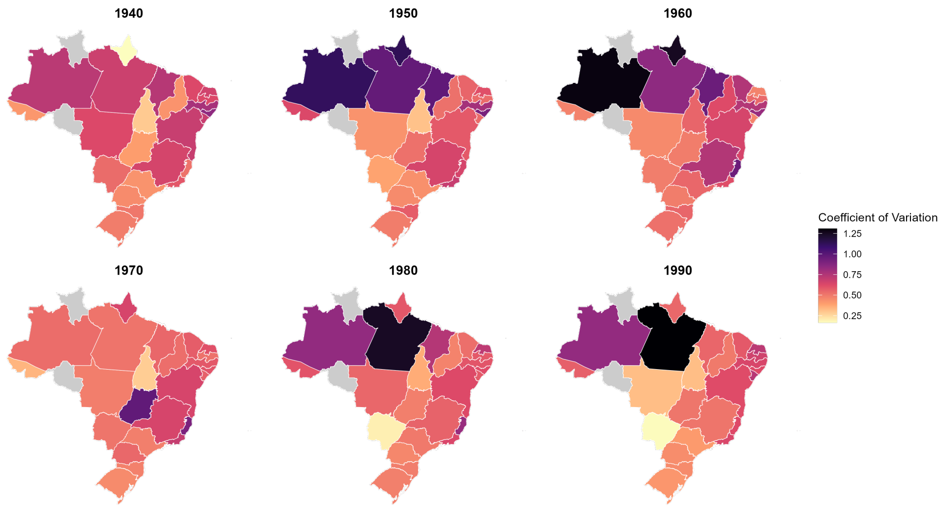
Source: Author's own elaboration.

Figure 8: Theil Evolution by State: 1940-1990



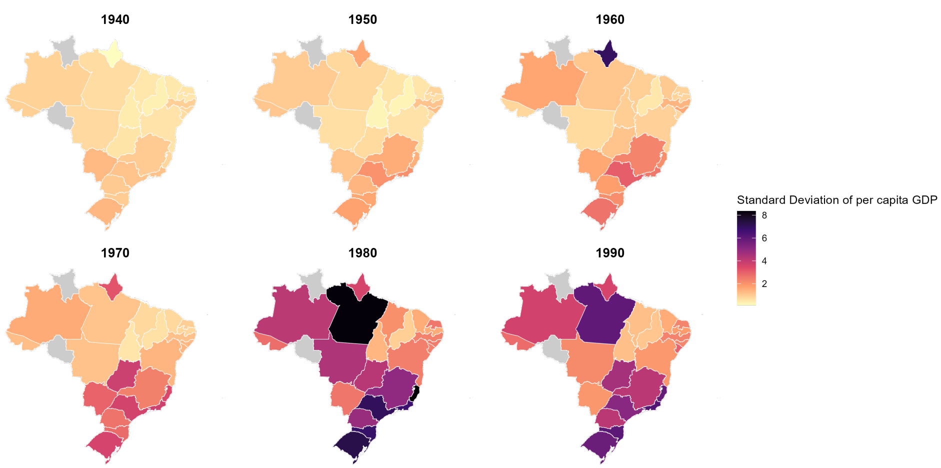
Source: Author's own elaboration.

Figure 9: Coefficient of Variation Evolution by State: 1940-1990



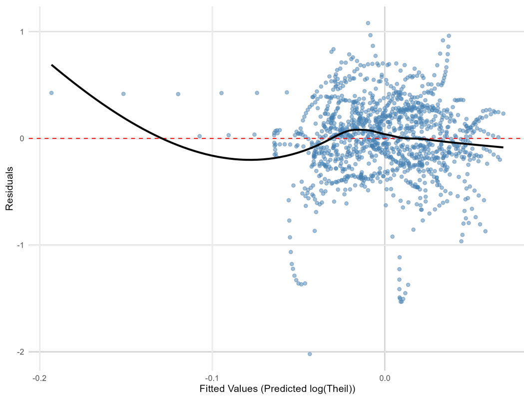
Source: Author's own elaboration.

Figure 10: Standard Deviation of per capita GDP Evolution by State: 1940-1990



Source: Author's own elaboration.

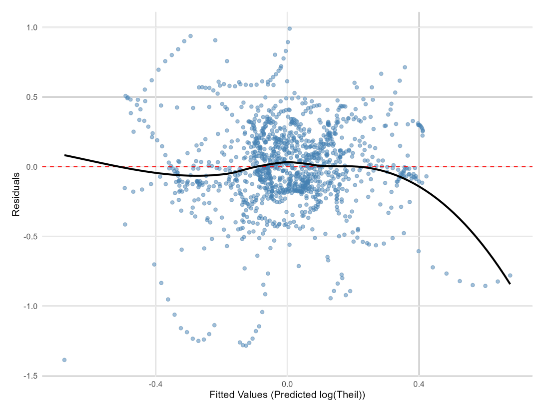
Figure 11: Matrix of Residual vs Fitted Values Plots without interaction for different models



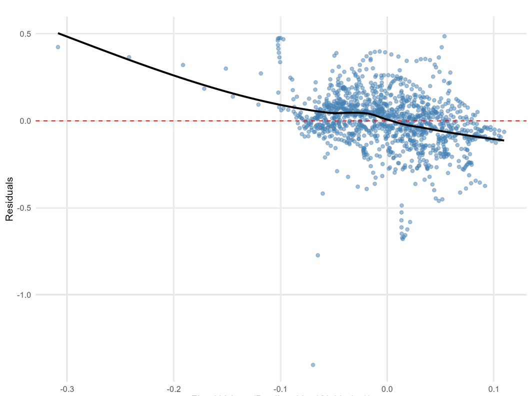
(a) Residual vs Fitted Values for Theil Index

(b) Residual vs Fitted Values for Coefficient of Variation

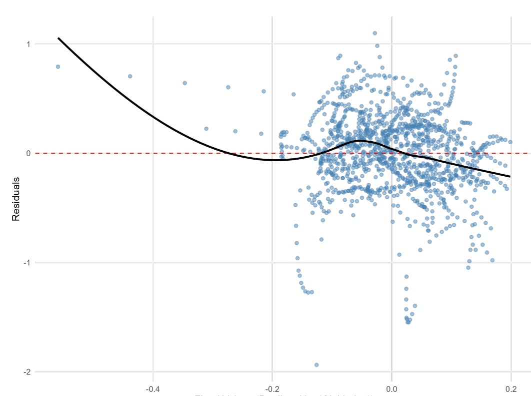
(c) Residual vs Fitted Values for Coefficient of Variation with Controls



(d) Residual vs Fitted Values for Theil Index with Controls



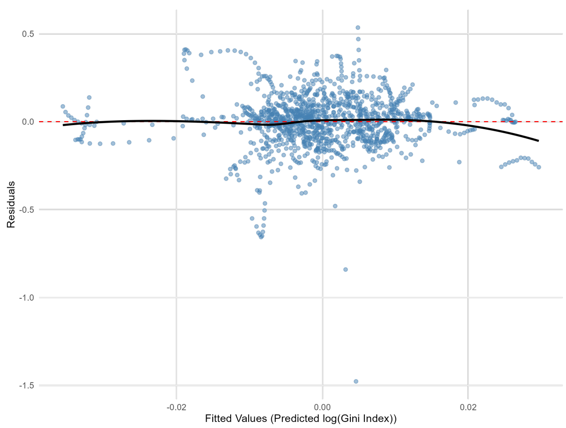
(e) Residual vs Fitted Values for Gini with Instrumented Market Access



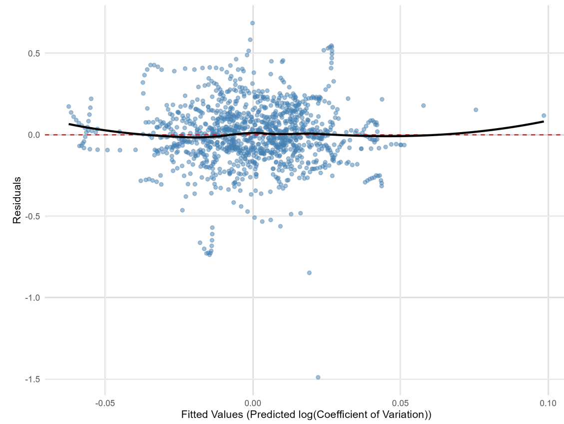
(f) Residual vs Fitted Values for Theil with Instrumented Market Access

Source: Author's own elaboration.

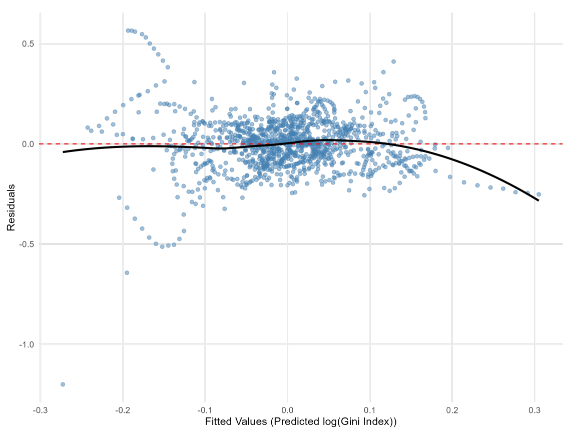
Figure 12: Residual vs Fitted Values Plots for Interaction Models



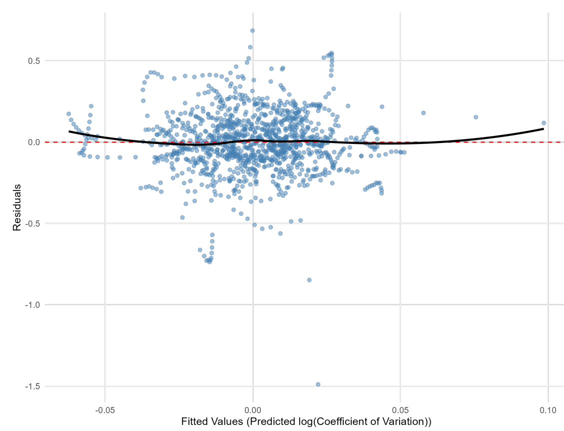
(a) Residual vs Fitted for Gini with Interaction



(b) Residual vs Fitted for the Coefficient of Variation with Interaction



(c) Residual vs Fitted for Gini with Interaction and CVs



(d) Residual vs Fitted for the Coefficient of Variation with Interaction and CVs

Source: Author's own elaboration.

Appendix B: Additional Tables

Table 7: OLS Estimates with Control Variables

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access	0.002	0.022**	−0.059***
	(0.007)	(0.009)	(0.017)
Urban Ratio	0.221***	0.241***	0.596***
	(0.057)	(0.070)	(0.136)
Agricultural Share	−1.322***	−1.536***	−2.971***
	(0.072)	(0.088)	(0.171)
Observations	1,224	1,224	1,224
R^2	0.247	0.231	0.236
Adjusted R^2	0.197	0.180	0.185
F Statistic (df = 3; 1147)	125.611***	114.902***	118.096***

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

Table 8: OLS Estimates of Inequality Measures on Market Access (pop)

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access (pop)	0.029***	0.050***	−0.091***
	(0.010)	(0.012)	(0.023)
Observations	1,224	1,224	1,224
R^2	0.008	0.015	0.013
Adjusted R^2	−0.056	−0.049	−0.050
F Statistic (df = 1; 1149)	8.704***	17.425***	15.476***

Note: All estimates are controlling for state fixed-effects and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

Table 9: Interaction Effect of Market Access (pop) and Treatment on Inequality Measures

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access (pop)	0.070***	0.090***	−0.029
	(0.013)	(0.016)	(0.031)
log Market Access (pop)× DummyState	−0.034***	−0.033***	−0.051***
	(0.007)	(0.009)	(0.017)
Observations	1,224	1,224	1,224
R^2	0.027	0.027	0.021
Adjusted R^2	−0.037	−0.036	−0.043
F Statistic (df = 2; 1148)	15.721***	16.135***	12.332***

Note: All models include state and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

Table 10: Interaction of Market Access (pop) and Treatment with Controls

	Dependent variable:		
	log Gini	log Coefficient of Variation	log Theil
	(1)	(2)	(3)
log Market Access	0.054***	0.071***	−0.067**
	(0.011)	(0.014)	(0.027)
Urban Ratio	0.191***	0.187***	0.500***
	(0.055)	(0.069)	(0.131)
Agricultural Share	−1.356***	−1.568***	−3.216***
	(0.072)	(0.089)	(0.169)
log Market Access (pop)× DummyState	−0.041***	−0.042***	−0.068***
	(0.006)	(0.008)	(0.015)
Observations	1,224	1,224	1,224
R^2	0.275	0.249	0.274
Adjusted R^2	0.227	0.199	0.226
F Statistic (df = 4; 1146)	108.911***	94.992***	108.303***

Note: All models include state and state-time fixed effects.
*p<0.1; **p<0.05; ***p<0.01

Source: Author's own elaboration.

Table 11: Summary of Inequality Metrics by State for Selected Years

Sigla	gini_1940	gini_1970	gini_1990	theil_1940	theil_1970	theil_1990	cv_1940	cv_1970	cv_1990	sd_1940	sd_1970	sd_1990
AC	0.2615385	0.2064856	0.2855114	0.0803797	0.0477744	0.0818393	0.4186927	0.3428157	0.5671652	0.9253108	1.2368791	2.794990
AL	0.3941330	0.3176939	0.2982054	0.2874775	0.2011963	0.2303999	0.8425404	0.5968539	0.6227357	0.8135781	1.3395641	2.280769
AM	0.3476159	0.2268442	0.3184154	0.2819629	0.2213593	0.2685913	0.7248652	0.5350266	0.8319740	0.8796963	1.5588534	3.691968
AP	0.1172414	0.4574575	0.3862434	0.0049916	0.0235615	0.0200809	0.1658043	0.6469425	0.5462306	0.1202082	3.2314780	3.613316
BA	0.3053226	0.3250956	0.2702132	0.3491824	0.3071075	0.3277053	0.6900171	0.6460174	0.6184650	0.5757447	1.3923857	1.908842
CE	0.2661898	0.2601242	0.2441420	0.2568685	0.3233174	0.2971247	0.6270243	0.5764860	0.5850637	0.4942380	0.7758479	1.547382
ES	0.1916939	0.3200042	0.2531832	0.1319009	0.3300743	0.0703075	0.5184863	0.9022041	0.6229263	0.7863169	3.5016796	5.905542
GO	0.2176905	0.3483879	0.2527999	0.0684439	0.3973757	0.1467031	0.3987416	0.9765692	0.5113426	0.5654587	3.8210249	4.594483
MA	0.2764346	0.2332406	0.2090125	0.2381383	0.1531469	0.2171332	0.7361820	0.5503729	0.5513519	0.5250387	0.7719272	1.216845
MG	0.2826619	0.3033863	0.2655443	0.1994308	0.2639852	0.1888486	0.6503119	0.6463415	0.5129113	0.9987518	2.3931247	4.170489
MS	0.3190204	0.2482563	0.1009250	0.1364295	0.1179816	0.0113456	0.5379580	0.5117217	0.1716075	1.3010307	2.9534213	1.935996
MT	0.3298514	0.2837901	0.1903284	0.1304654	0.0539198	0.0098894	0.6276595	0.4875049	0.3210173	0.7510992	1.1662202	2.183274
PA	0.2969272	0.2581200	0.4535177	0.2941322	0.1704166	0.1895744	0.6768453	0.5149583	1.3076315	0.7045537	1.1229309	5.986092
PB	0.2969022	0.2442326	0.2489702	0.2138946	0.1643833	0.2586107	0.6295252	0.5206569	0.6652810	0.5217190	0.8046753	1.810729
PE	0.3732831	0.2735069	0.2775122	0.3679612	0.2989712	0.3086742	0.7787653	0.5877365	0.6394805	0.9428690	1.1917596	2.344813
PI	0.2075140	0.2484173	0.2124269	0.1010921	0.1188805	0.1908795	0.4285302	0.5308260	0.5004145	0.3711548	0.6315355	1.075752
PR	0.2399116	0.2711012	0.2234832	0.1403091	0.1321106	0.1123512	0.4510848	0.5472946	0.4463050	1.0227409	2.6632724	4.170069
RJ	0.3115129	0.3257941	0.2900695	0.1226300	0.1485804	0.1684018	0.5898977	0.6935619	0.6379452	1.1729526	3.7872562	6.285261
RN	0.2811511	0.2831483	0.3108020	0.2425520	0.2344113	0.2397931	0.6624175	0.6072345	0.7224893	0.5902310	0.9612989	2.320488
RS	0.2376746	0.2404557	0.2136005	0.1591301	0.1077290	0.1089558	0.4914691	0.4518401	0.4197626	1.3623794	3.6313087	5.762524
SC	0.2739955	0.2522621	0.2405602	0.1751946	0.1002028	0.0799099	0.5060715	0.4801218	0.4517807	0.9862643	2.8064208	6.078812
SE	0.3604251	0.2826886	0.3017720	0.2716327	0.2260486	0.2375880	0.6891165	0.5662628	0.7352121	0.7298072	1.3828947	2.988987
SP	0.2288518	0.2485989	0.2199776	0.0822607	0.1381410	0.0820032	0.4288349	0.4845331	0.4085392	1.1457912	3.6501190	5.146817
TO	0.1523260	0.1627128	0.1876216	0.0158369	0.0219960	0.0300576	0.2920501	0.2838058	0.3211509	0.4220124	0.5463261	1.210739

Source: Author's own elaboration.