

DISCUSSION PAPER

3195

**INTERGENERATIONAL MOBILITY
AND REGIONAL INEQUALITY
IN BRAZIL**

**EDUARDO CENCI
DANIEL LOPES
LEONARDO MONASTERIO**



**INTERGENERATIONAL MOBILITY AND
REGIONAL INEQUALITY IN BRAZIL¹**

EDUARDO CENCI²

DANIEL LOPES³

LEONARDO MONASTERIO⁴

1. Acknowledgments: We thank the seminar participants at the FGV-SP and at the Ninth World Congress of Cliometrics for helpful comments and suggestions. We are also grateful to the anonymous reviewer from IPEA for valuable insights that improved the paper. Any remaining errors are our own.

2. Assistant Professor at the Department of Economics of the University of Utah. E-mail: e.cenci@utah.edu.

3. Researcher at the Industry Observatory of the National Confederation of Industry (CNI). E-mail: daniel.lopes@sesicni.com.br.

4. Researcher at the Department of Regional, Urban and Environmental Studies and Policies of the Institute for Applied Economic Research (Dirur/Ipea); professor in the Master's program in Economics at the Brazilian Institute of Teaching, Development and Research (IDP), in Brasilia; and researcher at the National Council for Scientific and Technological Development (CNPq). E-mail: leonardo.monasterio@ipea.gov.br.

Federal Government of Brazil

Ministry of Planning and Budget

Officer Simone Nassar Tebet

ipea Institute for Applied
Economic Research

A public foundation affiliated to the Ministry of Planning and Budget, Ipea provides technical and institutional support to government actions – enabling the formulation of numerous public policies and programs for Brazilian development – and makes research and studies conducted by its staff available to society.

President

LUCIANA MENDES SANTOS SERVO

Director of Institutional Development

FERNANDO GAIGER SILVEIRA

**Director of Studies and Policies of the State,
Institutions and Democracy**

LUSENI MARIA CORDEIRO DE AQUINO

Director of Macroeconomic Studies and Policies

CLÁUDIO ROBERTO AMITRANO

**Director of Regional, Urban and Environmental
Studies and Policies**

ALEXANDRE DOS SANTOS CUNHA

**Director of Sectoral Studies and Policies,
of Innovation, Regulation and Infrastructure**

PEDRO CARVALHO DE MIRANDA

Director of Social Studies and Policies

LETÍCIA BARTHOLO DE OLIVEIRA E SILVA

Director of International Studies

KEITI DA ROCHA GOMES

Chief of Staff

SHEILA CRISTINA TOLENTINO BARBOSA

**General Coordinator of Press and
Social Communication**

GISELE AMARAL DE SOUZA

Ombudsman: <https://www.ipea.gov.br/Ouvidoria>

URL: <https://www.ipea.gov.br>

Discussion Paper

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

© Institute for Applied Economic Research – ipea 2026

C395 Cenci, Eduardo

Intergenerational mobility and regional inequality in
Brazil / Eduardo Cenci, Daniel Lopes, Leonardo Monasterio. –
Brasília : Ipea, Mar. 2026.

31p. : grafs. – (Discussion Paper ; n. 3195).

ISSN 1415-4765

Inclui referências bibliográficas.

1. Mobilidade Intergeracional. 2. Desigualdade Regional.
3. Sobrenomes. 4. Desenvolvimento Socioeconômico. 5.
Brasil. I. Lopes, Daniel. II. Monasterio, Leonardo. III. Instituto de
Pesquisa Econômica Aplicada. IV. Título.

CDD 330.981

Ficha catalográfica elaborada por Elisângela da Silva Gomes de Macedo CRB-1/1670

How to cite:

CENCI, Eduardo; LOPES, Daniel; MONASTERIO, Leonardo. **Intergenerational mobility and regional inequality in Brazil**. Brasília: Ipea, Mar. 2026. (Discussion Paper, n. 3195). DOI: <https://dx.doi.org/10.38116/td3195-eng>.

JEL: J62; N36.

Ipea publications are available for free download in PDF (all) and ePUB (books and periodicals).

Access: <https://www.ipea.gov.br/portal/publicacoes>

The opinions expressed in this publication are of exclusive responsibility of the authors, not necessarily expressing the official views of the Institute for Applied Economic Research and the Ministry of Planning and Budget.

Reproduction of this text and the data contained within is allowed as long as the source is cited. Reproduction for commercial purposes is prohibited.

CONTENTS

ABSTRACT

SINOPSE

1 INTRODUCTION	6
1.1 Contributions of Güell and other authors and the Informational Content of Surnames	6
1.2 Contributions and related literature	7
2 METHODS	8
2.1 Surnames, ancestry, and ethnicity in Brazil	8
2.2 Estimating the ICS	9
2.3 Comparing ICS estimates	10
2.4 Correlating ICS estimates with current and historical variables	14
3 DATA	14
3.1 Data sources	14
3.2 Descriptive statistics	16
4 RESULTS	18
4.1 Correlations between ICS and current outcomes	18
4.2 Correlations between ICS and historical variables	21
5 DISCUSSION	24
6 CONCLUSION	25
REFERENCES	26
APPENDIX A	29

ABSTRACT

We estimate the long-term intergenerational mobility (IM) in Brazil, by mesoregion, utilizing the methodology developed by Güell, Mora and Telmer (2015) and Güell et al. (2018). Benefiting from the unique features of Brazil, such as data availability on income, education, full names, our study employed a machine learning algorithm to estimate surname-based ancestry. This allows us to bypass limitations of previous studies and provide an initial exploration of the link between intergenerational mobility (IM) and inequality in a middle-income, Latin American country. We also examined how IM spatially aligns with historical factors and current socioeconomic development indicators in Brazil. Our findings connect past land inequality and slavery to lower long-term mobility, which is linked to lower current income per capita and educational attainment. Unlike other IM studies in Brazil, we found no clear geographic patterns. Additionally, there is no distinct relationship between inequality and social mobility, or the Gatsby Curve, at the mesoregional level.

Keywords: intergenerational mobility; surnames; inequality

SINOPSE

Neste trabalho, estimamos a mobilidade intergeracional (MI) de longo prazo no Brasil, por mesorregião, utilizando a metodologia desenvolvida por Güell, Mora e Telmer (2015) e Güell et al. (2018). Beneficiando-nos das características únicas do Brasil, como a disponibilidade de dados sobre renda, educação e nomes completos, nosso estudo empregou um algoritmo de aprendizado de máquina para estimar a ascendência baseada em sobrenomes. Isso nos permite contornar as limitações de outros papers e fornecer uma análise preliminar da ligação entre a MI e desigualdade em um país latino-americano de renda média. Também examinamos como a MI está correlacionada espacialmente com fatores históricos e indicadores atuais de desenvolvimento socioeconômico no Brasil. Nossos resultados relacionam a desigualdade de terras do passado e a escravidão à menor mobilidade de longo prazo, que está ligada à renda per capita e ao nível de instrução mais baixos. Ao contrário de outros estudos de MI no Brasil, não encontramos padrões geográficos claros. Além disso, não há uma relação distinta entre desigualdade e mobilidade social, ou a Curva de Gatsby, em um nível mesorregional.

Palavras-chave: mobilidade intergeracional; sobrenomes; desigualdade.

1 INTRODUCTION

Economic inequality in developing economies, such as Brazil, has been extensively studied. There is a wealth of research that dives deep into its complexity and consequences. However, one aspect that is as significant, if not more so, hasn't received the same attention in Brazil: intergenerational mobility (IM). IM can act as a measure of societal progress and fairness in the long term, yet its examination is underrepresented in Brazil's economic studies. This paper aims to fill this gap.

Most studies on intergenerational mobility use a traditional approach: researchers estimate the correlation between outcomes of parents and their children, then build a measure of IM that goes in the opposite direction of such correlation (higher correlation means lower mobility). This approach has limitations. It often uses information on parents and children only, capturing mobility between no more than two generations. Studies may observe children of rich parents moving to lower classes even in the absence of actual mobility, due to regression to the mean or statistical noise. Another limitation is measuring mobility using occupation status or income, not wealth. Second-generation descendants may earn less and have low-paying occupations, even when they hold high-status positions in society. In other words, the traditional approach tends to overestimate long-term social mobility, while the rich of the past would still be the rich of the future, even though some generations have seen temporary declines. Finally, a more mundane and yet important limitation is data availability. The traditional approach requires data from both parents and their children. Such data must be consistently collected over the space of a generation, which is rare, or use recall methods, which are imprecise.

1.1 Contributions of Güell and other authors and the Informational Content of Surnames

Güell, Mora and Telmer (2015) overcome many of the limitations in the traditional approach by proposing a method that estimates long-term intergenerational mobility using cross-sectional data. If social mobility is low, rare names will be concentrated in certain parts of the income distribution. Conversely, in a society with high mobility, surnames are well distributed through all income levels. The authors apply this method to estimate intergenerational mobility in Spain using the 2001 census of Catalonia and introduce a new measure of such mobility, the Informational Content of Surnames (ICS). The ICS is the difference in the explanatory power of two regressions. The first is the "actual" regression of a socioeconomic outcome like income or education on demographic characteristics and a vector of dummy variables for all existing surnames in the sample. The second is a "fake" regression where the same surname dummies are

shuffled and randomly attributed to the individuals in the sample. The difference in the quality of fit between the two regressions – the difference between their R^2 statistics, specifically – gives a measure of how much information on the distribution of income in a population is captured by the distribution of surnames in that population.¹

In a related work, Güell et al. (2018) took another step and investigate how IM, proxied by the ICS, correlates with various socioeconomic indicators across Italian provinces. The authors find that the ICS varies greatly within Italy and is positively associated with various outcomes. The higher the IM of an Italian province, the better the economic, educational, and social performance. The authors note that given the centralized nature of the Italian government, these correlations are unlikely to result from differentiated local policies.

1.2 Contributions and related literature

We apply the methodology developed by Güell, Mora and Telmer (2015) to estimate IM in Brazil by leveraging information on the joint distribution of rare surnames and economic outcomes. We then investigate the relationship between IM and current regional inequality and other socioeconomic indicators of income, education, health, and social capital, thus reproducing for Brazilian regions the investigation done by Güell et al. (2018) using Italian provinces. Finally, we investigate the relationship between IM and historical variables measuring land inequality, income, and intensity of slavery and immigration, thus connecting past and present indicators of socioeconomic development of different areas in Brazil through their long-term social mobility. Our analysis is carried out at the level of mesoregions, an administrative division between state and municipality, which roughly corresponds, in size and number, to provinces in Italy.²

An issue as significant as measuring long-term IM involves investigating the deep roots of its variation. Although a straightforward causal identification strategy is not feasible, it seems appropriate to explore how IM correlates with key aspects of Brazil's socioeconomic formation. Later in section 4.2, we present the historical background and discuss the results of regressing the Informational Content of Surnames (ICS) on historical variables, which measure the deep-rooted aspects of Brazil's social formation. We improved upon previous studies by exploring unique features of the Brazilian context, namely, the availability of schooling and income outcomes in datasets informing surnames, detailed measures of ancestry and internal migration, and wide regional variation.

1. We provide more detail on the ICS estimation in Güell, Mora and Telmer (2015) and in our context in section 2.

2. Brazil has 5,570 municipalities distributed across 558 microregions, 137 mesoregions, and 27 Federative Units (26 states plus the Federal District).

And we add a new step to the investigation by looking at the correlation between the ICS and different historical factors.

Our study contributes to the broad literature on IM, and to the study of social mobility and regional inequality. In particular, we use a new method to study IM in Brazil, the object of study of several papers on intergenerational social mobility (Pero and Szerman, 2008; Ribeiro, 2012; Britto et al., 2022). Although the methods and mobility estimates in these studies vary, their conclusions align in suggesting that intergenerational mobility in Brazil is low when compared to more developed economies (Ribeiro and Scalon, 2001; Bourguignon et al., 2007; Torche, 2014). Our findings corroborate this conclusion and suggest a link between (poor) current socioeconomic development and (low) social mobility. Interestingly, we do not find a negative correlation between inequality and social mobility as Güell et al. (2018) found in Italy and Britto et al. (2022) found in Brazil. In other words, the “Great Gatsby Curve” (Krueger, 2012) we find is mostly flat.

We also contribute to the literature linking past inequality and institutional quality measures to current outcomes in Brazil (Naritomi et al., 2012; Wigton-Jones, 2020). We find that low IM correlates with historical measures of land inequality and the concentration of enslaved people in the population.

The remainder of our paper is organized as follows. Section 2 details the ICS estimation, provides information on the distribution of surnames in Brazil, and lists the steps taken to measure and compare the ICS across the Brazilian mesoregions. Section 3 presents our data sources, discusses our sample selection criteria, and shows descriptive statistics. Section 4 discusses the results from our investigation of the correlation between ICS and current socioeconomic indicators and historical variables. This section also motivates our historical analysis. Section 5 concludes.

2 METHODS

2.1 Surnames, ancestry, and ethnicity in Brazil

Because surnames reflect ancestry, which in turn reflects ethnicity, race, or migrant status – and those can explain differences in income across individuals and places (Cenci et al., 2021) – it is important that we control for ancestry in our estimations. Otherwise, our measure of the informational content of surnames could be capturing the role of ancestry in explaining the distribution of wages, not IM. This section details how we measure and control for ancestry in our setting.

Like in many Portuguese- and Spanish-speaking countries, children in Brazil are typically given two surnames: the mother's and the father's.³ And they typically pass only one surname to their children (the father's). The mother's surname survives for only one generation and tends to be the second-to-last surname in a person's name.⁴ We leverage this naming structure with two surnames, using the last one (the father's) to measure the ICS and the second-to-last one (the mother's), together with information on race and country of origin, to construct a measure of ancestry and thus control for ethnicity.

Specifically, we use the surname-based classification of ancestries created by Monasterio (2017) and further developed by Lopes et al. (2017) to attribute an ancestry to each person in our data based on their second-to-last surname. Simply put, the classification algorithm matches a person's surname to historical sources where immigrants' surnames are linked to a country or region of origin. These become ancestry groups (e.g., Italian, Japanese, or Eastern European). The large group of Iberian surnames (Portuguese and Spanish) is further divided into categories following the major racial designations in Brazil (e.g., white, black, or mixed). Our analysis excludes all foreign-born (less than 0.5% of individuals in the data) so that our measure of ICS does not reflect current migrant status.

2.2 Estimating the ICS

The paragraphs and equations below illustrate the estimation of the ICS in Güell, Mora and Telmer (2015), which we apply to the Brazilian context and data with minor adaptations. Let y_{is} be the outcome of interest for individual i of surname s , X_{is} be a vector of demographic characteristics, D be the S -vector of actual surname-dummies ($D_s = 1$ if individual i has the surname s), and F be the "fake" surname vector constructed after shuffling surnames in the sample. First, two regression equations are estimated, saving R^2 for each (labeled R_{actual} and R_{fake}). Next, we define the ICS as $\text{ICS} \equiv R_{\text{actual}} - R_{\text{fake}}$, which is a "moment of the joint distribution of surnames and economic well-being that measures the incremental ICS" (Güell, Mora and Telmer, 2015, p. 698). The three equations below summarize the ICS estimation procedure that we repeat for each mesoregion in Brazil.

$$y_{is} = \gamma'X_{is} + b'D + \epsilon_{is} \quad (1)$$

3. The order of surnames differs, however. In Spanish-speaking countries, the father's surname comes first, while in Brazil, Portugal, and other Portuguese-speaking countries, it comes last.

4. Regarding name changes after marriage, until 1977, Brazilian civil law mandated that a married woman adopt her husband's second surname. Later on, adoption became optional, and in 2002, both men and women had the option to adopt their spouse's surname.

$$y_{is} = \gamma'X_{is} + b'F + \epsilon_{is} \quad (2)$$

$$\text{ICS} = R^2_{\text{actual}} - R^2_{\text{fake}} \quad (3)$$

Importantly, the model in Güell, Mora and Telmer (2015) establishes an inverse relationship between ICS and intergenerational mobility (IM). The higher the ICS, the higher the positive correlation between the outcomes of parents and their children, and the lower the economic mobility across generations. We will return to this point when interpreting the correlations found in our empirical analysis.

In our estimations, the outcome y_{is} is the log of the monthly salary and the vector of individual characteristics, X_{is} includes age, age squared, and dummies for educational levels. We note that our inclusion of education controls in the basic specifications that estimate the ICS is an important departure from the estimations in Güell, Mora and Telmer (2015) and Güell et al. (2018), which do not control for education (likely because education is the outcome of interest in the original study). In fact, the range of ICS estimates we find for Brazilian mesoregions is generally smaller than what Güell et al. (2018) find for Italian provinces. However, if we did not control for education, our estimates and theirs are much closer in magnitude (see table A.1 in the appendix A).

Our basic regression does not include the usual female and race dummies because, in line with the literature on long-term mobility, we restrict our analysis to males.⁵ Also, we already use the information on race (and the mother's surname) to control for ancestry in some specifications. Specifically, we estimate and report two measures of the ICS throughout this study. In addition to the ICS, we use the ICS Ancestry, which includes ancestry dummies in equations (1) and (2). The two measures are, in fact, highly correlated ($\rho > 0.96$), as shown in figure A.1 in the appendix A.

2.3 Comparing ICS estimates

The distribution of surnames in a given region is the result of a complex historical process that includes marriage, fertility, and migration. And our measure of intergenerational mobility, the ICS, comes from the potentially different distribution of surnames in each Brazilian mesoregion. Therefore, before comparing the ICS estimates across these mesoregions and drawing conclusions between their correlations with past and

5. As in other countries, social and legal norms regarding surname adoption in Brazil make it hard to link a woman's surname to her ancestry.

DISCUSSION PAPER

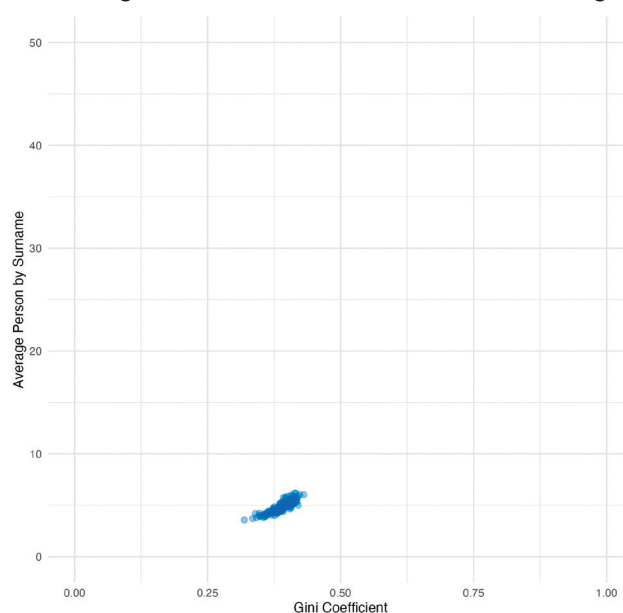
present socioeconomic outcomes, we must ensure that surname distributions across Brazilian mesoregions are similar. If that is the case, we can attribute differences in the ICS estimates to underlying differences in IM.

Following Güell et al. (2018), we take advantage of the fact that surname distribution can be approximated by a Pareto distribution, which, in turn, is fully characterized by two moments: the surname frequency (number of persons per surname, PPS) and the Gini coefficient. We then restrict the sample of individuals in our data to those with surnames whose frequency is below a certain threshold. Finally, we check how these two moments of these restricted surname distributions compare across the Brazilian mesoregions.

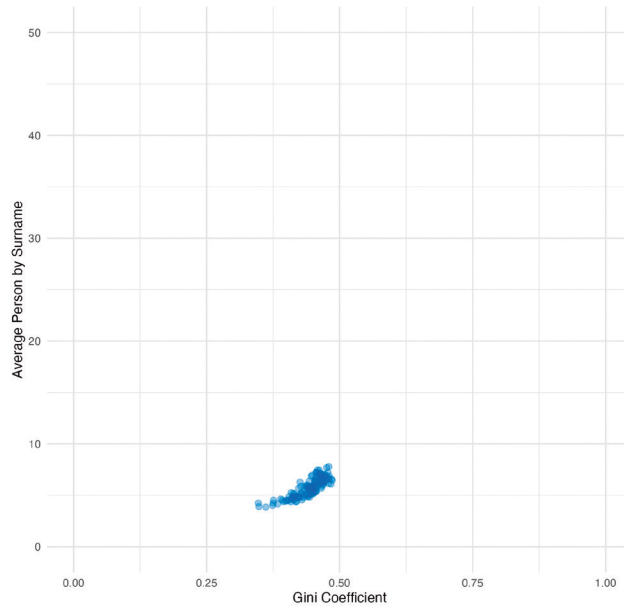
The results of this exercise can be seen in figure 1, which shows the PPS and Gini of the surname distribution for the 137 mesoregions in Brazil. We show the figures for samples with a maximum PPS of 20, 30, 50, or Inf. (no cutoff). Our main study samples use the cutoff 30, which provides a satisfactory balance between comparability and sample size. We note that by restricting our study sample in such a way, we mostly drop individuals with uninformative surnames (because the relatively rare surnames are the ones contributing the most to the estimation of the ICS). Therefore, we do not lose much information. Nonetheless, we calculate the ICS for alternative samples and repeat our main empirical exercises using these different samples and estimates. These are shown in the appendix A.

FIGURE 1**Surname distributions across mesoregions: PPS versus Gini for different PPS cutoffs**

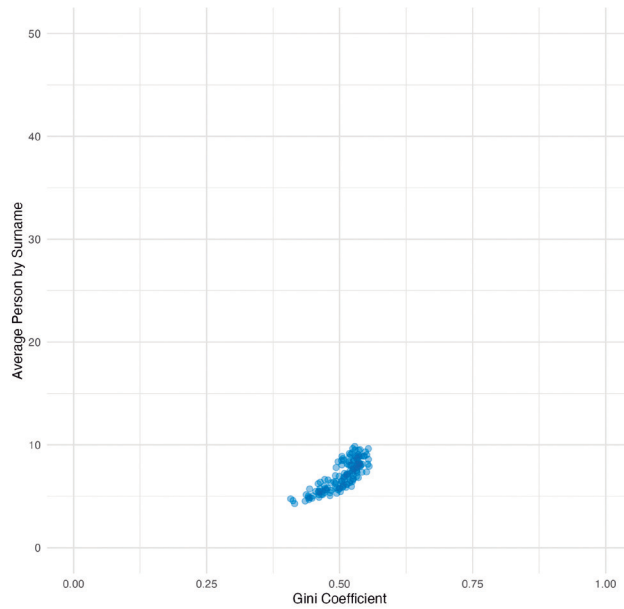
1A – Threshold for the Highest Number of Individuals Sharing the Same Surname: 20



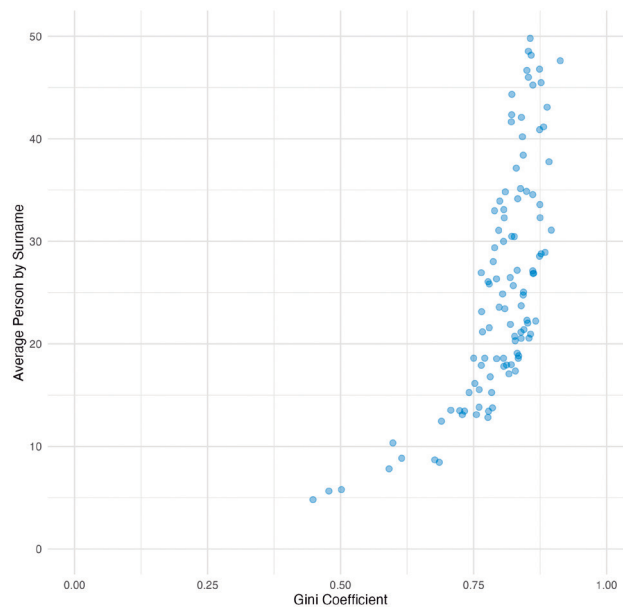
1B – Threshold for the Highest Number of Individuals Sharing Surname: 30



1C – Threshold for the Highest Number of Individuals Sharing the Same Surname: 50



1D – Threshold for the Highest Number of Individuals Sharing the Same Surname: Inf



Authors' elaboration.

Note: Figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

Migration, both foreign and domestic, is another potential confounding factor in our comparison of ICS and socioeconomic outcomes across Brazilian mesoregions. Foreign immigrants are more likely to have unique surnames and distinct economic outcomes than the general population. Domestic migrants, on the other hand, may have more common surnames, but, like foreign immigrants, are likely to be positively self-select (Cenci, 2021). If migration is correlated with ICS by chance, or if it follows consistent regional patterns and alters the ICS directly (e.g., sending regions will likely have lower mobility), then our analysis may be compromised. Fortunately for us, international immigration in Brazil has been low for several decades, and we can identify and exclude all foreign-born from our analysis without losing many observations. Moreover, we can use Census data to calculate immigration and emigration rates at the mesoregion level. We can, therefore, address concerns of migration confounding our results by including internal immigration rates as controls in some specifications, as well as dummies for the five macro-regions in Brazil.⁶

6. A different strategy, implemented by Güell et al. (2018), is to construct a Local Degree measure of surnames and keep in the sample only the individuals whose surnames are common in a given area. We do not use such a strategy in Brazil because we have information on migration flows and because using the Local Degree restriction reduces our number of observations considerably.

2.4 Correlating ICS estimates with current and historical variables

After estimating the ICS for all Brazilian mesoregions – and after establishing that higher ICS means lower social mobility – we can investigate how the differences in IM across regions in Brazil correlate with current income inequality and other socioeconomic indicators. To do so, we regress different socioeconomic outcomes on the ICS estimates. We vary the ICS estimate used (with and without ancestry controls) and include different controls in our specifications (immigration rates, macro-region dummies, and GDP per capita). We weigh all our regressions using the number of individual observations in each mesoregion used in our ICS regressions.

We also investigate the correlation between ICS and different historical variables. In this exercise, however, we do not use the ICS as an explanatory variable to gauge how much socioeconomic development indicators today can be explained by long-term intergenerational mobility. Instead, we flip the variables in the regression using ICS as the dependent variable and the historical measures as the explanatory variables. The purpose of this exercise is to gauge how much of the long-term mobility (or lack thereof) can be explained by historical factors such as land inequality or the intensity of slavery. Our historical analysis requires adjustments because the level of aggregation for historical variables follows municipal boundaries existing in 1872 or 1920. Even though we use mesoregion, an administrative division two levels above the municipality, present and past boundaries do not match. We circumvent this issue by first aggregating historical outcomes at the minimum possible level that matches the current municipality boundaries. These are the Minimum Comparable Areas (MCAs), created by Reis et al. (2008). Next, we attribute the ICS estimate in each of the 137 Brazilian mesoregions to all municipalities in each mesoregion. Then, we regroup these municipalities into their respective MCAs. If an MCA incorporates municipalities from different mesoregions and ICS estimates, we use a weighted average of the ICS using the current population as weights. In the end, we have 916 different MCAs with information on all historical variables and with regrouped ICS estimates.

3 DATA

3.1 Data sources

The primary source of data used in this study is the Annual Social Data Registry (Relação Anual de Informações Sociais – RAIS), a yearly report of all job links submitted by employers in Brazil to comply with labor regulations. These reports are used by the government to manage unemployment benefits and allowances for low-income employees,

as well as to produce statistics on the formal sector, making RAIS a reliable annual census of formally employed workers in Brazil.⁷ The RAIS data has detailed information on workers, their jobs, and their employers. Crucially for our study, the data informs each worker's full name along with their salaries, hours worked, and demographic characteristics such as gender, race, age, and education level.

Our study uses data from RAIS 2019, the most recent available at the time of our analysis. We impose a few restrictions on the individual-level data. As mentioned before, we keep only males. We also keep only those between 25 and 65 years old, have an active labor contract on December 31st, and work 40 hours or more per week. We exclude workers with missing information on their name or race, foreign-born workers, disabled, public servants, and the military. After these restrictions, our data still has over 4.5 million individual observations (unique workers) and over 238,000 unique surnames, which we use to estimate the ICS within each mesoregion. Specifically, we leverage the workers' last surname (the father's) to compute the ICS following equations (1) to (3), and their second-to-last surname (the mother's) to identify and assign to each worker an ancestry dummy to be used as control variable in our regressions.

We supplement our data sources using information from the Brazilian census and household surveys (Pesquisa Nacional por Amostra de Domicílios – PNAD) to obtain indicators of socioeconomic development such as income per capita, educational attainment, and life expectancy. We also used data from the Electoral Justice (TSE) and the Institute for Applied Economic Research⁸ to build variables such as voter turnout and homicide rates.

Finally, our historical data come from various sources. The spatial distribution of slavery is derived from the 1872 Demographic Census. This was the first Brazilian Census to employ reliable methods and achieve comprehensive coverage, encompassing all 643 municipalities existing in the country at that time (Botelho, 2005). At this point, approximately 15% of Brazilians were legally enslaved. Although slavery had lost its relevance in the northeast of Brazil, it continued to play a significant role in the coffee-growing regions of the southeast.

The data on land concentration were derived from the Brazilian 1920 census, which is the oldest source of high-quality, nationwide data on land ownership. We used the

7. In future versions of this study, we intend to expand the data on surnames and individual outcomes using the unified registry of welfare program beneficiaries in Brazil (CadUnico), thus including in our sample workers not employed in the formal sector.

8. Available at: <http://www.ipeadata.gov.br>.

Gini index of land concentration calculated by Wigton-Jones (2020). The 1872 Census provides data on foreigners, but we decided that it would be more appropriate to use the data from the 1920 Census.⁹ Data on GDP per capita by municipality in 1920 comes from Ipeadata.¹⁰

3.2 Descriptive statistics

Table 1 gives an overview of our individual-level data distributed across mesoregions. Our main sample keeps only individuals with surnames whose frequency within the mesoregion (measured by persons per surname – PPS) was smaller or equal to 30. Table A.1 in the appendix A shows similar information for samples determined by different surname frequency cutoffs. We note that the median ICS estimate we find in Brazil is similar in magnitude to what Güell, Mora and Telmer (2015) find in Spain and about half the magnitude of what Güell et al. (2018) find for Italian provinces. However, if we did not control for education in our ICS estimations, our estimates would be much closer in magnitude to those of Güell et al. (2018) (see table A.1, which includes ICS estimates computed without including education dummies as controls).

TABLE 1
Sample Information and Estimated ICS

Variable	Mean	SD	Percentiles		
			10	50	90
Unique Individuals (Obs.)	12,514	21,232	1,406	5,466	34,842
Unique Surnames	2,414	4,264	213	943	6,945
Persons Per Surname (PPS)	5.74	0.928	4.524	5.776	6.982
Gini of Surname Distribution	0.442	0.029	0.405	0.448	0.474
Estimated ICS	0.032	0.018	0.008	0.032	0.054
Estimated ICS Ancestry	0.028	0.017	0.006	0.029	0.049

Authors' elaboration.

Note: Sample information for all 137 Brazilian mesoregions considering a sample of individuals in RAIS 2019 with surnames whose frequency (measured by PPS) was smaller or equal to 30.

9. The quality of the 1890 and 1910 Censuses is considered very low, so they are rarely used. Monasterio and Reis (2007) found no correlation between the presence of foreigners in 1872 and manufacturing activities in 1920, while Reis (2017) argued that the number of foreigners in 1920 impacted regional growth rates between 1920 and 2000.

10. Available at: <http://www.ipeadata.gov.br>.

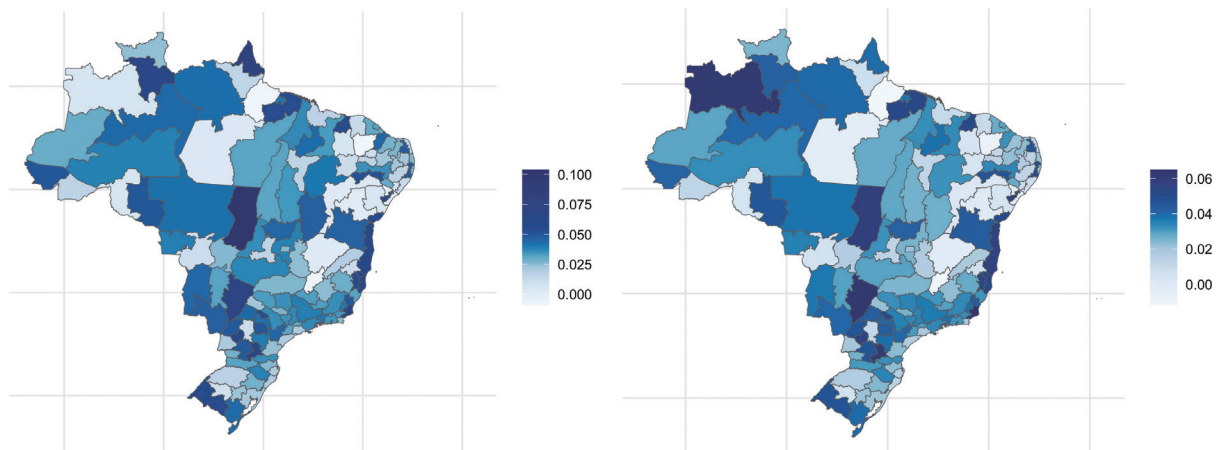
Figure 2 shows the geographic distribution of the ICS across mesoregions in Brazil. Curiously, we do not see any clear pattern along, for example, the poorer regions in the North/Northeast and the richer regions in the South/Southeast.

FIGURE 2

Geographic distribution of ICS and ICS ancestry estimates – Brazilian Mesoregions

2A – ICS

2B – ICS Ancestry



Author's elaboration.

Note: The figure shows ICS estimates all 137 Brazilian mesoregions considering a sample of individuals in our database with surnames whose frequency (measured by persons per surname – PPS) was smaller or equal to 30; the estimation procedure is described in equations (1) to (3) in section 2; the ICS Ancestry includes ancestry dummies as controls in the estimation; figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

Table 2 presents summary statistics of the indicators used as outcomes in our analysis of the correlation between the ICS and current socioeconomic development. Almost all indicators are constructed using data from the last Census in Brazil (2010) by the Atlas Brasil project.¹¹ There is a long list of indicators available at the municipality level in Brazil, which can be aggregated (using population weights) at the mesoregion level for our study. We selected nine indicators that represent some of the most relevant information in four dimensions: income, education, and health. We then complement these with three indicators of inequality, social capital, and crime. We use the Theil index because it is the only inequality measure that could be aggregated from the municipality to the mesoregion level. We obtain similar results using the standard deviation of wages computed from our individual-level data.

11. Exceptions are voter turnout and homicide rates, which are averages from 2012-2018 and 2016-2019, respectively.

TABLE 2
Summary statistics of current socioeconomic indicators

Variable	Mean	SD	Percentiles		
			10	50	90
Income					
(log) Income per capita	6.34	0.46	5.68	6.43	6.86
Poverty (%)	20.33	15.28	4.08	15.34	42.65
Unemployment (%)	7.01	2.17	4.33	6.83	9.61
Education					
Years of education	9.47	0.83	8.56	9.51	10.52
High school graduate (%)	32.05	9.02	19.83	31.84	44
Adult illiteracy (%)	16.83	10.93	5.55	12.89	34.32
Health					
Life expectancy	73.66	2.14	70.47	74.26	76.12
Fertility (children/woman)	2.14	0.47	1.69	2.04	2.68
Infant mortality (per 1,000)	17.98	5.38	12.22	16.22	25.16
Inequality, Social Capital, and Crime					
Theil index	0.52	0.1	0.4	0.51	0.63
Voter turnout (%)	83.61	3.3	79.3	83.8	87.58
Homicide rate (per 100,000)	29.5	15.52	10.89	26.94	51.05

Authors' elaboration.

Note: Summary statistics for all 137 Brazilian mesoregions.

4 RESULTS

4.1 Correlations between ICS and current outcomes

Table 3 shows the results of our regressions of current socioeconomic outcomes on the ICS estimates. All results shown in the table use a sample restricted to individuals whose surname frequency is below 30 persons per surname in the mesoregion. The dependent variables are shown in the rows. The explanatory variable of the results shown in the first three columns is the simple ICS estimate. In the last three columns, the explanatory variable is the ICS Ancestry estimate. All specifications include dummies for the five macro-regions in Brazil (North, Northeast, Southeast, South, and Center-West). Subsequent specifications also add immigration rates and the GDP per capita in 2018 as controls, as noted at the bottom of the table.

DISCUSSION PAPER

TABLE 3
Correlations between ICS estimates and current socioeconomic indicators

Explanatory Variable	ICS (1)	ICS (2)	ICS (3)	ICS Ancestry (4)	ICS Ancestry (5)	ICS Ancestry (6)
Income						
(log) Income per capita	-4.612* (1.795)	-4.474* (1.794)	-3.769** (1.268)	-4.444* (1.926)	-4.349* (1.922)	-3.717** (1.359)
Poverty (%)	-3.846 (34.425)	-0.509 (34.280)	-8.634 (31.131)	-19.087 (36.741)	-16.913 (36.551)	-24.197 (33.150)
Unemployment (%)	4.369 (10.012)	6.555 (9.556)	8.256 (9.094)	2.672 (10.701)	4.102 (10.210)	5.613 (9.719)
Education						
Years of education	-7.409* (3.303)	-7.688* (3.297)	-7.308* (3.241)	-6.937+ (3.545)	-7.112* (3.537)	-6.768+ (3.474)
High school graduate (%)	-95.681* (46.568)	-89.896+ (46.074)	-73.452* (35.379)	-92.527+ (49.895)	-88.626+ (49.273)	-73.895+ (37.811)
Adult illiteracy (%)	-6.186 (31.493)	-8.343 (31.517)	-17.255 (27.225)	-13.367 (33.630)	-14.787 (33.616)	-22.754 (29.018)
Health						
Life expectancy	-3.668 (5.941)	-4.095 (5.943)	-2.311 (5.016)	-2.879 (6.352)	-3.154 (6.347)	-1.557 (5.354)
Fertility (children/woman)	2.878* (1.321)	2.731* (1.311)	2.476* (1.233)	2.544+ (1.419)	2.444+ (1.406)	2.214+ (1.322)
Infant mortality (per 1,000)	-5.884 (12.623)	-5.279 (12.665)	-7.736 (11.920)	-7.916 (13.480)	-7.515 (13.508)	-9.712 (12.707)
Inequality, Social Capital, and Crime						
Theil index	-1.325* (0.519)	-1.154* (0.459)	-0.994** (0.359)	-1.509** (0.553)	-1.395** (0.486)	-1.252** (0.378)
Voter turnout (%)	-0.265+ (0.134)	-0.272* (0.134)	-0.275* (0.134)	-0.251+ (0.144)	-0.256+ (0.144)	-0.259+ (0.144)
Homicide rate (per 100,000)	117.757 (79.186)	122.131 (79.382)	114.798 (78.663)	138.689 (84.443)	141.470+ (84.573)	134.897 (83.765)
Controls						
Region fixed effects	Y	Y	Y	Y	Y	Y
Immigration rate		Y	Y		Y	Y
GDP per capita			Y			Y

Authors' elaboration.

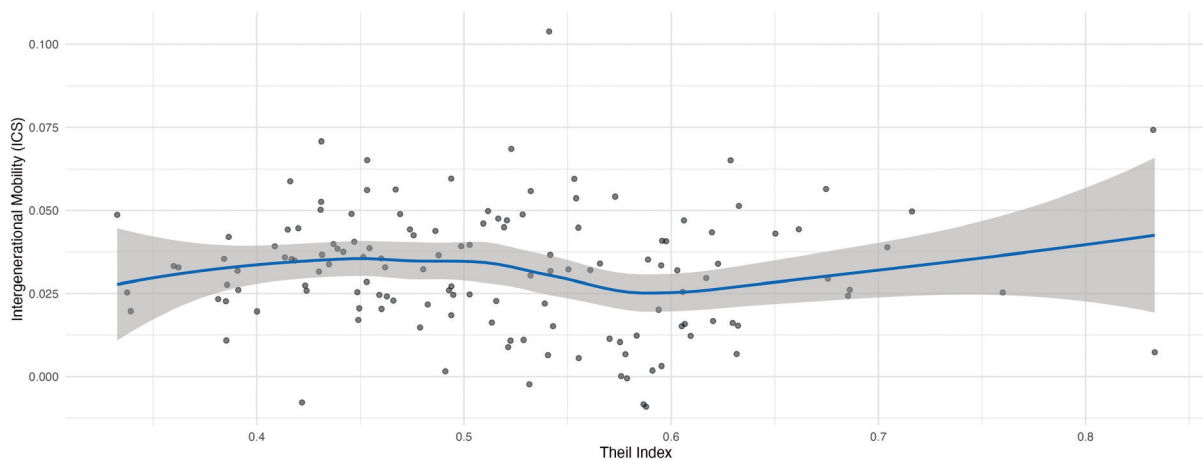
Note: Dependent variables are shown in the rows. The explanatory variable is the ICS estimate in columns one to three, and the ICS Ancestry estimate in columns four to six; the number of observations is N = 137 in all specifications except for the "Voter turnout" outcome, which has N = 136; all regressions are weighted by the number of observations (individuals) used in the ICS estimation in each mesoregion; robust standard errors are in parentheses; symbols denote: + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

We cannot interpret the magnitudes of the coefficients in our regressions directly, but we know that positive coefficients reflect a negative correlation between mobility (as measured by the ICS) and the outcome of interest, whereas negative coefficients show a positive correlation. For example, the coefficients in the first row of table 3 show that, across different specifications, higher ICS is negatively correlated with income per capita. Because higher ICS means lower IM, we conclude that the distributions of income and social mobility at the mesoregion level in Brazil are negatively correlated. This suggested, perhaps intuitively, that places with lower social mobility tend to be also poorer on average. Other significant correlations between indicators of socioeconomic development and intergenerational mobility show up in table 3. Higher ICS (lower IM) is correlated with lower educational attainment and lower rates of high-school graduation. Higher ICS is also associated with higher fertility rates and lower voter turnout.

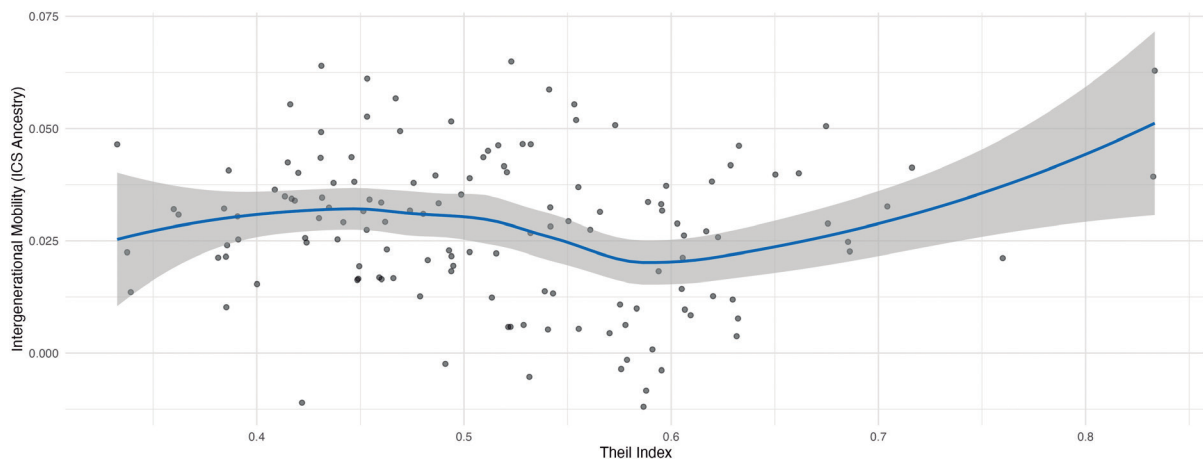
Curiously, we find a negative correlation between ICS (low IM) and current inequality (measured by the Theil index), whereas other studies find a positive correlation between such variables, showing that inequality in the present is correlated with low social mobility (Güell et al., 2018; Britto et al., 2022). By plotting the values for the ICS estimates and Theil index, we can fit a curve that would show the relationship between long-term social mobility and current inequality – the so-called “Great Gatsby curve.” The curve we obtain in our analysis, as seen in figure 3, is mostly flat.

FIGURE 3

The Brazilian Great Gatsby Curve: ICS and ICS ancestry estimates versus Theil index
3A – Gatsby Curve: ICS versus Theil index



3B – Gatsby Curve: ICS Ancestry versus Theil index



Authors' elaboration.

Note: The figure shows the values of Theil index and ICS estimates all 137 Brazilian mesoregions considering a sample of individuals in our database with surnames whose frequency (measured by persons per surname – PPS) was smaller or equal to 30; the estimation procedure is described in equations (1) to (3) in section 2; the ICS Ancestry includes ancestry dummies as controls in the estimation; figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

4.2 Correlations between ICS and historical variables

Traditional Brazilian historiography asserts that the economic formation of the country was based on a tripod of large-scale agricultural property, cash crops for export, and slave labor (Prado Júnior, 1942). Recent studies have shown that even during the colonial period, Brazil's economy was far more intricate and diverse than previously believed. Nevertheless, one cannot deny the profound influence exerted by these colonial institutions (Engerman and Sokoloff, 2002; Naritomi et al., 2012).

Slavery was the most important institution in the economic history of Brazil. From the mid-16th to late 19th century, Brazil was the largest importer of enslaved Africans during the trans-Atlantic slave trade. The economy of colonial Brazil was heavily dependent on enslaved labor, especially in the sectors that propelled the country's development, namely sugar, gold, and, later, coffee production. The institution of slavery not only provided the labor force for the exploitation of these valuable resources, but it also shaped the social and economic structures that still have repercussions today (Laudares and Caicedo, 2022).

Land concentration in Brazil has its roots in the colonial era, particularly in the Portuguese Crown's practice of granting large tracts of land, or "sesmarias," to favored

individuals, religious institutions, or military orders (Dean, 1971). In the Imperial era, the 1850 Land Bill (Lei de Terras) essentially solidified the existing concentration of land ownership; it did not provide for land redistribution or offer affordable means for the less affluent to acquire land. Today, despite the attempts at land reform, in many areas of Brazil, especially those of older occupation, the land tenure structure still reflects the characteristics of the original occupation of these territories.

At the end of the 19th century, a large number of immigrants, often encouraged by government subsidies, arrived in Brazil and transformed its economy. This wave of immigration, particularly in the aftermath of the abolition of slavery in 1888, was a demographic shock that reshaped Brazil's socioeconomic landscape. In a country where a mere 18.7% of the Brazilians over five years old were literate in 1872 (Senra, 2006), human capital was obviously the scarce economic resource. The influx of European and Japanese immigrants, many of whom brought with them a range of skills, became a vital asset for early 20th modernization and industrialization. Indeed, between 1872 and 1920, more than 3.2 million foreigners arrived in Brazil (Levy, 1974). In 1920, 5.1% of the population were foreigners or naturalized. In 1920, only 23% of Brazilians could read or write, while 52% of foreigners were literate by that time (DGE, 1929).¹²

There is a broad consensus that the influx of immigrants positively influenced the overall Brazilian economy (Carvalho Filho and Monasterio, 2012; Rocha et al., 2017; Reis, 2017), yet the implications in terms of social mobility are debatable. On the one hand, a significant number of immigrants ascended the social ladder, assuming roles at the upper echelons of society – positions traditionally reserved for established families. On the other hand, public education in Brazil has historically been subpar, and this lack of quality schooling posed a substantial barrier to social mobility for many. Thus, the very intergenerational transmission of human capital has caused their descendants to form a new elite in Brazil.

With these considerations in mind, we turn to the results of our historical analysis. Table 4 shows the results of the ICS estimate regressed on historical variables. As discussed in section 2, we calculated the ICS for each mesoregion in the present using a sample of individuals whose surname frequency is below 30 persons per surname in the mesoregion. We then assigned this ICS estimate to all municipalities within the microregion and re-aggregated the municipal ICS values, weighted by the present population, attributing them to each of the minimum comparable areas (MCAs) used in our historical analysis. Results using ICS estimates calculated with samples using alternative cutoffs or no cutoff at all are shown in table A.2 in the appendix A.

12. Also, Stolz et al. (2013) have shown that the numeracy skills of immigrants, especially non-Iberians, were much higher than that of Brazilians.

DISCUSSION PAPER

We show correlations between the ICS estimates in the present (the dependent variable) and four historical variables (explanatory variables): the Gini index of land inequality in 1920, the share of foreign-born people in the population in 1920, the share of enslaved people in 1872, and the GDP per capita in 1920. Our table shows results for regressions that include one historical outcome at a time (columns one to four), and all of them together (column five). In column six, we repeat the last specification adding dummies for the five macro-regions as controls.

TABLE 4
Correlations between Historical Variable and ICS Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: ICS						
Land Gini 1920	0.010*** (0.003)				0.009*** (0.003)	0.005+ (0.003)
Foreigners 1920		0.000 (0.000)			0.000 (0.000)	0.002*** (0.000)
Slavery 1872			0.027*** (0.003)		0.029*** (0.004)	0.040*** (0.004)
Ln GDP per capita 1920				0.000 (0.001)	-0.003* (0.001)	-0.003** (0.001)
Observations	928	933	923	941	916	916
R ²	0.013	0.000	0.061	0.000	0.083	0.155
Dependent Variable: ICS Ancestry						
Land Gini 1920	0.009** (0.003)				0.008** (0.003)	0.005* (0.003)
Foreigners 1920		0.000 (0.000)			0.001* (0.000)	0.002*** (0.000)
Slavery 1872			0.030*** (0.003)		0.031*** (0.003)	0.038*** (0.004)
Ln GDP per capita 1920				0.001 (0.001)	-0.002* (0.001)	-0.003** (0.001)
Dependent Variable: ICS Ancestry						
Observations	928	933	923	941	916	916
R ²	0.011	0.002	0.084	0.003	0.101	0.148
Region fixed effects						Y

Note: Explanatory variables are shown in the rows. The dependent variable is the ICS estimate in panel A and the ICS Ancestry estimate in panel B; the ICS for each MCA is attributed in the following way: (1) estimate the ICS of each mesoregion, (2) attribute this estimate to all municipalities in the mesoregion, and (3) attribute the average ICS (weighted by population) of all municipalities within an MCA to that MCA; all regressions are weighted by the number of observations (individuals) used in the ICS estimation in each MCA.; robust standard errors are in parentheses; symbols denote: + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Once more, we cannot draw direct conclusions from the magnitudes of the estimates due to the nature of our dependent variable, the ICS. We can, nonetheless, interpret the direction of such coefficients. What these coefficients show is a consistent association between past inequality measured by the Gini of land distribution in 1920 and higher ICS estimates today, that is, low estimates of long-term intergenerational mobility. This suggests that places in Brazil that started off as more unequal in the past offered fewer chances of social mobility in the following decades through today.

Notably, we see the same pattern with slavery. The places in Brazil that had a high concentration of enslaved people are more likely to correspond to areas with high ICS estimates (low IM) today. We also find some evidence that GDP per capita in the past is negatively correlated with the ICS, whereas the share of foreign-born is positively correlated, but these estimates are less consistent across specifications.

5 DISCUSSION

Our analysis of intergenerational mobility in Brazil reveals important patterns linking historical conditions to current socioeconomic outcomes. The historical context of Brazil suggests that unequal access to education, healthcare, and formal labor markets may have created persistent barriers to upward mobility, particularly in regions with historical land inequality and higher concentrations of formerly enslaved populations.

Areas with higher land concentration in 1920 and higher shares of enslaved people in 1872 show lower intergenerational mobility today. These historical conditions likely created institutional barriers to accessing quality education, healthcare, and formal employment that persisted across generations. As shown by Wigton-Jones (2020), areas with high land concentration in Brazil historically invested less in public goods such as schools and healthcare facilities.

The observed pattern – where low mobility correlates with lower income, education, and life expectancy – may therefore be understood as the result of historically limited access to opportunity-enhancing institutions and services. This interpretation is consistent with migration patterns observed in Brazil, where regions offering fewer opportunities experienced selective outmigration, potentially further reducing the average mobility of non-migrants who remained in these areas.

Our approach has limitations for causal inference. When examining correlations between ICS and socioeconomic indicators, we face potential endogeneity concerns. Lower intergenerational mobility could cause lower socioeconomic outcomes, but the reverse is also possible – regions with poor economic performance may offer fewer

opportunities for mobility. Additionally, unobserved regional characteristics could influence both mobility and economic outcomes.

We addressed these concerns in several ways. First, by including macro-region fixed effects, we control for broad regional differences in development patterns. Second, by including immigration rates as controls, we account for selective migration that might affect both surname distributions and economic outcomes. Third, our use of historical variables from 1872 and 1920 provides temporal precedence to potential causes. While these approaches strengthen our analysis, we acknowledge that definitive causal claims would require additional research designs beyond our study.

6 CONCLUSION

This study is the first attempt to estimate long-term social mobility in Brazil using surnames. By applying the method proposed by Güell, Mora and Telmer (2015) and utilizing a database that encompasses all workers in Brazil's formal sector, we were able to estimate the ICS across different geographies. Our contribution to the literature is further enhanced by the incorporation of an algorithm that enables the identification of surname ancestry, thereby allowing us to account for this characteristic in the estimation of the ICS.

We show that ICS is associated with various contemporary socioeconomic indicators. It is inversely correlated with income per capita, years of schooling, and voter turnout, even after accounting for full controls, and positively correlated with fertility rates. Furthermore, our regressions show a negative correlation between ICS (the inverse of intergenerational mobility) and current inequality (measured by Theil's index), a result contrary to most studies.

Our findings also deviate from typical patterns found in similar studies. We did not observe a clear geographic pattern based on common regional divisions such as those found in Italy (Güell et al., 2018) or Brazil (Britto et al., 2022). Our estimate of the Great Gatsby curve, which usually shows a positive correlation between inequality and intergenerational immobility, revealed a mostly flat trend.

We make a novel contribution by investigating the correlation between the ICS estimates and several historical variables in Brazil. These historical regressions are relevant for two reasons. Firstly, the results are so robust that they suggest that the analysis of the distribution of surnames is indeed capturing a social phenomenon. In this sense, the results support the idea that the method introduced by Güell, Mora and Telmer (2015) has a more general application, and can also be applied in middle-income

countries with historical trajectories very different from those of European countries. Secondly, the results indicate that slavery and land concentration over 100 years ago are correlated with social immobility in Brazil. Future studies will need to investigate these potential causal mechanisms more deeply, examine their validity, and elucidate the precise ways in which they influence the current socioeconomic landscape in Brazil.

For regions seeking to improve social mobility prospects, our results suggest that historical constraints continue to influence current outcomes, but targeted interventions in education access, labor market formalization, and political inclusion might help overcome these historical legacies. However, the persistence of immobility patterns in certain regions indicates that substantial structural changes may be necessary to break long-standing barriers to mobility.

REFERENCES

- BOTELHO, T. R. Censos e construção nacional no Brasil Imperial. **Tempo Social, revista de sociologia da USP**, v. 17, n. 1, p. 321-341, June 2005.
- BOURGUIGNON, F.; FERREIRA, F. H. G.; MENENDEZ, M. Inequality of opportunity in Brazil. **Review of Income and Wealth**, v. 53, n. 4, p. 585-618, Dec. 2007.
- BRITTO, D. G. *et al.* **Intergenerational mobility in the land of inequality**. Bonn: Iza, Sept. 2022. (IZA Discussion Paper, n. 15611).
- CARVALHO FILHO, I. D.; MONASTERIO, L. Immigration and the origins of regional inequality: government-sponsored European migration to southern Brazil before World War I. **Regional Science and Urban Economics**, v. 42, n. 5, p. 794-807, Sept. 2012.
- CENCI, E. **The wage premium of migrants and return migrants**: internal migration in Brazil. [S. l.: s. n.], 2021. (Working paper).
- CENCI, E.; LOPES, D. A. F.; MONASTERIO, L. M. Internal migration and the spread of long-term impacts of historical immigration in Brazil. [S. l.: s. n.], 2021. (Working paper).
- DEAN, W. Latifundia and land policy in nineteenth-century Brazil. **Hispanic American Historical Review**, v. 51, n. 4, p. 606-625, Nov. 1971.
- DGE – DIRECTORIA GERAL DE ESTATÍSTICA. **Recenseamento do Brazil**: realizado em 1 de setembro de 1920 – população. Rio de Janeiro: Typ. da Estatística, 1929.
- ENGERMAN, S. L.; SOKOLOFF, K. L. Factor endowments, inequality, and paths of development among new world economics. **Economía**, v. 3, n. 1, p. 41-109, 2002.

GÜELL, M.; MORA, J. V. R.; TELMER, C. I. The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. **The Review of Economic Studies**, v. 82, n. 2, p. 693-735, Apr. 2015.

GÜELL, M. *et al.* Correlating social mobility and economic outcomes. **The Economic Journal**, v. 128, n. 612, p. F353-F403, Jul. 2018.

KRUEGER, A. B. The rise and consequences of inequality in the United States. **Speech at the Center for American Progress**, v. 12, n. 3, 2012.

LAUDARES, H.; CAICEDO, F. V. Tordesillas, slavery and the origins of Brazilian inequality. *In: ECONOMICS OF CULTURE AND INSTITUTIONS, 2022, Cambridge, Estados Unidos. Proceedings...* Cambridge, Estados Unidos: NBER, 2022.

LEVY, M. S. F. O papel da migração internacional na evolução da população brasileira (1872 a 1972). **Revista de Saúde Pública**, v. 8, n. suppl., p. 49-90, June 1974.

LOPES, D. A. F.; SILVA FILHO, G. A.; MONASTERIO, L. Culture, institutions and school achievement in Brazil. *In: ENCONTRO NACIONAL DE ECONOMIA, 45., 2017, Natal. Anais...* Niterói: Anpec, 2017.

MONASTERIO, L. Surnames and ancestry in Brazil. **PloS One**, v. 12, n. 5, 2017.

MONASTERIO, L. M.; REIS, E. The roots of regional inequalities in Brazil, 1872-1920. *In: ECONOMIC HISTORY ASSOCIATION MEETINGS, 67., 2007, Austin. Proceedings...* Austin: Economic History Association, 2007.

NARITOMI, J.; SOARES, R. R.; ASSUNÇÃO J. J. Institutional development and colonial heritage within Brazil. **The Journal of Economic History**, v. 72, n. 2, p. 393-422, June 2012.

PERO, V. L.; SZERMAN, D. Mobilidade intergeracional de renda no Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 38, n. 1, p. 1-36, Apr. 2008.

PRADO JÚNIOR, C. **Formação do Brasil contemporâneo**. São Paulo: Brasiliense, 1942.

REIS, E. Historical perspectives on regional income inequality in Brazil, 1872-2000. *In: BÉRTOLA, L.; WILLIAMSON, J. (ed.). Has Latin American inequality changed direction?* Cham: Springer International Publishing, 2017. p. 143-170.

REIS, E. *et al.* **Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000**. Rio de Janeiro: Ipea/Dimac, 2008.

RIBEIRO, C. A. C. Quatro décadas de mobilidade social no Brasil. **Dados**, v. 55, n. 3, p. 641-679, 2012.

RIBEIRO, C. A. C.; SCALON, M. C. Mobilidade de classe no Brasil em perspectiva comparada. **Dados**, v. 44, n. 1, p. 53-96, 2001.

ROCHA, R.; FERRAZ, C.; SOARES, R. R. Human capital persistence and development. **American Economic Journal: Applied Economics**, v. 9, n. 4, p. 105-136, Oct. 2017.

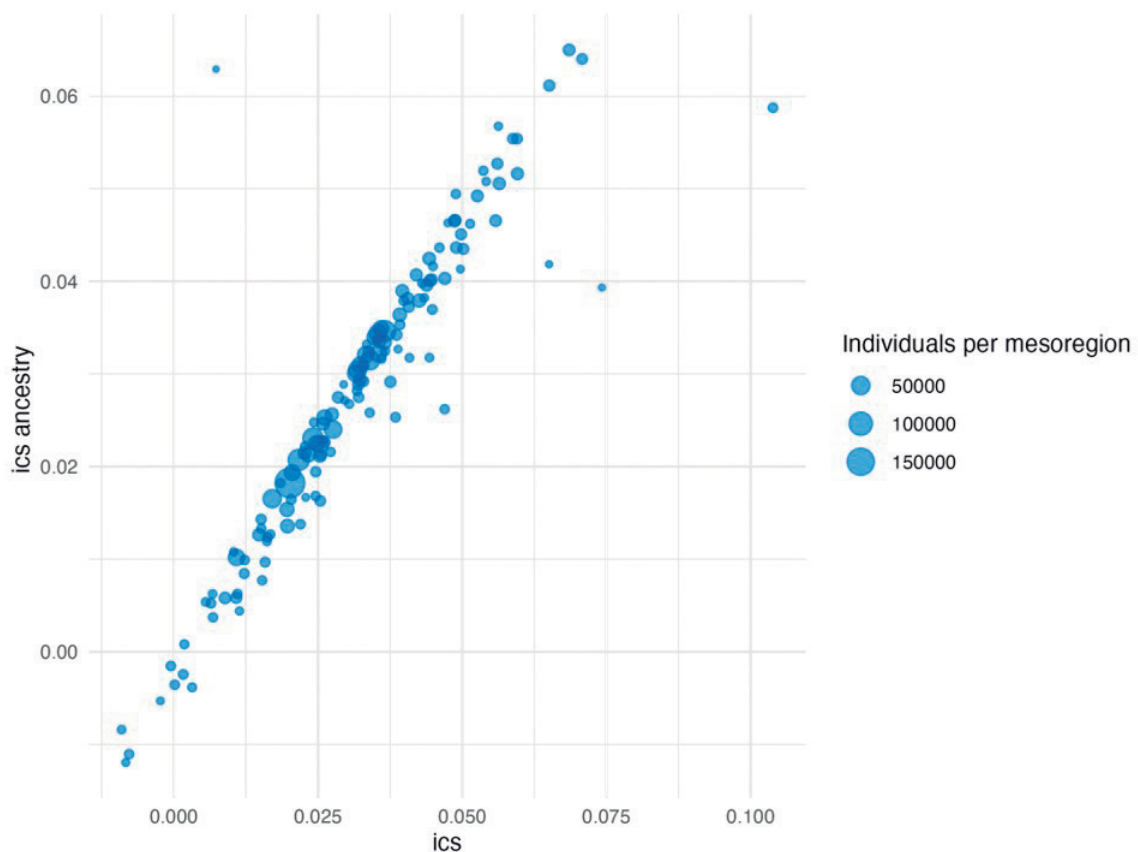
SENRA, N. C. **História das estatísticas brasileiras**. Rio de Janeiro: IBGE, 2006.

STOLZ, Y.; BATEN, J.; BOTELHO, T. Growth effects of nineteenth-century mass migrations: “Fome Zero” for Brazil? **European Review of Economic History**, v. 17, n. 1, p. 95-121, Feb. 2013.

TORCHE, F. Intergenerational mobility and inequality: the Latin American case. **Annual Review of Sociology**, v. 40, p. 619-642, 2014.

WIGTON-JONES, E. Legacies of inequality: the case of Brazil. **Journal of Economic Growth**, v. 25, n. 4, p. 455-501, 2020.

APPENDIX A

FIGURE A.1**Scatter plot of Informational Content of Surnames versus ICS ancestry estimates**

Note: ICS estimations follow equations (1) to (3) in section 2 and use data from a sample of individuals in Annual Social Data Registry (Relação Anual de Informações Sociais – RAIS) 2019 with surnames whose frequency (measured by persons per surname – PPS) was smaller or equal to 30; the ICS Ancestry includes ancestry dummies as controls in the estimation; the correlation between the two measures is above 0.96; figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher’s note).

TABLE A.1**Sample information and estimated ICS: alternative sample cutoffs**

Variable	PPS ≤ 20 - Mean	PPS ≤ 20 - SD	PPS ≤ 30 (main) - Mean	PPS ≤ 30 (main) - SD	PPS ≤ 50 - Mean	PPS ≤ 50 - SD	All (no cutoff) - Mean	All (no cutoff) - SD
Unique Individuals (Obs.)	10,657	18,529	12,514	21,232	15,017	24,513	94,672	226,606
Unique Surnames	2,339	4,157	2,414	4,264	2,478	4,348	2,621	4,500
Persons Per Surname (PPS)	4.867	0.598	5.74	0.928	7.05	1.429	36.928	24.98
Gini of Surname Distribution	0.388	0.021	0.442	0.029	0.505	0.033	0.824	0.085
Estimated ICS	0.035	0.02	0.032	0.018	0.028	0.016	0.013	0.011
Estimated ICS Ancestry	0.03	0.02	0.028	0.017	0.025	0.015	0.012	0.01
Estimated ICS (no education)	0.063	0.03	0.058	0.026	0.051	0.024	0.024	0.011

Note: Sample information for all 137 Brazilian mesoregions; the first three groups of columns consider samples of individuals in RAIS 2019 with surnames whose frequency measured by the number of persons per surname (PPS) was smaller than or equal to 20, 30, or 50. The last two columns consider a sample of all individuals in RAIS 2019. In all cases, the samples consider only males between 25 and 65 years old with an active labor contract on December 31st that work 40 hours or more per week; they exclude workers with missing names or race, foreign-born, disabled, public servants, and military; the ICS in the last row is estimated without including education and ancestry dummies in the log salary regressions.

DISCUSSION PAPER

TABLE A.2
Correlations between historical variables and ICS estimates, alternative sample cutoffs

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Sample (PPS cutoff):	20	20	30	30	None	None
Dependent Variable: ICS						
Land Gini 1920	0.012*** (0.003)	0.007* (0.003)	0.009*** (0.003)	0.005+ (0.003)	-0.005*** (0.001)	-0.003*** (0.001)
Foreigners 1920	0.000 (0.000)	0.002*** (0.000)	0.000 (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Slavery 1872	0.033*** (0.004)	0.047*** (0.004)	0.029*** (0.004)	0.040*** (0.004)	-0.003* (0.001)	0.001 (0.001)
Ln GDP per capita 1920	-0.003* (0.001)	-0.004** (0.001)	-0.003* (0.001)	-0.003** (0.001)	-0.002*** (0.000)	-0.002*** (0.000)
Observations	916	916	916	916	916	916
R ²	0.093	0.185	0.083	0.155	0.302	0.452
Dependent Variable: ICS Ancestry						
Land Gini 1920	0.011*** (0.003)	0.008** (0.003)	0.008** (0.003)	0.005* (0.003)	-0.002+ (0.001)	0.000 (0.001)
Foreigners 1920	0.001* (0.000)	0.002*** (0.000)	0.001* (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.002*** (0.000)
Slavery 1872	0.036*** (0.004)	0.045*** (0.004)	0.031*** (0.003)	0.038*** (0.004)	0.000 (0.001)	0.003** (0.001)
Ln GDP per capita 1920	-0.002* (0.001)	-0.003** (0.001)	-0.002* (0.001)	-0.003** (0.001)	-0.002*** (0.000)	-0.002*** (0.000)
Observations	916	916	916	916	916	916
R ²	0.114	0.170	0.101	0.148	0.169	0.352
Region fixed effects		Y		Y		Y

Note: Explanatory variables are shown in the rows. The dependent variable is the ICS estimate in panel A and the ICS Ancestry estimate in panel B. The samples used in each regression are identified by their PPS (persons per surname) cutoffs: 20 (columns one and two), 30 (columns three and four), or none (columns five and six). The cutoff 30 is the default used in the paper. For more details, see the notes in table A.1; the ICS for each MCA is attributed in the following way: (1) estimate the ICS of each mesoregion, (2) attribute this estimate to all municipalities in the mesoregion, and (3) attribute the average ICS (weighted by population) of all municipalities within an MCA to that MCA; all regressions are weighted by the number of observations (individuals) used in the ICS estimation in each MCA; robust standard errors are in parentheses; symbols denote: + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Head of the Publishing Department

Denise Pimenta de Oliveira

Assistants to the Head of the Department

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

Supervision

Alice Souza Lopes

Typesetting

Aeromilson Trajano de Mesquita

Augusto Lopes dos Santos Borges

Cristiano Ferreira de Araújo

Daniel Alves Tavares

Danielle de Oliveira Ayres

Leonardo Hideki Higa

Cover design

Aline Cristine Torres da Silva Martins

Graphic design

Aline Cristine Torres da Silva Martins

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea – Brasilia

Setor de Edifícios Públicos Sul 702/902, Bloco C

Centro Empresarial Brasilia 50, Torre B

CEP: 70390-025, Asa Sul, Brasília-DF

Ipea's mission
To enhance state decision-making and public debate.



ipea Institute for Applied
Economic Research

MINISTRY OF
**PLANNING AND
BUDGET**

