Universidade Federal de Pernambuco

Centro de Ciências Sociais Aplicadas

Departamento de Economia

Programa de Pós-graduação em Economia

Intergenerational Mobility in the Land of Inequality

Lucas Warwar

Recife

2023

Lucas Warwar

Intergenerational Mobility in the Land of Inequality

Dissertação apresentada ao Programa de Pós-graduação em Economia do Departamento de Economia da Universidade Federal de Pernambuco como requisito parcial para a obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Breno Ramos Sampaio

Co-orientador: Dr. Diogo Britto

Recife

2023

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

W299i Warwar, Lucas
Intergenerational mobility in the land of inequality / Lucas Warwar. - 2023.
120 folhas: il. 30 cm.

Orientador: Prof. Dr. Breno Ramos Sampaio.
Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, CCSA, 2023.
Inclui referências e apêndices.

1. Mobilidade intergeracional. 2. Desigualdade. 3. Programas de transferência condicional de renda. I. Sampaio, Breno Ramos (Orientador). II. Título.

336 CDD (22. ed.) UFPE (CSA 2023 – 027)

Lucas Warwar

Intergenerational Mobility in the Land of Inequality

Dissertação apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Mestre em Economia.

Aprovado em: 23 de Março de 2023

Banca Examinadora:

Prof. Dr. Breno Ramos Sampaio (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Joana Naritomi
London School of Economics

Prof. Dr. Gianluca Violante
Princeton University

Recife

2023

ABSTRACT

The dissertation comprises two papers on social mobility in Brazil. The first (Chapter 1) is coauthored with Diogo Britto, Alexandre Fonseca, Paolo Pinotti, and Breno Sampaio and provides the first estimates of intergenerational mobility of income (IGM) in Brazil using administrative registries and income tax data. Besides characterizing mobility in great detail, the chapter also proposes new approaches to deal with informal income measurement – a central issue to study mobility in developing countries but until now largely neglected by the literature, focused on developed countries. The chapter presents a snapshot of the distribution of opportunities for children born in Brazil at the end of the 1980s, depicting higher income persistence and lower mobility than in developed countries.

In turn, Chapter 2 is a solo-authored draft of ongoing work which pretends to proceed from a snapshot to a film. Specifically, it presents the first results of a project aiming to assess the effect of large social programs on IGM. In particular, the chapter uses similar populational datasets to the previous paper to estimate the long-term impacts of Programa Bolsa Família (PBF), the largest conditional cash transfer (CCT) program in the world. Despite preliminary, results suggest that children who had exposure to PBF accumulate more years of education, display better labor market outcomes, and have lower rates of teenage pregnancy, social welfare receipt, and incarceration. In doing so, PBF promoted social mobility and contributed to a decrease in regional inequality within the country.

Keywords: Intergenerational Mobility; Inequality; Conditional Cash Transfers.

RESUMO

A dissertação compreende dois artigos sobre mobilidade social no Brasil. O primeiro (Capítulo 1) é escrito em coautoria com Diogo Britto, Alexandre Fonseca, Paolo Pinotti e Breno Sampaio, e fornece as primeiras estimativas de mobilidade intergeracional de renda (IGM) no Brasil, utilizando registros administrativos e dados fiscais de renda. Além de caracterizar a mobilidade em grande detalhe, o capítulo propõe novas abordagens para lidar com a mensuração de renda informal - uma questão central para o estudo da mobilidade em países em desenvolvimento, mas até agora amplamente negligenciada pela literatura, focada em países desenvolvidos. O capítulo apresenta um retrato da distribuição de oportunidades para crianças nascidas no Brasil no final dos anos 80, retratando uma maior persistência de renda e menor mobilidade do que em países desenvolvidos.

Por sua vez, o Capítulo 2 é um rascunho escrito individualmente de um trabalho em andamento que pretende passar de um retrato a um filme. Especificamente, apresenta os primeiros resultados de um projeto que tem como objetivo avaliar o efeito de grandes programas sociais em mobilidade social. Em particular, o capítulo utiliza conjuntos de dados populacionais semelhantes ao artigo anterior para estimar os impactos de longo prazo do Programa Bolsa Família (PBF), o maior programa de transferência condicional de renda (TCC) do mundo. Apesar de preliminares, os resultados sugerem que crianças que tiveram exposição ao PBF acumulam mais anos de educação, apresentam melhores resultados no mercado de trabalho e têm taxas mais baixas de gravidez na adolescência, recebimento de assistência social e encarceramento. Ao fazer isso, o PBF promoveu a mobilidade social e contribuiu para uma diminuição na desigualdade regional dentro do país.

Palavras-chave: Mobilidade Intergeracional; Desigualdade; Transferência Condicional de Renda.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1 INTERGENERATIONAL MOBILITY IN THE LAND OF INEQUALITY

1.1 Introduction

Intergenerational mobility (IGM) is a long-standing interest in social sciences and the public debate. The extent to which children's opportunities are determined by their parents' income is a relevant question from both equity and efficiency perspectives. Moreover, evidence on the actual degree of IGM can shift preferences for redistributive policies, as recently shown by (2).

However, accurately measuring IGM is challenging, as it requires linking incomes across generations for a large number of individuals. This was recently made possible by the increasing availability of population registry data for the US and a few other high-income countries (e.g., see 3; 4; 5; 6; 7; 8; 9). However, similar data are typically not available for medium- and low-income economies where, in addition, labor informality is more widespread. At the same time, understanding IGM is arguably most relevant in such contexts, which are typically characterized by high inequality and poverty, and weak welfare systems.

In this paper, we estimate IGM in Brazil, a large developing country displaying extreme inequality in socioeconomic conditions; in 2019, the Gini index was as high as 0.53, the 9$^{\text{th}}$ highest worldwide. Previous evidence of a positive relationship between inequality and intergenerational persistence of income – the "Great Gatsby Curve", as labeled by (10) – suggests that social mobility may be particularly low in Brazil.

Indeed, there is extensive evidence of Brazilian families holding fortunes and positions of power for centuries, such as the "*Quatrocentão*" ("four-hundred-years") families in the State of São Paulo. These families descend directly from the Portuguese colonizers who landed in the 16th century to search for gold mines and trade indigenous people as slaves. In turn, early genealogical studies by (11) and (12) portray the persistence of elite families over centuries in Brazil. However, we know very little about the actual extent of intergenerational income mobility in Brazil beyond this largely anecdotal evidence.

To conduct our analysis, we combine rich individual-level data from multiple population-wide administrative registries and large-scale surveys, including unique tax and residential address data supplied by the Brazilian tax authority for the first time for research purposes. In addition to estimating IGM at the national level, these data allow us to document in detail how income mobility varies by groups and fine geographical units, and to study the role of causal place effects for upward mobility.

Importantly, we develop several methods to address two key challenges in the estimation of IGM in Brazil that are common to other developing countries. First, family links for the entire population are not readily available in any single registry, for which reason we combine different administrative datasets to link parents to children at the individual level. Second, a significant portion of the Brazilian economy is informal, and informal income is not reported in administrative registries (as of 2019, almost one in two Brazilian workers are employed in the informal sector, see 13). To address this challenge, we train machine learning (ML) models on rich, large-scale survey data to impute informal income in our main sample.[1] Our main dataset eventually provides total income – the sum of formal and informal income – for a large, representative sample of 1.3 million children born between 1988-1990 as well as both of their parents.

We use these data to estimate linear regressions of the percentile income rank of children on the rank of their parents in the national income distribution. Following (3), we focus on the slope coefficient of the rank-rank regression as our main measure of (inverse) *relative mobility*: the steeper the slope, the lower the relative mobility. We also calculate the expected rank of children born to below-median income parents, which we use as our main measure of *absolute mobility.*

The estimated slope coefficient equals 0.55, meaning that a 10 percentile increase in parental income is associated with an average 5.5 percentile increase in child income during adulthood. There is stronger persistence at the very top of the income distribution, portraying the success of rich Brazilian families in maintaining their economic status across generations. In terms of absolute mobility, children born to below-median income parents are expected to reach the 35th income percentile in adulthood. A transition matrix between parental and child income quintiles shows that only 2.5% of children born to parents in the bottom quintile surge to the top quintile, and only 4% of those born to parents in the top quintile fall to the bottom quintile. In turn, almost one in two children born in the bottom and top quintiles remain at the same quintiles when adults. We show that our main results are unaffected by several robustness checks that address measurement and estimation issues, notably sample selection, attenuation, and life-cycle

[1] We use a similar approach to impute formal non-labor income – namely, formal income other than labor such as dividends and capital gains – for time periods when tax data are not available.

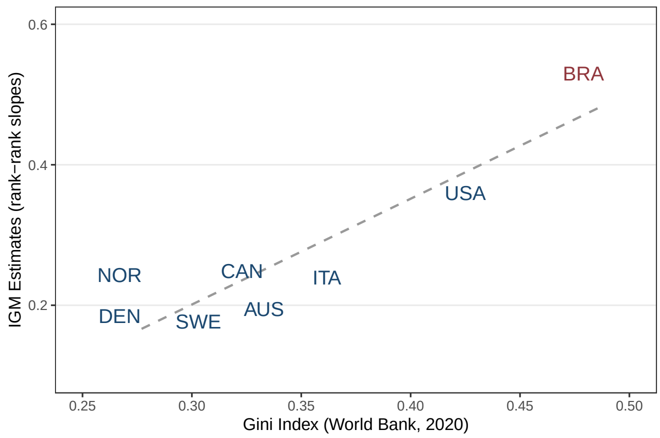
bias.

Our methodology is crucial for obtaining these results. We show that relying exclusively on administrative data results in a much flatter rank-rank regression, whereby the slope equals .35, similar to US mobility levels (3). Such attenuation reflects measurement error due to administrative data neglecting informal income for a large share of individuals at the bottom of the distribution and also, in some years, formal non-labor income (i.e., dividends and other types of capital income). In addition, we provide evidence that our data-driven predictions for income unreported in administrative data are highly accurate. In particular, a simulation exercise indicates that any measurement error arising from the imputation process leads only to a quantitatively small attenuation bias in income persistence.

Our main findings are also robust to using two alternative, novel approaches that rank parents and children without the need to impute unobserved income components. The first method exploits the fact that Brazilian workers move very frequently between formal and informal jobs (14; 15). Accordingly, this approach ranks them on the average income earned during long periods of formal employment, which is precisely recorded in administrative employment data, as a measure of individual-specific "productivity". The second approach ranks parents and children on a "neighborhood-based"income measure, defined by the average (formal) income across 300 thousand census tracts, leveraging data on more than 500 million residential addresses. The rationale for this second approach is that residential choices strongly correlate with socioeconomic status, particularly in highly unequal contexts such as the Brazilian one. Each method has advantages and disadvantages, but the rank-rank curve estimated using either of these two approaches is largely consistent with the one obtained using our baseline method. Importantly, these alternative approaches may be viable in other contexts characterized by a paucity of data on informal income.

Although comparisons of IGM estimates across countries must be interpreted with caution (5; 16), social mobility in Brazil appears to be much lower than in any other country for which similar estimates are available. In particular, the rank-rank slope is estimated at 0.34 in the US (3), and ranges from 0.19 to 0.22 in Australia, Canada, Italy, and Scandinavian countries (4; 5; 6; 7; 8; 9). Figure Figura 1 plots these estimates against the Gini index of income inequality. The graph conveys two important facts: first, both intergenerational persistence of income and inequality are much higher in Brazil than the other (richer) countries; and second, Brazil lies perfectly on the Great Gatsby Curve depicted by other countries.[2]

Figura 1: The Great Gatsby Curve



Notes: The figure plots the relationship between the Gini index (horizontal axis) and relative mobility (vertical axis) using this paper's estimates for Brazil and available rank-based mobility estimates for developed countries. The latter are obtained from (7) (Australia), (6) (Canada), (9) (Denmark), (4) (Italy), (5) (Norway), (8) (Sweden), and (3) (US).

Within Brazil, upward mobility widely varies based on individual characteristics, notably gender and race. A girl born to below-median income parents ranks on average 14 percentiles below boys born with the same parental income, and this gap is unaffected when we restrict the comparison to siblings. In turn, whites rank on average 7 percentiles above non-whites with the same parental income, and the gap is larger for below-median income families. While these results are broadly in line with previous evidence on differences in IGM by race in the US (18; 19), they are all the more remarkable in the context of Brazil, where non-whites are not a minority group but instead represent about half of the population.

We also document a large degree of assortative mating in Brazil. For instance, college-educated individuals are about 3.5 times more likely to marry other graduates relative to a counterfactual scenario in which mating is fully random. In turn, assortative mating plays an important role for household-level IGM. If mating was indeed random, absolute and relative mobility would increase by 16% and 42%, respectively, in our sample.

We also explore the association of parental income with a broad array of long-term

[2]See (17) for a discussion on the factors driving the relationship between inequality and income mobility. In line with previous evidence for the US and Italy (see, respectively, 3; 4), we also document a within-country Great Gatsby Curve, as mobility is inversely correlated with income inequality across Brazilian areas.

outcomes. Children born to poorer parents are at higher risk of dying at an early age, experiencing teen pregnancy, being the victim of a crime, and receiving social assistance from the government. In turn, the same children receive less formal education, and are less likely to be firm owners and find employment in prestigious occupations (e.g., as a lawyer or doctor). These results uncover patterns of welfare persistence across generations along several dimensions beyond individual income.

Finally, we show that social mobility dramatically varies across the vast and heterogeneous Brazilian territory, unveiling a mobility divide between the wealthier Center-South regions and the poorer Northern regions. Southern regions colonized by European immigrants in the late XIX century and Center-Western regions that recently experienced a "soy boom-driven"economic growth exhibit the highest degrees of upward mobility. At the opposite, children born to below-median income parents in the Amazon or Northeast semi-arid regions face the worst outcomes in the country.

When correlating absolute mobility with an array of local characteristics across Brazilian regions, education quality has by far the strongest predictive power, followed by indicators on family structure, demographics (including the racial composition), household characteristics and the local infrastructure. These factors partially overlap with those found for the US and Italy as key mobility predictors (4; 20). Nonetheless, in Brazil education quality stands out as the single most important factor. Although entirely correlational, this analysis may offer insights for an emerging body of literature studying mobility determinants (e.g., see 21; 22; 23).

Motivated by these stark regional divides, we estimate causal place effects on absolute mobility leveraging variation in age at move among the children of migrating families (24). Movers converge linearly to the income of permanent residents in the destination area at a rate of 2.4% per year of childhood exposure, meaning that children moving at birth to a place where they are expected to rank 10 percentiles higher will increase their rank by 5.76 percentiles on average due to causal place effects.[3] Hence, these effects explain more than half of the regional differences in absolute mobility across Brazil. These results are unaffected when comparing siblings moving at different ages, thus absorbing selection into migration and other family-effects. In turn, overidentification tests show that movers' outcomes converge to those of permanent residents from their same

[3] In this analysis, we measure income at the age of 24. Hence, exposure from birth to the age of 24 implies a $24 \times 2.4\% = 57.6\%$ convergence.

cohort, gender, and race, providing compelling evidence that we indeed identify causal place effects.

We contribute to a recent body of literature estimating IGM using rich administrative registries in different countries. Starting with the seminal paper by (3) in the US, this literature has focused exclusively on rich countries, mainly due to data constraints (2, review existing estimates for five developed economies).[4] We are the first to measure income mobility in a large developing country characterized by high inequality and widespread poverty leveraging detailed population-wide data. This setting allows us to characterize mobility patterns across different groups and fine geographical areas, and to study several long-term outcomes beyond income. Instead, previous analyses for developing countries have largely been based on survey data, or they have used education as a proxy for income due to data limitations and the prevalence of informality (e.g., see 1; 29; 30; 31; 32; 33; 34; 35).[5]

Our second contribution is thus methodological, as we devise new approaches to overcome such measurement challenges. In particular, we show how supervised ML can be used to predict informal income, we propose simulation exercises to quantity potential biases due to remaining measurement error, and we develop two novel methods for ranking parents and children on economic status without the need to impute informal income. These tools can be adapted to estimate IGM in other countries – including many developed economies – that are also characterized by a large informal sector.[6]

Finally, we contribute to the literature studying the impact of places on social mobility (e.g. 20; 24; 38) and – more broadly – long-term outcomes (e.g. 39; 40; 41). Our estimates of causal place effects – the first ones available for a developing country – confirm that heterogeneity of opportunities across areas explains a large portion of the total variation in intergenerational mobility, in line with previous evidence from the US and other rich countries.[7]

The remainder of the paper proceeds as follows. Section 1.2 briefly introduces the

[4](25) and (26) review previous studies relying mainly on household surveys, while (27) and (28) consider alternative approaches.

[5]On the relationship between informality and economic development, see, e.g., (36) and (15).

[6](37) estimate that, during the period 1991-2015, one fourth of Italian GDP is produced in the informal economy, and the size of the informal sector accounts for as much as 15% of GDP in countries like Canada, Denmark, Norway, and Sweden.

[7]The review by (42) shows that evidence on place effects is largely confined to the US and a few other rich countries, with (43) being a notable exception in South Africa.

Brazilian context, followed by Section 2.2 describing our data, family linkage and income measurement methods, and our main IGM measures. Section 1.4 presents our main IGM estimates at the national level and by subgroups. We explore geographic variation in mobility in Section 1.5 and estimate causal place effects in Section 1.6. Finally, Section 1.7 concludes.

1.2 Institutional Background

Brazil is the fifth largest country in the world by area and the sixth by population size, hosting nearly one-third of the population in Latin America. The country comprises 27 states and 5,570 municipalities. It has historically been characterized by extreme socioeconomic inequality. In 1990 – roughly the period when our cohorts of children were born – the Gini index was as high as 0.60, placing Brazil as the fifth most unequal country in the world, and the first one outside Africa. Although inequality has subsequently followed a mildly decreasing trend, the Gini index remained as high as 0.53 in 2019. According to official estimates, the top 10% of the population holds 43% of the country's income (13), compared to 31% in the US, 29% in China, and around 25% in European countries.[8]

The country's colonial past, characterized by the boom and bust of short-spanned extractive economic cycles and over 350 years of slavery, bestowed strong regional and social disparities. For instance, GDP per capita is about 40% lower in the Northern regions relative to the more developed Center-South, and the gap in income per capita between white and non-white households is over 35%. Non-whites represent nearly half of the population but account for 64% of the unemployed, 67% of the incarcerated population, and 75% of the beneficiaries of *Bolsa Família* cash transfers. The homicide rate ranges from above 50 per 100k inhabitants in poorer states such as Roraima and Ceará to below 12 in the richest states such as São Paulo and Santa Catarina. These facts further motivate an analysis of mobility across subgroups and geographical areas.

Like in most low- and middle-income countries, the labor market is characterized by a large degree of informality. Labor turnover is also very high, with 70% of formal jobs lasting less than a year, and it is common for workers to turnover between the formal and informal sector (14; 15). In line with this, 82.8% of men in our data have held at least

[8] Estimates based on the World Bank's Poverty and Inequality Platform (44).

one formal job over their lifetime, but about 40% of workers are employed in the informal sector in a given year. For the purpose of our analysis, it is thus crucial to appropriately measure informal income.

The bulk of income taxes in Brazil is collected on formal labor income, although around half of formal workers are fully exempted from filing income taxes because they earn below the first tax bracket (BRL 22,847 in 2019).[9,10] For the same reason, most informal workers would not pay taxes even if they had an official contract, since the majority of them earn below the first tax bracket. Dividends are fully exempted from income taxes.[11]

Individual income taxes are exclusively levied by the federal government and marginal tax rates range from 7.5% to 27.5%. Tax filings are mandatory for individuals with earnings above the first tax bracket, for all firm owners and for all individuals with any capital gains, any stock market operations, or property wealth above BRL 300,000.[12] Individuals filing taxes must report all (formal) income sources, including tax-exempted ones.

1.3 Data and measurement

Following the recent literature (e.g. 3; 4; 45), we measure social mobility by the relationship between the income rank of children and their parents'. In turn, this approach requires (i) linking one or more cohorts of children to parents at the individual level, and (ii) measuring their individual income. Constructing such data for Brazil faces two main challenges, which are common in the context of developing countries. First, comprehensive registries of family links (of the type available, e.g., for Scandinavian countries) are not readily available for Brazil. Second, a large portion of income is earned in the informal economy and, as such, it is not reported in administrative registries. We next describe how we overcome these challenges by combining several sources of individual-level data to recover family relationships, and training supervised ML models on large-scale survey data to impute informal income.

[9] Throughout the paper, we refer to BRL at 2019 prices. In 2019, the purchasing power parity rate was 2.28 relative to the US dollar.

[10] For instance, in 2015 only slightly more than 27 million tax forms were filled in a universe of over 60 million formal workers.

[11] For simplicity, throughout the paper we refer to all types of withdrawals by firm owners as dividends.

[12] Starting in 2010, a small share of firm owners receiving dividends below 40,000 BRL were no longer required to file taxes.

1.3.1 Family links

We aim to link each child's unique person code (*CPF*) to their parents'. Our starting point is dependent claims in individual tax returns data for the 2006-2020 period, provided by the Brazilian tax authority (*Receita Federal do Brasil*). Parents report children aged 0-24 for the purpose of tax deductions, in which case we can directly link them to each other through the unique person codes available in these data.[13] However, only one-third of Brazilians – mainly in the upper part of the income distribution – file taxes every year (unlike in the context of rich countries, where much larger shares of the population file taxes). Therefore, we rely on additional data sources to link children who are not claimed by their parents in the tax data.

We link unclaimed children to their mothers using the Brazilian person registry (*Cadastro de Pessoas Físicas*), which covers the entire population and is provided by the Brazilian tax authority. All individuals are identified by their person code, full name, and mother's full name. If the mother can be uniquely identified by her name – as is the case for 52% of Brazilians – we link the child's person code to her mother's based on the mother's name.[14] Since fathers' names are not available in the person registry, we rely on a welfare registry (*Cadastro Único*) to link children to their fathers. The registry covers around two-thirds of the Brazilian population and contains the father's name for all individuals, along with person codes.[15] Since it provides the informative basis for administering social programs such as *Bolsa Família*, the registry mainly covers the low and middle parts of the income distribution. We implement the same procedure as before, linking children to their fathers conditional on the father having a unique name in the country, so that we can precisely identify his person code.

Overall, 49% and 25% of the children of the 1988-1990 cohorts can be linked to their mother and father, respectively.[16] Our main sample is defined by 1.34 million children who can be linked to both parents, accounting for around 15% of the entire 1988-1990

[13] Children aged 22-24 can only be reported if they are enrolled in technical school or higher education.

[14] The share of individuals with a unique name in the country is large because Brazilians typically carry one or more surnames from both their parents. These individuals are easily identified in the person registry as the latter includes both names and person codes. (46) show that individuals with unique names do not strongly differ from the overall population along several characteristics.

[15] We combine yearly snapshots of this registry for the 2011-2020 period, with 135.6 million individuals in total.

[16] Using younger cohorts reduces the period in which we can measure their income as adults, whereas using earlier cohorts reduces our sample because tax data on dependent claims starts in 2006. Nevertheless, we show in Appendix A.2.1.1 that our main findings remain similar when using additional cohorts.

cohorts.

In Appendix A.1.1, we report the share of children from each cohort who can be linked to each parent using the procedure described above and present an extended sample based on a less conservative linkage procedure, which we use for robustness. In Appendix A.1.2, we show that our main sample is fairly representative of the population in terms of several individual characteristics, given that our procedure relies on complementary data sources covering different parts of the income distribution, namely tax registry on the upper part of the income distribution, person registry on the entire distribution (for mothers), and welfare registry on the lower and middle part of the distribution (for fathers). We also show that our main findings are robust to using additional cohorts and re-weighting the sample to eliminate any remaining differences in characteristics between our working sample and the general population.

1.3.2 Income

Our baseline estimates of social mobility rank children and parents on total income, defined as the sum of formal and informal income. Accounting for informal income is crucial given the large size of the informal labor market (about 40% of all jobs). For this purpose, we develop a novel approach leveraging rich, large-scale survey data and ML methods to estimate income that cannot be readily observed in administrative data.

Formal income. Tax records cover all sources of formal income, including both labor and non-labor components, earned by an individual in a given year.[17] However, tax data are not always available for two reasons: first, only about one-third of Brazilians file taxes each year; and second, tax data are available from 2006 onwards, limiting our ability to measure parental income until our main sample cohorts – born in 1988-1990 – are aged 16-18.[18]

Whenever tax records are unavailable for a given individual in a given year, we measure formal income as the sum of a labor and non-labor component. The first component – formal labor income – is directly available from administrative employment data covering the population of formal jobs for the 1985-2019 period (*Relação Anual de In-*

[17]This also includes tax-exempted income, since all sources of (formal) income must be reported in tax filings.

[18]We show that our results remain similar when measuring parental income only for years when tax data are available (Section A.2.1).

formações Sociais, RAIS).[19] The second component – formal non-labor income – includes dividends, rents, interests, and capital gains, which are not available in administrative registries other than tax data. We thus follow an imputation procedure to predict formal non-labor income leveraging survey data sources. The procedure is the same one used to input informal income, which we describe next.

Informal income. While the Brazilian administrative registries allow us to accurately measure formal income, they do not contain – by their very nature – any information on informal income. We measure the latter using individual-level data from two large-scale surveys: the *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a large longitudinal household survey covering about 400,000 individuals per year for the 1992-2019 period; and Population Census surveys covering 10% of the population in 1991, 2000, and 2010. Both surveys are collected by the Brazilian Institute of Geography and Statistics (IBGE), which has a long tradition of measuring informal income for the purpose of estimating national aggregates that account for the large unofficial sector.

We impute informal income based on a rich array of individual characteristics available in both administrative registries and the survey data. This is a typical prediction problem, for which reason we use random forests (RF), a supervised ML algorithm that endogenously splits the space of covariates to generate predictions for a given outcome. Specifically, we implement the algorithm developed by (49), see Appendix A.1.3 for details.

We grow a separate RF to predict informal income in each year from 1991 to 2019 using survey data as training datasets. The vector of predictors includes a wide array of individual characteristics: state of residence (27), a dummy identifying metropolitan regions, gender, age, race (white vs. non-white), education dummies (4), and occupation category (dummies for formal worker, formally self-employed, and firm owner, while informal workers are the residual category).

After training the model, we predict informal income for all individuals in our main sample, including formal workers and owners who may earn part of their total income in the informal sector. We repeat the same process for estimating formal non-labor income, which is necessary for measuring total formal income when tax data are not available. Appendix A.1.3 provides details on the estimated model. The error in predicting income

[19] RAIS has been extensively used in previous research on the Brazilian labor market, see e.g. (47) and (48).

out-of-sample is very small, in line with similar exercises in the literature (e.g., 50).

Nevertheless, the prediction exercise naturally entails some degree of measurement error that may bias our mobility estimates. In Section 1.4.1, we propose a novel simulation exercise to quantify such potential bias. This exercise supports the notion that measurement error leads only to attenuation bias – consistent with classical measurement error – and that such bias is quantitatively small.

Main sample. In the main analysis, we measure the income of children born in 1988-1990 in the period 2015-2019, when they are 25-31 years old, and relate it to the income of their parents at the time when children were 3-18 years old.[20] For both children and parents, we focus on the average annual income throughout the measurement period. The median parental and child annual income is BRL 47,068 and BRL 19,730, respectively, while the share of total income held by the top decile is around 40% for both populations. Table Tabela 1 displays descriptive statistics for the full sample and separately by gender and race.

Figure Figura 2 shows that the distribution of total income in our sample matches well the distribution of total income in the PNAD survey. This is a remarkable result, given that the latter distribution relies exclusively on survey data for a smaller, representative sample of the population, while we measure total income for a much larger number of individuals using administrative data on formal income and data-driven methods to impute other sources of income.

1.3.3 Alternative approaches to rank incomes

Imputing informal income is crucial for correctly ranking children and parents in their respective distributions, yet any imputation method entails in principle some degree of measurement error. We thus introduce two alternative, novel metrics that rank parents and children on economic status while waiving the need for imputing income.

Productivity-based measure. The key insight motivating our first approach is that

[20] Three years old is the earliest age at which we can measure parental income for our oldest cohort born in 1988, since survey data in the format that we use is available from 1991. In turn, we measure child income setting a five-year window as late as possible. Income data from the tax authority cover children in our main cohorts over the period 2015-2019, and their parents over the period 2006-2010. In addition, RAIS data are available until 2019. In Appendix A.2.1, we provide robustness exercises showing that our main results are not affected by life-cycle bias and alternative income definitions.

Tabela 1: Income Distribution Statistics

	Parents			Children		
	5%	50%	95%	5%	50%	95%
All	16,044	47,068	249,468	8,515	19,730	102,068
Males	9,509	31,997	193,521	10,604	22,226	117,414
Females	5,202	13,046	64,874	7,736	16,005	87,762
White	19,906	53,931	282,546	10,419	22,274	111,560
Non-white	13,797	32,917	187,921	7,388	16,267	81,658

Notes: The table reports the income at the 5^{th}, 50^{th}, and 95^{th} percentiles of both parents and children. The first row refers to the entire sample, while the other rows present separate statistics by gender and race. In the columns for "Parents", the entries for "Males"and "Females"report individual incomes of fathers and mothers, respectively, while all other entries report household income. The columns for "Children"always report individual income. All amounts are displayed in 2019 BRL.

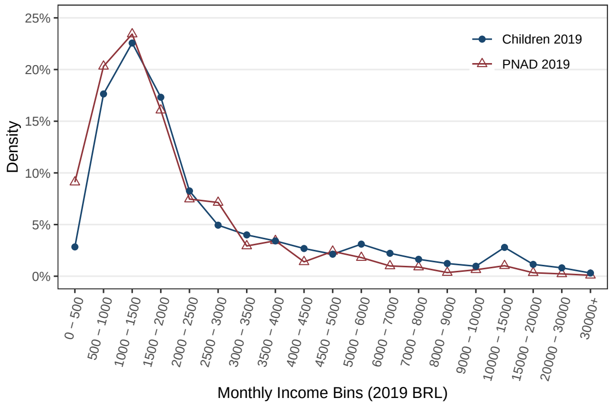
a large share of individuals in Brazil frequently turnover between the formal and informal sector. In particular, about 80% of individuals in our main sample hold at least one formal job throughout their career. We thus rank individuals based on the average income earned during employment spells in the formal labor market as a measure of their individual-specific productivity. The underlying assumption behind this approach is that the average productivity during employment spells in the formal sector – as measured by formal earnings – is a reasonable proxy for individual productivity when employed in the informal sector. The key advantage of this productivity-based method is that it only requires high-quality data on formal jobs.

Neighborhood-based measure. Our second approach ranks parents and children based on the average income in the census tract in which they reside. Census tracts are small geographical areas designed to cover homogeneous groups of about 400 families throughout the country. The rationale for this measure is that residential choices are strongly correlated with income, particularly in poorer countries characterized by high inequality and socioeconomic spatial segregation. In addition, neighborhoods have a major impact on living standards and access to opportunities, as determined by access to public goods, job opportunities, and exposure to violence (e.g., see 51; 52). In our data, variation between census tracts explain 29% of total variation in formal labor income. Another important advantage of this measure is that it may better capture the high living standards of individuals benefiting from inherited wealth or living on in-kind and informal family donations.

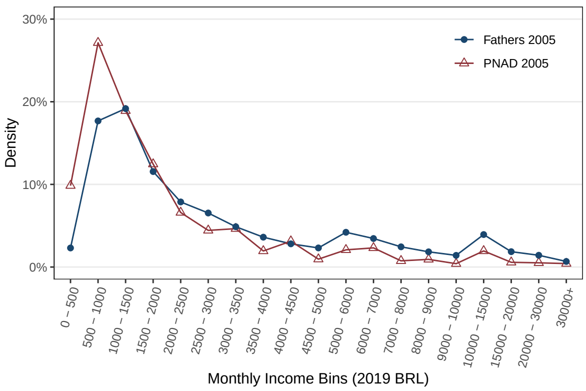
To implement this strategy, we geocode unique data from the Brazilian tax authority that track residential addresses for the entire population in the 2000-2020 period – a total of more than 500 million addresses – and assign them to a census tract using shape files provided by IBGE.[21] We then measure the average income in each location as the average labor income of residents holding formal jobs.

Figura 2: Income Distribution

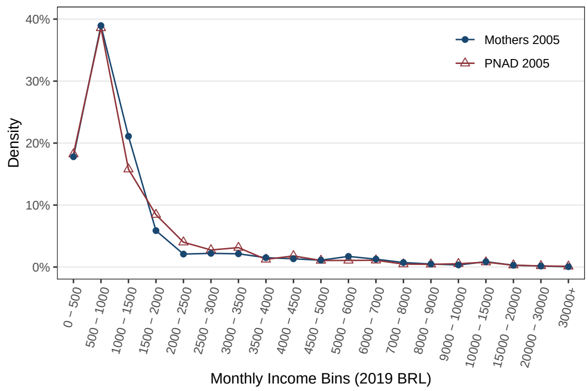
(a) Children



(b) Fathers



(c) Mothers



Notes: The graphs compare the income distribution in our main sample (blue dots) and in the PNAD survey (red triangles), separately for children, fathers, and mothers. The PNAD sample includes "children" born in 1988-1990 that were interviewed in the wave 2019, and "parents" of children born in 1988-1990 that were interviewed in the wave 2005. PNAD sample weights are used to compute the income distribution in PNAD data.

1.3.4 Mobility measures and other outcomes

Following the recent literature (e.g., 3; 4), we focus on the relationship between children and parents' income ranks, as originally proposed by (45). Since this relationship tends to be linear, it can be summarized by a few statistical parameters that can be

[21] Specifically, we consider the place of residence for children in 2000, when they are aged 10-12, and the place where they live when adults in 2019.

compared across areas and groups. Specifically, we estimate the following linear regression:

$$y_i = \alpha + \beta p_i + \epsilon_i \tag{1.1}$$

where y_i and p_i are, respectively, the income percentile rank of child i and her parents' at the national level, ordered from 1 to 100. Child ranks are measured relative to their own cohorts, and parents' ranks are measured relative to other parents with children from the same cohorts.

The estimated parameters in equation (1.1) provide us with two IGM measures. The slope coefficient β measures the (inverse) *relative mobility* of children born to parents who are 1 percentile apart in the parental income distribution. A higher β means a wider gap between the two, thus implying lower IGM. In a perfectly mobile society, the rank-rank slope would equal zero as children's long-term outcomes would be unrelated to parental income.

The intercept α equals the expected rank for children at the bottom of the parental income distribution. Combining α and β, one can recover the expected rank for children born at any point of the income distribution. Following previous literature (e.g., 3), we focus on the expected rank of children born in below-median income families as our main measure of *absolute mobility*, which we also refer to as *upward mobility* throughout the paper. In turn, the latter equals the expected rank for children whose parents are in the 25^{th} percentile of the income distribution (i.e. $\alpha + 25 \times \beta$). This measure is particularly useful to characterize geographical variation in mobility patterns, as it compares the outcomes of children born in different regions of the country while holding constant parental income.

In addition, we construct transition matrices from parental income quintiles to child income quintiles. In particular, we focus on the chances of *escaping poverty* – defined as the probability that children born to parents in the bottom quintile do not belong to the same quintile when adults –, and on the probability that children move from the bottom to the top quintile of the income distribution within one generation (53).

Finally, we estimate intergenerational income elasticities (IGE), defined by the correlation of children's and parents' log incomes. IGE allows for a comparison with earlier survey-based studies in Brazil and – more generally – earlier studies of social mobility (see, e.g., 26; 32; 54). Some disadvantages of IGE compared to our main measures are that (i) the log-log relationship between parental and child income is typically non-linear, so that it cannot be easily summarized by a few parameters and compared across different areas and groups; (ii) it cannot easily accommodate zero income; and (iii) it is more strongly affected by life-cycle bias (55).

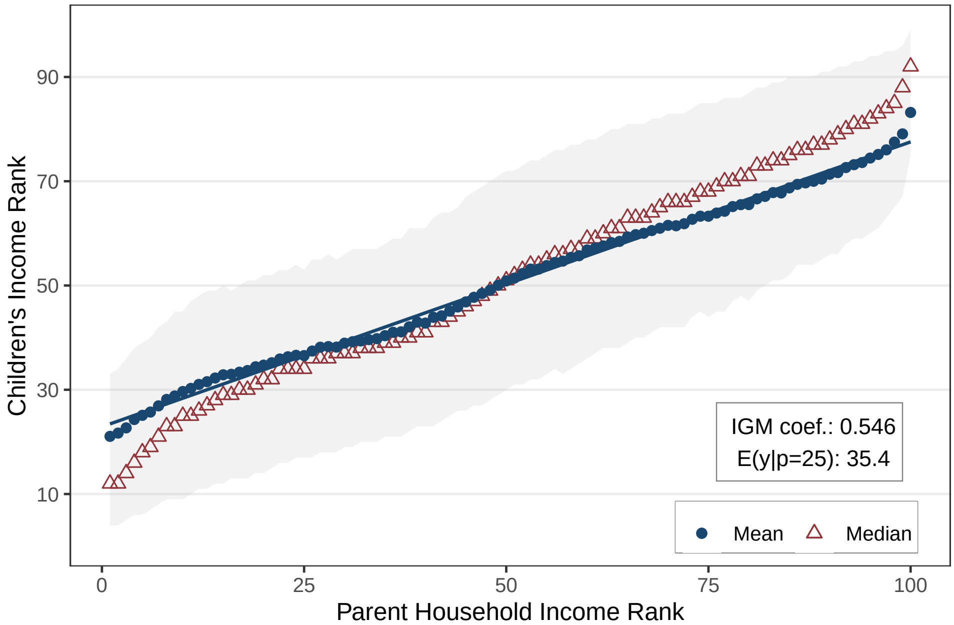
We also document the association between parental income and several children's long-term outcomes beyond income, namely education, access to prestigious occupations, victimization, mortality, and teenage pregnancy. In Appendix A.1.4, we describe all data sources used in the paper and how we link them to our main sample.

1.4 Income mobility at the national level

1.4.1 IGM estimates

Figure Figura 3 plots the average and median income rank in adulthood for children born to parents in each percentile of the parental income distribution, along with the inter-quartile range (i.e., the range between the 25th and 75th percentiles of the children's income distribution). The ranks are based on our main measure of total income, described in Section 1.3.2. The rank-rank relationship is approximately linear, with the exception of the very top percentiles of the distribution, which exhibit a steeper slope. Although similar patterns have been documented for Canada, Denmark, Italy, Norway, Sweden, and the United States (3; 5; 56), for the case of Brazil the change in slope is more concentrated at the very top of the distribution.

Figura 3: Baseline Mobility Curve in Brazil



Notes: The figure shows the relationship between parental and child income ranks at the national level, for our main sample (1988-1990 cohorts). For each parental income percentile, it plots the mean (blue dots), median (red triangles) and inter-quartile range (shaded area) of child income rank during 2015-2019, i.e. at the age of 25-31. Parental income is the sum of the father's and mother's average income when children are aged 3-18 years old. The figure also displays our absolute ($\alpha + \beta * 25$) and relative mobility (β) measures based on Equation (1.1).

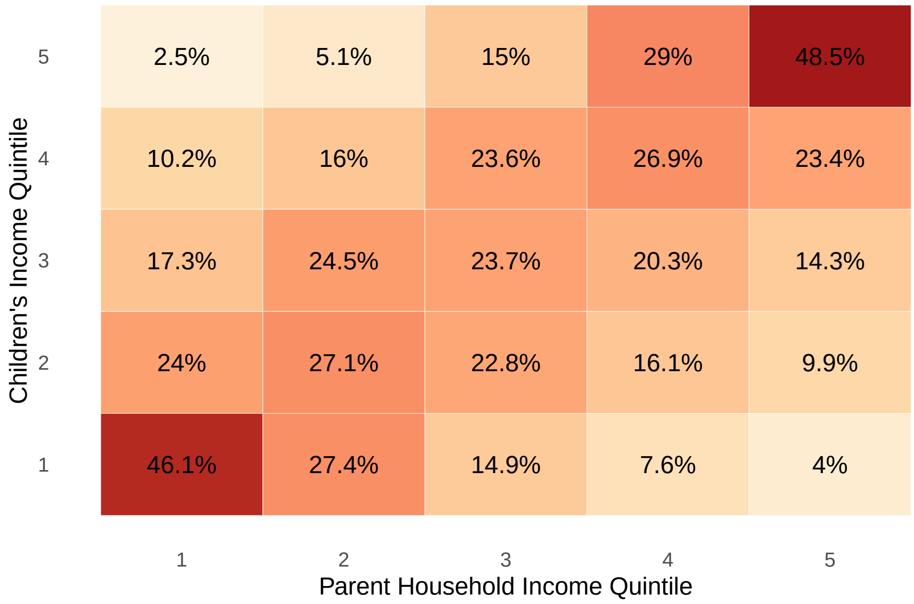
The rank-rank slope coefficient in Equation (1.1) equals 0.546, meaning that a 10 percentile increase in parental income is associated on average with a 5.46 percentile increase in children's income in adulthood. Based on this estimate, it would require seven generations for a family starting in the 25$^{\text{th}}$ percentile to reach the same rank of a family in the 75$^{\text{th}}$ percentile.[22]

Regarding absolute mobility, a child born to parents in the 25$^{\text{th}}$ percentile is expected to reach the 35$^{\text{th}}$ percentile in adulthood. Figure Figura 3 also shows that – even conditional on parental income – there is considerable variation in children's outcomes. For instance, the inter-quartile range of child ranks for parents at the 25$^{\text{th}}$ percentile is [17, 53].

Figure Figura 4 shows the transition matrix between quintiles of the parental and child income distributions. The probability of raising from the bottom to the top quintile within one generation is only 2.5%, and the probability of falling from the top to the

[22] Assuming that permanent income over generations is an $AR(1)$ process, the number of generations N required for families that are Δ percentiles apart to converge to the same percentile solves the equation $\beta^N \Delta = 1$, where β is the rank-rank slope coefficient (4). This back-of-the-envelope calculation might be a lower bound given that recent empirical estimates find a stronger correlation between the grandparents' and grandchildren's incomes than an $AR(1)$ process would suggest (57; 58).

Figura 4: Transition Probability Matrix by Quintile



Notes: The figure shows the probability that children born to parents in a given quintile of the parental income distribution (horizontal axis) move to a given income quintile in adulthood (vertical axis). Darker red tones indicate higher probabilities.

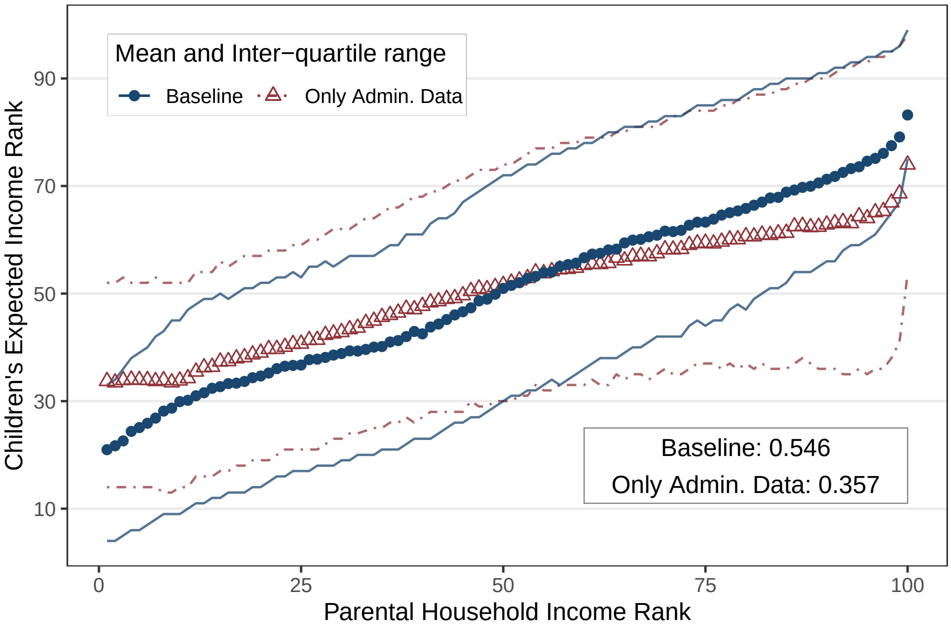
bottom is only 4%. In turn, roughly half of the children born in the bottom fail to escape poverty, remaining at the bottom, and half of the children born in the top remain at the top.

1.4.2 Measurement error

In Figure Figura 5, we show that an alternative estimate of the rank-rank curve relying exclusively on administrative data is considerably flatter than the baseline curve: the slope coefficient decreases from .546 to .357. The most likely explanation for this result is attenuation bias from measurement error, as administrative data neglect informal income and, also, formal non-labor income in the years for which tax data are not available. Consistently with this conjecture, measurement error is particularly strong for informal workers with zero formal income, showing up in a flat pattern over the bottom 10% of the parental income distribution. In addition, the inter-quartile range is substantially larger at the upper side of the parental income distribution, which is likely due to measurement error from neglecting dividends and other sources of non-labor income for years in which they are not reported in tax data. Therefore, imputing all these sources of income unreported in administrative data is crucial for correctly estimating IGM.

At the same time, any imputation will naturally generate some degree of measu-

Figura 5: Mobility Curve Comparison: Baseline vs. Administrative Income Data Only



Notes: The figure shows our baseline mobility curve displayed in Figure Figura 3 (blue dots) and the mobility curve obtained when solely relying on administrative data sources to measure income (red triangles). For each parental income percentile, we plot the mean child income rank during 2015-2019, i.e. when the cohorts of children in our main sample (1988-1990) where aged 25-31, along with the interquartile range. Parental income is the sum of the father's and mother's average income when the child is aged 3-18 years old. For each curve, the figure also displays the estimated β coefficient in Equation (1.1).

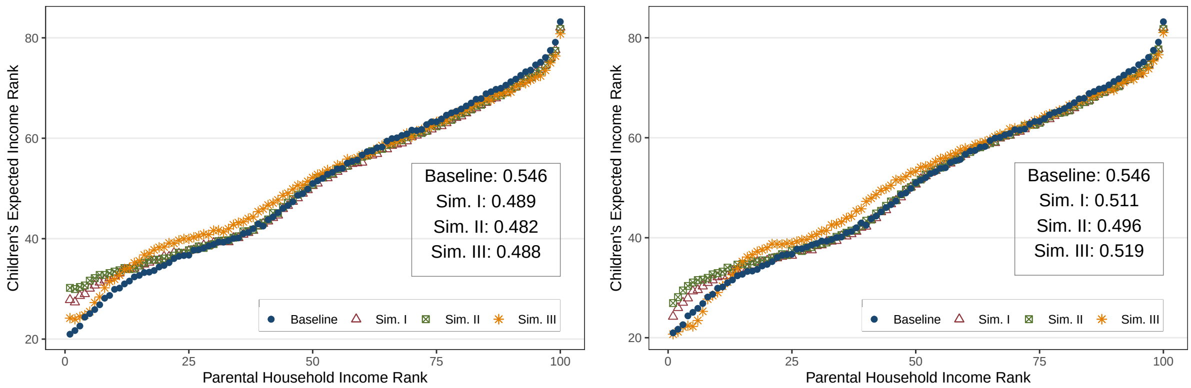
rement error. While classical measurement error would lead to attenuation bias, other forms of prediction errors could in principle affect our estimates. To gauge the direction and magnitude of any potential bias, we simulate errors in the imputation process by re-sampling the prediction errors estimated in our baseline analysis and compare the rank-rank curve on simulated data with our baseline estimate.

Specifically, we set our main sample – used to construct Figure Figura 3 – as our *benchmark sample*, i.e., we treat each income component in this main sample as the ground truth. We then generate the *simulated sample* by adding measurement error to predicted income components in the benchmark sample, namely informal income and formal non-labor income (when tax data are not available). The error is drawn with replacement from the actual distribution of prediction errors in our baseline RF models, i.e., the difference from predicted income values to those reported in the survey data (see Appendix Figure Figura 19). Importantly, we stratify this process by year, income deciles, and occupation category, so that the error distribution is allowed to vary across time and groups. The left panel of Figure Figura 6 shows the mobility curve based on our *benchmark sample*, and our *simulated sample*. Relative to the benchmark, the simulated curve is slightly flatter, with the rank-rank slope decreasing from .546 to .489 (Sim. I). The same graph shows that these results barely change when we progressively add individual characteristics to stratify the sampling error – namely, age and gender (Sim. II), and education and region of residence (Sim. III). Overall, the attenuation bias is quantitatively small and consistent with classical measurement error in our main income measures.

So far, the exercise accounts for errors in predicting income reported in the survey data. However, some degree of reporting error naturally arises in survey data, and the latter may correlate with income (e.g., see 59; 60). To account for this possibility, we repeat the second part of our simulation exercise drawing prediction errors from the comparison between predicted income and a precisely measured counterpart in administrative data. Specifically, we predict formal labor income for our entire benchmark sample using our RF algorithm on survey data, and draw the error distribution from the difference between these predictions and the precisely measured counterpart observed in the formal employment data. As such, the predictions from our RF algorithm are compared to high-quality income data from administrative registries. The underlying assumption is that the prediction error for formal income is informative on the error for imputed income that cannot be observed in the administrative data sources. The right panel of Figure Figura 6 shows the mobility curve based on our *benchmark sample*, and our *simulated sample* following this refined procedure. Again, the results indicate that the measurement error only leads to a small degree of attenuation bias, with the rank-rank slope moving from .546 to .496.

Figura 6: Measurement error, simulation

(a) Prediction Errors Relative to Survey Data (b) Prediction Errors Relative to Admin Data



Notes: The figure plots simulated mobility curves adding noise to imputed income components (red triangles, green squares, yellow stars), along with the baseline mobility curves (blue dots). On panel (a), noise is drawn from the distribution of prediction errors in our RF model, relative to income reported in survey data, while on panel (b) measurement error is drawn from the distribution of prediction errors of the RF model against administrative data. In each simulation exercise (I, II, and III), noise is drawn from error distributions stratified by an increasing set of characteristics: year, income decile, and occupation category for Sim. I (red triangles); plus age and gender for Sim. II (green squares); plus education and region of residence for Sim. III (yellow stars). For each curve, the figure also displays the estimated β coefficient in Equation (1.1).

Overall, these exercises establish that measurement error arising from the imputation procedure only attenuates our estimates, and the resulting bias is quantitatively small. The low sensitivity of our estimates can be attributed to three features of our methodology. First, the RF described in Appendix A.1.3 – based on high-quality survey data – generates small prediction errors. Second, we measure income over many years and most individuals' incomes are computed as an average between income from administrative data and imputed income, further attenuating the impact of prediction errors or misreporting in survey data. Third, we employ rank-based measures, meaning that small disturbances in income will not dramatically alter the final percentile rank of an individual. If anything, these results even reinforce our conclusion that IGM in Brazil is very low.

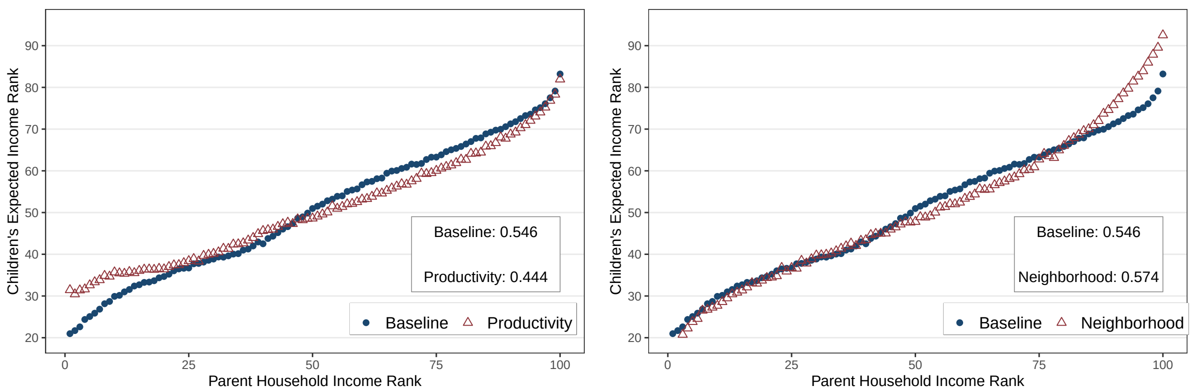
In Appendix A.2.1, we show that our IGM estimates are not significantly affected by other sources of bias in IGM measurement, namely selection, life-cycle, and attenuation bias. In addition, they are largely unaffected when exclusively focusing on years when tax data are available to measure parental (and child) income, and when varying income and occupation definitions used in the imputation process. In Appendix A.2.2, we show that focusing on household income to measure child ranks also has little impact on our results.

1.4.3 Alternative mobility measures

Figures Figura 7a-Figura 7b show the robustness of our main results to ranking parents and children on the “productivity”and “neighborhood-based”measures described in Section 1.3.3. The rank-rank curves obtained using these alternative approaches are very similar to the baseline curve. This is a remarkable result given that our baseline approach imputes both informal income and formal non-labor income, while neither of the alternative approaches relies on any imputation. The productivity-based curve, which relies only on formal labor income, is only slightly flatter, yielding a rank-rank slope of .44. This is the case because the minimum wage is binding in the bottom part of the distribution and because omitting capital income contributes to flattening the curve in the upper half of the distribution. By contrast, the neighborhood-based curve has a slope of .57, and it is steeper in the top quartile of the parental income distribution. This is consistent with the intuition that ranking individuals on income may underestimate the high living standards of children raised in affluent families, who may enjoy amenities and transfers beyond the income that they produce.

Figura 7: Alternative Measures

(a) Productivity-based (b) Neighborhood-based



Notes: The figure plots mobility curves based on the productivity-based ranking (a) and neighborhood-based ranking (b), along with our baseline mobility curve. The productivity-based measure is based on the average formal labor income for parents and children in periods when they hold formal jobs. The neighborhood-based measure is based on the average formal income in the census tract in which children grew up (parental rank) and where they live as adults (child rank). Section 1.4.3 provides a detailed description of these measures. For each curve, the figure also displays the estimated β coefficient in Equation (1.1).

To the extent that individuals born in a given place may develop preferences for that area, the neighborhood-based measure could underestimate mobility since such preferences will mechanically create persistence in our analysis. We address this issue by restricting the analysis to individuals who live in a different postal code from the one in which they were raised. Although such selection is endogenous and should be interpreted with some caution, this exercise may offer some insights into the bias due to preferences for remaining in the same areas. Appendix Figure Figura 24 shows that dropping children who did not change area flattens the neighborhood-based measure, but the rank-rank slope remains as high as 0.48 and continues to show strong persistence in the top 20% of the distribution.

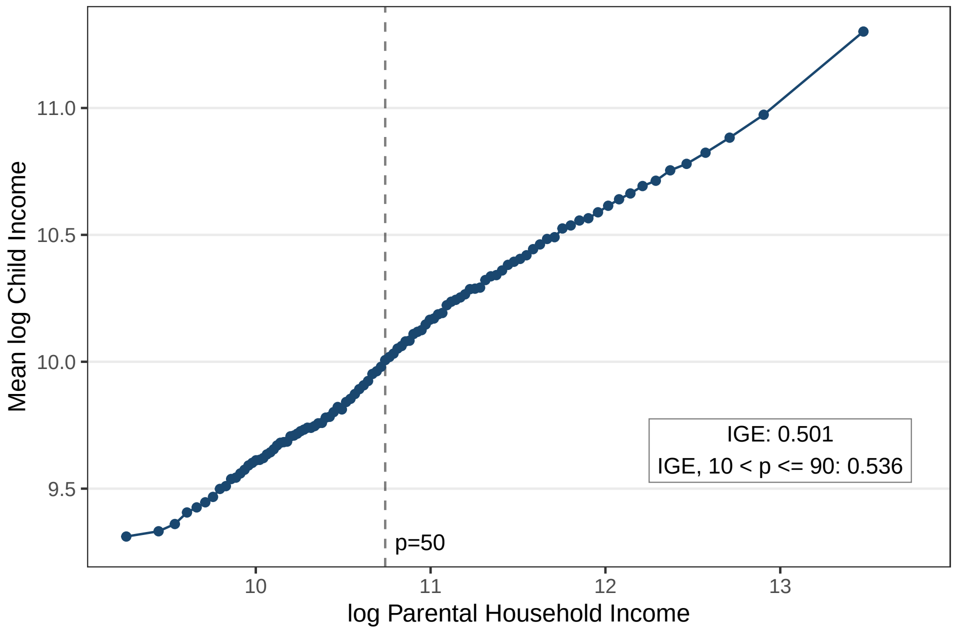
Overall, the results obtained using the three rankings – total income, productivity, and neighborhood income – are largely consistent with each other in revealing low levels of IGM in Brazil relative to previous evidence based on (developed) countries.

1.4.4 Cross-country comparisons

As already shown in Figure Figura 1, the Brazilian rank-rank slope of .546 is much larger than existing estimates for rich countries, which lie in the range of .19-.34. Absolute upward mobility is also lower, as below-median income children reach an income rank around 6 percentiles lower in Brazil than in the US. We reach similar conclusions when comparing the full mobility matrix across income quintiles. For instance, children born in the bottom income quintile in Brazil have only a 2.5% chance of reaching the top quintile, while the same figure is three times larger in the US (7.5%) and 4-6 times larger in Italy (11.2%) and Sweden (15.7%).

The stark contrast between Brazil and developed countries is also evident when we turn to the intergenerational income elasticity (i.e., the log-log relationship between parent and child income) as an alternative measure of income persistence. We estimate an IGE coefficient of .50, significantly larger than the estimates available for high-income countries, e.g., (3) finds .34 for the US and (4) .23 for Italy. (32) finds an even larger IGE of .69 in Brazil using survey data and instrumenting parental income by education. Such a higher estimate may reflect – among other things – the fact that parental education increases child income through other mechanisms beyond parental income.

Figura 8: Intergenerational Income Elasticity (IGE)



Notes: The figure plots the relationship between child and parental log income for our main sample (1988-1990 cohorts). For each level of log parental income (100 bins), it plots the mean log child income during 2015-2019, at the age of 25-31. It also reports the estimated IGE slope across all individuals and when restricting to parents between the 10^{th} and 90^{th} income percentiles. The vertical dashed line marks the median income in the parental income distribution.

Comparing rank-rank coefficients across countries requires some caution because they are based on different underlying income distributions. Specifically, it is mechanically more challenging to move up children by a given number of percentiles in more unequal income distributions. In Appendix A.2.4.1, we follow previous literature (4; 61) and focus on a different absolute mobility measure that does not depend on the underlying income distribution, namely the probability that a son earns more than his father (Figure Figura 25).[23] We find an average probability of .56 in Brazil, similar to the available estimates for the US (.55) and Italy (.53). A counterfactual exercise shows that the similar rate is partially driven by the faster GDP growth in the Brazilian economy relative to these two countries, as the probability that a son earns more than his father would decrease to .48 if income growth across generations in Brazil were equivalent to that observed for the US. The counterfactual analysis also shows that the joint distribution of parental and child income, and within-generation income inequality play a smaller role in explaining cross-country differences.[24]

[23]We focus on father-son income so that our estimates are comparable to available estimates for the US (61) and Italy (4). They focus on similar birth cohorts to ours (1988-1990 vs. 1980-1982 for the US and 1979-1983 for Italy). Father and sons average age when income is measured is 25-29 in our sample, relative to 30 and 36 in the US and Italy, respectively.

[24]Overall, these findings are in line with (62) who finds that aggregate growth is the main factor behind

Finally, no other previous paper provides similar IGM estimates for developing countries, although some have computed measures of educational mobility. (29) estimate that children born to parents in the bottom half of the education distribution in India are expected to reach the 36th percentile in the same distribution, which is similar to our estimate of absolute mobility for Brazil. We also estimate measures of upward and downward educational mobility used by (1) in Africa. Overall, we find that educational mobility displays large variation across Brazilian regions, but is generally comparable to the levels found for the most mobile African countries – see Appendix A.2.4.2 for the details on this analysis. Interestingly, even though income and educational mobility are positively correlated across Brazilian regions, we still observe large variation in absolute income mobility for given levels of educational mobility. This finding further motivates our mobility analysis based on income rather than education.

1.4.5 Mobility by gender and race

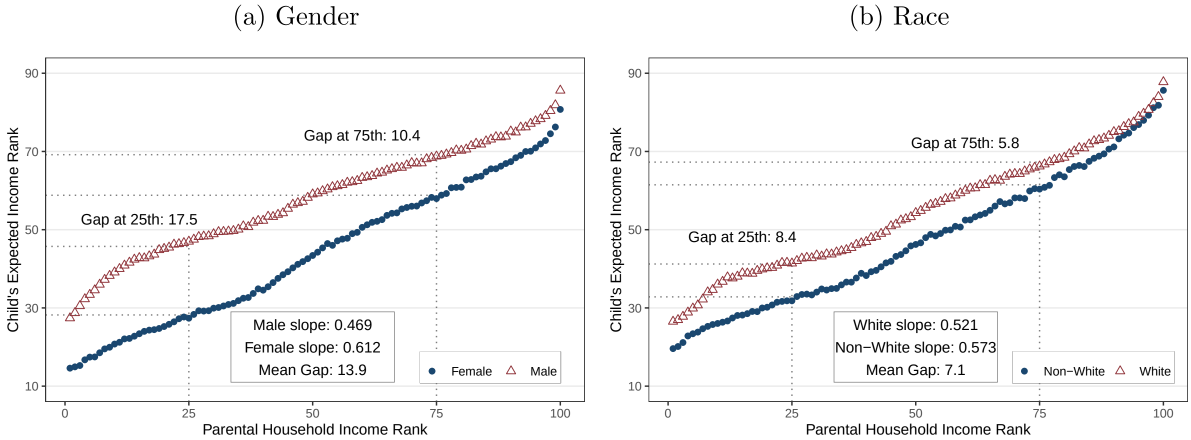
Opportunities depend not only on parental income but also race and gender, especially in a country characterized by strong segregation such as Brazil. Figures Figura 9a and Figura 9b show the gender- and race-specific mobility curves, respectively. Importantly, the ranks on both axes indicate the positions relative to all individuals within the same cohort (rather than separately by gender and race), so the graphs show between-group differences in child ranks keeping constant parental income.

Female children's income is on average 14 percentiles below males with the same parental income. This gap largely reflects gender differences in labor market participation and wages (Appendix Table Tabela 8). The mobility gap is virtually identical when restricting the same comparison to siblings, whereas the gap between siblings unconditional on gender is near zero (Appendix Table Tabela 9). Interestingly, the rank-rank slope is steeper for females than for males (.61 vs. .47). Consequently, the gender gap declines from 17 to 10 percentiles when moving from the 25th to the 75th percentile of the parental income distribution. Interestingly, girls exhibit higher absolute mobility than boys in household income upon marriage or cohabitation (Appendix Figure Figura 27), as on average girls marry partners with higher income.

Turning to differences by race, Figure Figura 9b shows that non-white children

changes in absolute mobility over time for several countries.

Figura 9: Mobility Curves by Gender and Race

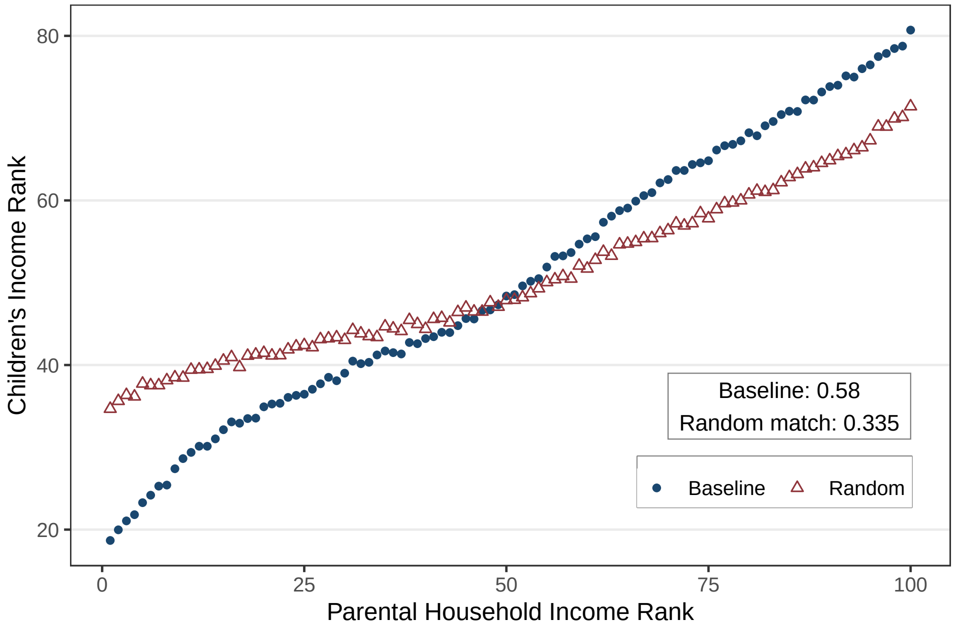


Notes: This figure plots separate mobility curves by gender (a) and race (b), for our main sample (1988-1990 cohorts). For each parental income percentile, it plots the mean child income rank in ages 25-31. Parental income is the sum of the father's and mother's average income when the child is aged 3-18 years old. The ranks on both axes indicate the income positions relative to all individuals within the same cohort (rather than separately by gender and race). For each curve, the figure also displays our relative mobility measure based on Equation (1.1), the between-group gap conditional on having parents at the 25^{th} and 75^{th} income percentiles, and the average between-group gap across parental income percentiles.

rank on average 7 percentiles lower than white children with the same parental income. Race-specific transition matrices show that non-whites born in the first income quintile are much more likely to remain at the bottom (52.8% vs. 33.7%) and less likely to climb to the top (2% vs. 3.4%) compared to white children (Appendix Figure Figura 30). Although differences are strongly reduced at the top of the distribution, non-whites born in the top quintile are twice as likely to fall to the bottom relative to white children (5.7% vs. 2.8%). The large mobility gap is particularly remarkable given that non-whites – mainly comprising black and mixed-race individuals – are far from a minority in Brazil, representing about half of the population. The gap by race in Brazil is similar to the black-white gap in the US, where the former group is a minority. Unlike the gender gap, looking at children's household income does not strongly influence the racial gap, as less than 30% of marriages in our sample are interracial (Appendix Figure Figura 31).

In addition, we document a high degree of assortative mating on education (following the definition of 63), for both parents and children; see Appendix Figure Figura 28. College-educated individuals are more than three times more likely to marry other college graduates than would be expected under a random assignment of partners. A similar pattern emerges when analyzing assortative mating over partners' income. Figure Figura 10 compares the baseline mobility curve with a hypothetical curve randomly matching children to partners/spouses (holding constant individual incomes). The counterfactual rank-rank slope in this alternative scenario would be 42% lower, and absolute mobility would increase by 16%. These results suggest that assortative mating may strongly contribute to inequality and intergenerational persistence.

Figura 10: Baseline vs. Random Mating Household Mobility Curve



Notes: The figure compares the baseline mobility curve (blue dots) with a counterfactual curve obtained by randomly assigning partners/spouses to children (red triangles). For each parental income percentile, it plots the mean household income rank measured during 2015-2019, at the age of 25-31. Parental income is the sum of the father's and mother's average income when the child is aged 3-18 years old. Both curves are constructed using our main sample restricted to married (or cohabiting) individuals. For each curve, the figure also displays the estimated β coefficient in Equation (1.1).

1.4.6 Variation in child ranks by parental income

In light of the substantial variation in child ranks for given levels of parental income – as shown by the wide inter-quartile range in Figure Figura 3 – we study which predetermined family and child characteristics may explain this variation. We run a linear regression for child rank on several of these factors, while linearly controlling for parental income; the results are reported in Appendix Figure Figura 32. The gender gap is not significantly affected when controlling for parental characteristics. Instead, adding these control variables reduces the racial gap from 7 to 2 income percentiles, indicating that racial differences can be explained by factors such as parental occupation, education, cohabitation and family size. These results also reveal that several characteristics display strong predictive power on child income. First, children to married parents – i.e., who we can identify as a couple in welfare registries or tax data – rank five percentiles higher.

Second, children to firm owners and college graduates rank around 3-5 percentiles higher, while children in large families – with more than five members – and children born to a father older than 40 rank about 3 percentiles lower. Instead, formal employment status and birth order have much smaller predictive power on child ranks.[25] Finally, children born in Center-South regions relative to those born in the Northern regions rank roughly ten ranks higher, in line with the substantial variation in mobility across space in Brazil (Section 1.5).

1.4.7 Parental income and children's long-term outcomes

Figure Figura 11 shows how parental income relates to a wide array of children's outcomes other than income. Figure Figura 11a plots college attainment over parental income ventiles, showing that it is convex over income: while children in the bottom ventile have almost no chances of completing college, roughly 80% of children in the upper ventile do so. Girls exhibit higher educational attainment than boys over the entire parental income distribution, yet they experience lower income later in life (Figure Figura 9a). Children in higher-income families are disproportionately more likely to hold prestigious occupations – such as doctors and lawyers – and this relationship is highly convex at the top (Figure Figura 11b). Figures Figura 15c, Figura 11d and Figura 11e show that low parental income is also strongly associated with markers of socioeconomic struggle. Children born to below-median income families are four times more likely to receive conditional cash transfers (*Bolsa Família*), five times more likely to become teenage mothers, and twice as likely to be the victim of a crime leading to hospitalization compared to richer children.[26] Finally, low parental income is associated with early mortality (Figure Figura 11f): children in low-income families are up to three times more likely to die before they turn 30.

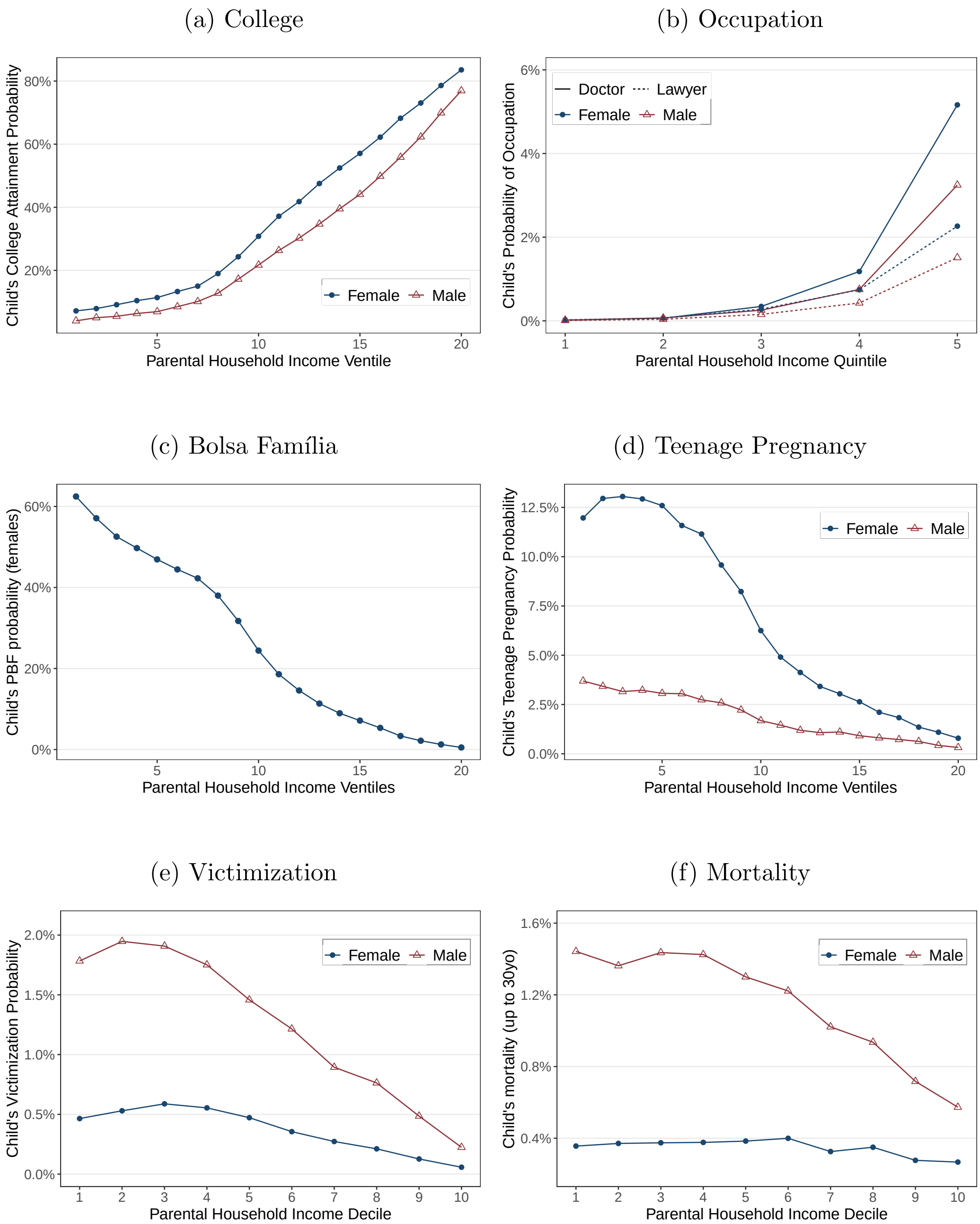
Therefore, parental income is associated with better outcomes along several dimensions beyond income. These results also suggest that income persistence may be explained (or amplified) by gaps in educational achievement and other factors that emerge early in life such as teenage fertility. The fact that all children outcomes are correlated in the expected direction with parental income rank and that most of these relationships are

[25] Parental occupation is defined by the modal occupation throughout the analysis period, when children are in ages 3-18.

[26] We measure victimization as the probability of hospitalization due to an assault.

very smooth bolsters our estimates of the rank-rank curve.

Figura 11: Long-Term Outcomes



Notes: This figure plots the relationship between parental income, measured when children are aged 3-18, and several children long-term outcomes in adulthood: college degree attainment (a), the probability of working as a doctor or lawyer (b), the likelihood of receiving *Bolsa Família* transfers when adult (c), teenage pregnancy rates (d), the probability of being hospitalized due to violent assault (e), and mortality rates (f).

1.5 Geographic variation in mobility

1.5.1 Geographical units and IGM measures

Brazil exhibits extreme variability in local socioeconomic conditions. We investigate social mobility across the 510 "immediate geographic regions" (IGRs), which are aggregations of neighboring municipalities sharing the same urban network and a common local hub (similar to the US commuting zones).[27] We assign children to the area in which they grew up, which we proxy by their father's place of residence (or, when the latter is missing, the mother's) in 2000, i.e. when children in our sample were aged 10-12.[28] Like in the main analysis, we rank parents and children relative to the national income distribution.

1.5.2 Regional mobility patterns

The rank-rank relationship between parental and child income remains linear within regions – see, e.g., the plots for Belo Horizonte and Fortaleza, two of the largest metropolitan areas in the country, in Appendix Figure Figura 33. Therefore, we can compare mobility between regions using the measures of relative and absolute mobility introduced in Section 1.3.4, which rely on such linearity.

Figure Figura 12 visualizes spatial variation in absolute mobility across IGRs. The map highlights three striking patterns. The first pattern is that absolute mobility strongly varies across regions, with the expected rank of below-median income children ranging between the 10^{th} and the 51^{st} percentile. More developed areas in the Center-South display significantly higher upward mobility relative to the less affluent North and Northeast regions. A natural concern is that this map reflects different costs of living across regions. In Appendix Figure A.3.2, we show that adjusting for prices does not alter the main patterns in the map.[29] In Appendix Table Tabela 10, we report mobility estimates for the 50 largest metropolitan areas of the country.

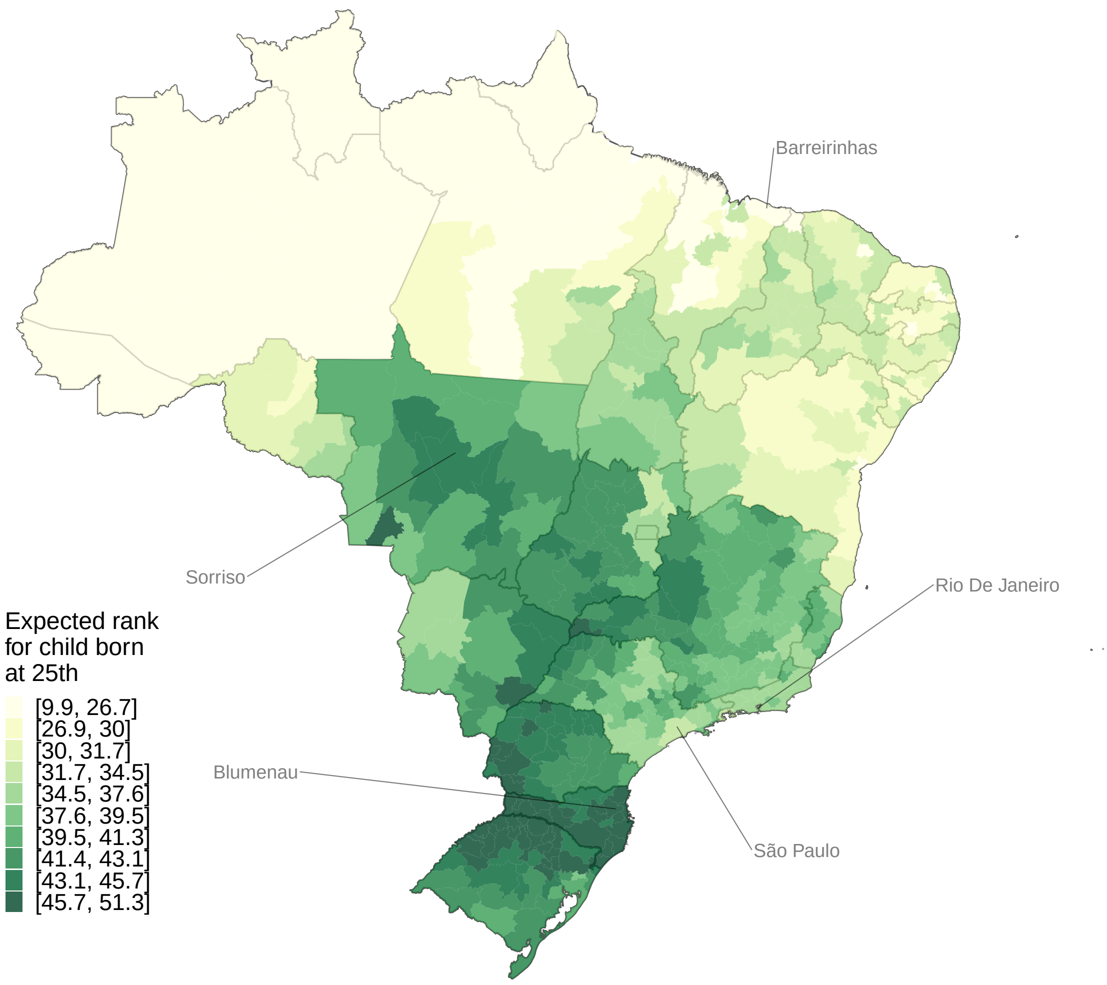
The second striking pattern is that several regions in the countryside display higher

[27] IGRs replaced the *microrregião* used in earlier studies on Brazil.

[28] In our main sample, the father and mother live in the same IGR in 83% of cases.

[29] The correlation between baseline and price-adjusted estimates of both absolute and relative mobility across regions is above .9. This high correlation is explained by the fact that, although prices significantly vary across regions, most children live in the same area where they grew up (or in areas with similar price levels).

Figura 12: Absolute Mobility Map: Predicted Rank for a Below-Median Income Child



Notes: The figure visualizes spatial variation absolute mobility (in deciles) across Brazil's 510 immediate geographical regions (IGRs) for our main sample (1988-1990). Parent and child incomes are ranked in the national income distribution and measured when children are aged 3-18 and 25-31, respectively. Absolute mobility indicates the expected rank for children in below-median income families, based on Equation (1.1). Darker green tones indicate higher absolute mobility. Children are assigned to IGRs according to the location of their fathers in 2000.

absolute mobility than large and rich metropolitan areas such as São Paulo and Rio de Janeiro. On the other hand, large metropolitan areas such as São Paulo and Rio de Janeiro offer excellent outcomes for high-income children but not for low-income children (see Appendix Table Tabela 10).

The third pattern is that the top 5% areas in terms of absolute mobility are all concentrated in a large mobility hotspot crossing three southern states: Paraná, Santa Catarina, and Rio Grande do Sul. This region has historically been characterized by the presence of agricultural communities established by European settlers maintaining a strong cultural heritage such as Caxias do Sul and Joinville, founded respectively by Italians and Germans. In such regions, below-median income children reach on average the 47^{th} percentile in adulthood and about 80% of children born in the bottom quintile escape poverty, transiting to higher income quintiles (see Appendix Table Tabela 10).

Appendix Figure Figura 35 shows, in addition, that areas with higher absolute mobility do not necessarily display higher relative mobility. While places with high absolute mobility in the Southern regions also tend to display higher relative mobility (lower rank-rank slopes), the correlation between the two is close to zero in the Northern regions. This indicates that high absolute mobility may be achieved by both producing parallel upward shifts in the mobility curve and breaking the correlation between parental and child income.

Finally, Appendix Figure Figura 36 documents a Great Gatsby curve within Brazil, and Appendix Section A.3.5 presents an analysis of the factors that better explain the substantial regional variation in mobility. Although entirely correlational, this analysis may inform future work aimed at understanding the causal determinants of upward mobility. Interestingly, we find that factors related to the quality of education provision yield by far the highest explanatory power on absolute mobility across IGRs, followed by indicators related to family structure, demographics (including the racial composition), household characteristics, and the local infrastructure. Although there is some overlap with the main mobility predictors found by (3) and (4) for the US and Italy, in Brazil the quality of education stands out as the strongest factor.

1.6 Causal place effects

Motivated by the stark regional disparities in IGM documented in the previous section, we next estimate the causal effect of the place where children grew up on their perspectives of upward mobility. To disentangle such effect from sorting, we compare migrant children (or siblings) who moved to new areas at different ages (24).

1.6.1 Data and research design

For this analysis, we use a sample that covers all children born during the 1983-1992 period that can be linked to their fathers. We distinguish between permanent residents and movers based on parents' residency in the 1992-2019 period. Like in Section 1.5, the geographical unit of analysis is the IGR. We track moves using formal employment data, because address coverage prior to 2000 is low in the person registry. Appendix A.4.1 provides details on the data construction.

Our empirical strategy and specifications closely follow (24) (see also 38, for an application to Australian data). We first characterize the predicted outcomes of permanent residents using rank-rank regressions for each cohort and region (see Appendix A.4.2 for additional details). We then use these estimates to compute the predicted rank difference for each mover based on the origin and destination region, the child's cohort, and parental income rank. Finally, we estimate causal place effects by relating movers' income rank at the age of 24 to their predicted difference in ranks across children moving at different ages.[30] Intuitively, to the extent that location exert causal effects, movers' outcomes should display greater convergence to that of permanent residents the earlier they move (and the longer they are exposed) to the destination place. Specifically, our main analysis is based on the following equation:

$$y_i = \alpha_{ocpa} + \sum_{a=1}^{33} b_a I_a(a_i = a)\Delta_{odpc} + \sum_{c=1983}^{1991} \kappa_c I_c(c_i = c)\Delta_{odpc} + \epsilon_i, \tag{1.2}$$

where y_i is the child's income rank at the age of 24; α_{ocpa} is a fixed effect by origin o, cohort c, parental income decile p, and age at move a; I_a and I_c are indicators for each

[30] Like (24), we focus on income at an earlier age relative to our main analysis (Section 1.4), so that we can measure income for (older) cohorts who move at older ages.

age at move a and cohort c; and Δ_{odpc} is the difference in permanent residents' predicted outcomes between origin o and destination d for parental income decile p and cohort c. The coefficients of interest b_a for $a \leq 24$ give the expected increase in rank associated with moving at age a to a destination with a 1 percentile higher predicted rank. Since we measure income at 24, moving at an older age cannot possibly have a causal effect on income, so b_a for $a > 24$ captures solely selection effects. The coefficients κ_c control for our varying ability to track moves across cohorts, ensuring that we only use within-cohort variation in the age at move.[31] One advantage relative to (24) is that we track moves from age 1 (instead of 11). This implies that we can flexibly study how convergence varies by age from early life, without the need to rely on linear extrapolations.

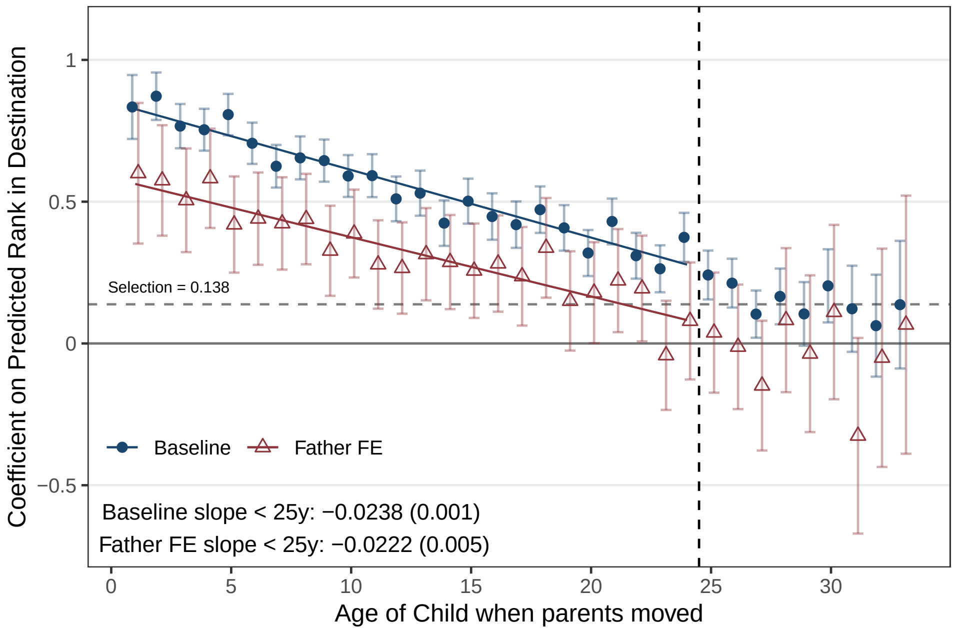
Our econometric specification effectively compares the extent of convergence to destination outcomes by children who move at different ages. The key identifying assumption for the analysis is that selection effects driving some children to move to better (or worse) areas are orthogonal to the child's age when they move. Importantly, fixed effects α_{ocpa} ensure that the estimation of convergence coefficients for moves at each age exclusively relies on variation between children who have the same parental income background, and belong to the same cohort and place of origin. We provide several pieces of evidence that strongly support our main identification assumption in the next Section 1.6.2.

1.6.2 Results

Causal place effects are summarized in Figure Figura 13. The series in blue displays the estimated coefficients b_a on the predicted difference in outcomes for each child given her origin, destination, cohort, and age at move. Supporting the intuition behind our identification strategy, the extent of convergence decreases with age, following a roughly linear pattern: children who move earlier to better places benefit more. The positive coefficients from the age of 25 purely reflect positive selection into migration, as child income measured at the age of 24 – by construction – cannot be affected by future moves. The flat pattern from the age of 24 supports our identifying assumption that parental selection into migration to a given destination does not vary with age.

[31] For instance, since we track parents' locations starting from 1993, we observe children born in 1983 moving from 10 years old onwards, while children born in 1992 since the age of 1. To avoid collinearity, we omit the indicator for the 1992 cohort.

Figura 13: Exposure Effect Estimates for Children's Income Rank in Adulthood



Notes: This figure plots the estimated b_a coefficients in equation (1.2) (blue dots) and in an alternative specification including family fixed effects in Appendix equation (A.6) (red triangles). The sample includes all father-linked children from the 1983-1992 cohorts whose father moved once between 1993-2019, and the dependent variable – child income – is measured at the age of 24 (dashed vertical line). Vertical bars report 95% confidence intervals. Each coefficient b_a indicates the degree of convergence to the outcomes of permanent residents in the destination relative to those at the origin. Coefficients b_a for moves until the age of 24 estimate causal place effects, while b_a coefficients for older ages estimate selection effects as moves when aged 25 or older cannot explain income at age 24. The slope of the blue and red lines, as estimated by linear approximations for b_a in equations (1.2) and (A.6), summarize convergence per year of exposure.

Since exposure effects decline linearly with age at move, we substitute the coefficients b_a with a linear counterpart to estimate an average convergence by year of exposure.[32] Each additional year of exposure to the destination area increases convergence in children's outcomes by .024 (baseline slope), meaning that children moving at birth to a place where they are expected to rank 10 percentiles higher will increase their rank in the national income distribution by $.024 \times 24 \times 10 = 5.76$ percentiles on average due to causal place effects.

These estimates suggest that about 57% of the substantial mobility gap across Brazilian regions (Section 1.5) is due to causal place effects. Therefore, some areas in Brazil offer significantly better opportunities for low-income children, notwithstanding the high levels of income persistence at the national level.

[32]Specifically, we substitute the non-parametric term $\sum_{a=1}^{33} b_a I_a(a_i = a)\Delta_{odpc}$ in eq. 1.2 with a linear counterpart $I(a_i \leq 24)(b_0 + (24 - a_i)\gamma)\Delta_{odpc} + I(a_i > 24)(\delta + (24 - a_i)\delta')\Delta_{odpc}$. The linear term is split at the age of 24, since moves above this age capture selection effects. γ is the main coefficient of interest identifying convergence by year of exposure. We repeat the same procedure to summarize convergence for an alternative specification used in the robustness analyses.

1.6.3 Robustness

Importantly, Figure Figura 13 shows that our results are virtually unaffected by the addition of family fixed effects, thus relying exclusively on within-family variation (Appendix A.4.3 provides the details on this specification). The latter rules out the possibility that our findings are driven by family selection over child age at move.

Our results are also robust to overidentification tests. They follow from the intuition that children's outcomes should converge to the average outcomes of their own group, whereas predicted outcomes of other groups are used as placebos. The results in Appendix A.4.4 confirm that children's outcomes converge to those of permanent residents with precisely the same age, gender, and race, while coefficients on other groups are generally an order of magnitude smaller, close to zero, and statistically insignificant. In addition, since children's outcomes in different areas not only differ at the mean but over the entire distribution, we show that movers' outcomes track different moments of the distribution of permanent residents' outcomes beyond the mean.[33] These tests address additional concerns such as the possibility that moves to better places are driven by different shocks producing positive effects on children that decrease with age, e.g., positive income or wealth shocks.[34]

1.7 Conclusion

In this paper, we provide the first estimates of income intergenerational mobility for a developing country using large-scale individual-level income data linking two generations. We find that income mobility in Brazil is much lower than the estimates available for developed countries: a 10 percentile increase in parental income is associated with a 5.5 percentile increase in a child's income, and 46% of children born from parents in the first income quintile remain in the same quintile as adults. Parental income is also strongly correlated with educational attainment, teenage pregnancy, occupation, mortality, and victimization. Moreover, we uncover wide disparities across areas, genders, and

[33] For instance, two areas may have the same mean child rank for low-income children but different probabilities that children end up in the top decile of the income distribution.

[34] Specifically, they indicate that following these shocks, parents would need highly accurate knowledge to select better places for our results to be driven by selection. Accordingly, for them to drive our main findings, parents would need to select places that offer better opportunities for children from the same cohort, gender and race. Finally, the potential shocks driving such selection process would need to replicate not only the mean outcomes but also the distribution of outcomes for children in the destination.

racial groups, depicting a *"land of inequality"* in which children's opportunities are deeply dependent on their parents' socioeconomic status, connections, and location. Causal place effects account for half of large regional disparities in mobility, which, in turn, strongly correlate with educational quality factors.

These results rely on administrative registries comparable to those available for developed countries, yet the presence of a large informal sector raises additional challenges towards accurately estimating income and social mobility. We address such challenges by developing new methods for imputing informal income and alternative measures for ranking individuals on socioeconomic status. These same methods could be applied to estimate income and social mobility in other contexts characterized by a large unofficial sector, as is typically the case in low- and middle-income countries.

This work is also relevant in public and policy debates. Even though Brazil has long been perceived as a place of high inequality and low mobility, hard evidence on IGM may contribute to shifting people's perceptions and potentially their preferences for distributive policies (2). Moreover, revealing dramatic penalties for long-neglected groups and places – in particular, non-whites and the North-Northeast of the country – can encourage public policies targeted at increasing access to opportunities. In particular, our results on causal place effects and drivers of mobility across regions can motivate placed-based policies aimed at improving the quality of public education provision in the poorest areas of Brazil.

2 CONDITIONAL CASH TRANSFERS IN THE LONG RUN

2.1 Introduction

In the 1990s, Brazil's Gini Index was 0.605, and 25% of the population was poor.[35] Children born around this time in the bottom quintile of the income distribution had only a 2.5% chance of rising to the top 20% as adults (64). Yet, many things changed since. During the 2000s the country witnessed a sustained period of economic growth and the social safety net was largely increased. Despite still being one of the most unequal countries in the world, currently only 5% of the population lives in poverty.

One of many factors regarded as drivers of such reduction in poverty is *"Programa Bolsa Família"* (PBF), the largest conditional cash transfer (CCT) in the world. Created in 2004, PBF provides poor families cash transfers conditional on children's school attendance and health checks. Being a CCT, PBF was conceived with two purposes. In the short term, it hoped to alleviate current poverty through monetary transfers to poor families. In the long term, conditionalities related to children's human capital accumulation aimed to improve the living conditions of future generations. While it is well established that PBF, and CCTs in general, have positive short-term effects (65; 66; 67), there is limited evidence of whether these programs are effective in breaking the intergenerational cycle of poverty (68). In the particular case of PBF, there is no causal link between the program implementation and the improvement of adult living conditions for beneficiary children, nor evidence of its (potential) impact on intergenerational mobility.

In this paper, we aim to study the long-term effects of PBF. We use population registries to track childhood exposure to PBF and long-term educational, labor market, and socioeconomic outcomes for nearly 30 million children born between 1979 and 1998. Since the choice and timing of entry in PBF is endogenous, we train a machine learning algorithm to identify which families are most likely to benefit from PBF based on predetermined characteristics (69). We then use a difference-in-differences design with mother fixed effects to compare children young enough to benefit from PBF's conditionalities to their older siblings in families with a high versus low probability of being beneficiaries of the program (70; 71). The intuition is that PBF's long-term effects should operate

[35] The source of both statistics is World Bank's Poverty and Inequality Platform. The poverty threshold considered is the international poverty line of 2017 USD 2.15 PPP.

through higher human capital accumulation, so any impacts on children out of secondary schooling age at the time of implementation (older than 17 in 2004) should be constant over age. We show support for such hypothesis with event-study specifications. While children with longer childhood exposure display increasing benefits for each additional year of exposure, we find constant patterns for those exposed at varying ages above 17.

Using family-level predicted enrollment as an instrument for children's cumulative exposure (72), we find that each additional childhood year in PBF is associated with 1.8 p.p. and 1.1 p.p. increases in high school and college degree attainment, respectively. As the median child in the treatment group spends 5 years in the program and baseline high school completion and college attainment are, respectively, 34% and 1.6%, these findings suggest that PBF resulted in 26.5% higher chances of completing high school and a 3.5 times larger likelihood of obtaining a college degree. Moreover, we find per-year decreases in teenage pregnancy (0.3 p.p.), overall hospitalizations (0.1 p.p.), adult enrollment in PBF (2.2 p.p.), and incarceration (0.2 p.p.); as well as increases in the probability of owning a firm (0.8 p.p.) and being formally employed at 25 years old (1.7 p.p.). These findings dovetail well with earlier evidence on PBF impacts in the short run (65) and more recent long-term effects of CCTs in Mexico (73) and Colombia (74).

After documenting the positive long-term effects of PBF on beneficiary children, we look for impacts on children growing up in non-exposed families. Several studies document spillover and general equilibrium effects of social programs, in particular cash transfers (75; 76; 77). For instance, nearly-eligible children whose classmates participate in the program can be impacted through peer effects. Likewise, if PBF causes large enough changes in the supply of skilled workers, that could shift returns to education and modify education decisions in high-income families. Finally, local labor market multipliers could benefit everyone living in a particular area highly-exposed to PBF.

In a similar research design, we compare siblings' differences among non-eligible children growing up in municipalities with varying exposure to PBF. To do so, we exploit the fact that the total number of PBF beneficiaries is capped at the national level and distributed across municipalities using an allocation rule based on local poverty rates (76). We recover the first distribution of PBF slots, which was decided in 2003 before the program began, and use it as a measure of the program's intensity. This pre-program number of slots (per capita) varies substantially across the country and is a great predictor of PBF's coverage over time. The intuition behind comparing siblings' differences across values of PBF's intensity is that, if present, spillover effects should be proportional to non-eligible children's age and indirect exposure to the program. Hence, effects should emerge for younger children living in places where PBF had a relatively larger presence.

We find that a 10 p.p. increase in municipality-level PBF coverage is associated with gains in high school completion (0.6 p.p.) and college attainment (0.3 p.p.). We document improvements of similar magnitude for all health and labor market outcomes explored earlier, pointing to sizeable spillover effects. To explore potential mechanisms, we estimate spillovers by mothers' education. Effects on high school completion only appear for children whose mothers did not complete high school. This group had small but positive exposure to PBF, so effects are likely driven by both direct effects and spillovers through neighbors and classmates (75; 78). The fact that we find null effects for high-educated mothers is a good placebo test, as they should not change decisions about their children's secondary schooling due to PBF. Yet, we find increases in college, employment, and entrepreneurship across all groups. This can indicate aggregate effects on the local labor market (76; 79) and/or changes in returns to schooling with broader implications to human capital accumulation (80; 81).

To assess how direct and spillover effects impact intergenerational mobility, we compute measures of educational mobility at the city-cohort level and compare cross-cohort trends in low- versus high-coverage municipalities. Again, the intuition is that if PBF resulted in improvements in mobility, such an effect should be proportional to the program's intensity. We find that a 10 p.p. larger PBF coverage is associated with up to a 2 p.p. (3.9%) increase in the probability of children of low-education mothers completing high school, and a 1 p.p. (2.1%) decrease in the intergenerational correlation on years of education. This suggests PBF fostered opportunities for poor children and reduced the intergenerational gap between low- and high-education families.

Since PBF targeted disadvantaged areas disproportionately, one potential threat to our research design is different cohort trends reflecting convergence between poorer and richer areas of the country. Likewise, one could be worried that other public policies contemporaneous to PBF confound our estimates. To address this, we include state-by-cohort fixed effects and a rich set of pre-determined municipal characteristics interacted with linear cohort trends in all baseline specifications. In addition to reassuring event studies, in robustness exercises we add more municipal controls relative to the baseline and show that our main independent variables do not correlate with the introduction of several public policies, although being great predictors of exposure to PBF. Finally, we include a whole set of municipality-by-cohort in our analysis of direct effects, which do not impact estimates.

Our results show evidence of individual-level direct and indirect impacts of one of the largest poverty-reduction programs in the world, contributing to many strands of the literature. First, it adds to the much-needed work on the long-term effects of CCTs (68; 73; 74) and cash transfers in general (82). Similarly, it speaks to the literature on the long-run impacts of the social safety net, focused on developed countries (70; 71; 72; 83). By documenting effects on non-eligible children, our work also relates to the literature on social programs' spillovers (75; 78; 79) and aggregate and general equilibrium effects (77; 80; 81; 84). Finally, it adds the first developing country evidence to the emerging literature on the long-term determinants of intergenerational mobility (21; 22; 23).

2.2 Background and Data

2.2.1 The *Programa Bolsa Família* (PBF)

Created in 2004, PBF is the largest conditional cash transfer program in the world. As of 2019, it reached 15 million families – roughly 30% of the population – with an average stipend of 180 BRL per family, costing 0.5% of the GDP annually. The program encapsulated and expanded previous social programs that were scattered through the country, and its roll-out was relatively quick and evenly distributed across regions.

Under PBF, every family with per capita income below a national poverty threshold (140 BRL) is eligible to a variable benefit that depends on the number and age of the children. As of 2019, families could accumulate up to five child (0-15 years old) and two youth (16-17 years old) monthly benefits of 32 and 38 BRL each, respectively.[36] Transfers are conditional on children's school attendance and vaccination. In addition to the variable benefit, families with per capita income less than half of the poverty threshold receive an unconditional monthly benefit of 70 BRL.

[36]At implementation, families could only receive up to two child benefits. The youth benefit and the higher child-limit were implemented in 2008. Monetary values of transfers and the poverty threshold are regularly adjusted for inflation.

The total number of slots for PBF beneficiaries is capped at the national level and then allocated across municipalities according to “municipal quotas” based on local poverty rates. These quotas are binding nationally but can be reallocated across municipalities upon leave/entry of beneficiaries (76). Municipalities track PBF’s eligibility and other socioeconomic characteristics through *Cadastro Único* (CadÚnico), a continuously-updated registry of the poor population. Poor families can self-register in CadÚnico or be included by the municipal social assistance service. Upon eligibility and availability of PBF slots in the municipality, transfers are preferentially entitled to the woman (mother) of the household. Conditionalities are enforced in a joint effort between municipalities and the federal government.

2.2.2 Expected Long-term Effects of PBF

Several studies document PBF’s positive short-term effects on children’s education and health outcomes, as well as overall decreases in inequality and poverty rates (65). The main goal of the program was that those translated into better socioeconomic conditions in adulthood. This could be observed by long-term improvements in health and labor market outcomes as well as changes in fertility decisions and criminal behavior. These effects are in line with the program’s purpose and the literature on the long-run effects of the social safety net (70; 73; 74; 83). Given the program design, the main mechanism for such improvement would be higher human capital accumulation.

Hence, such effects should appear for children whose educational decisions were most likely influenced by PBF’s transfers and conditionalities – i.e., those in eligible families who were in schooling age when the program began. Moreover, gains should be increasing on cumulative exposure. For instance, children treated in primary-education age should have a higher likelihood of finishing high school than the treated only during in high-school. Likewise, there should be null effects for children older than 18, since they did not have to comply with conditionalities and had already concluded secondary education at the start of the program. Nevertheless, non-treated cohorts in eligible families could have non-zero effects for outcomes determined after 18 years old. For instance, this group could change later human capital accumulation and labor supply decisions as their families start to receive cash transfers.

In addition, spillovers to children of non-eligible families could be present. For example, there could be neighborhood and classroom peer effects in education (75; 78) and multiplier effects that cause increases in consumption and employment in the local labor market (76; 77; 79). Likewise, if PBF increased poor children's education there could also be local general equilibrium effects on the supply of high-skill workers with broader implications on schooling and employment decisions (80; 81).

In this paper, we try to quantify PBF's potential direct and spillover long-term effects. First, we couple event-study and IV estimates to measure the effects of childhood exposure to PBF on education, health, criminal, and labor market outcomes. We then focus on non-eligible children to identify spillovers via municipal-level variation in program's intensity.

2.2.3 Data

We combine several administrative records to perform our analysis. Our sample of children is the Brazilian person registry (*Cadastro de Pessoa Física*, CPF), which covers the entire population and is provided by the Brazilian tax authority. We restrict the sample to children born between 1979 and 1998 (aged 5 to 24 years old when PBF started) and whose mothers are uniquely-identified by their names.[37] In this way, we can link them to their mothers' and retrieve her characteristics before the launch of PBF. Our final sample comprises over 27 million children, around half of the cohorts studied.

We track individuals' participation in PBF using administrative records on the program's payments from 2004 to 2020. We observe children's educational and labor market outcomes from administrative employment data covering the population of formal jobs for the 2002-2019 period (*Relação Anual de Informações Sociais*, RAIS) coupled with CadÚnico snapshots from 2011 to 2020 and the Brazilian firm registry (CNPJ). We observe health outcomes in the universe of hospitalizations at the national health service (SIH-SUS) from 2000 to 2020. Finally, we measure criminal behavior using the universe of criminal prosecution cases collected by a private firm, and incarceration data from the Ministry of Justice (DEPEN).[38] We define the variables used as outcomes in thee following sections in Appendix Table Tabela 14.

[37]Due to accumulation of surnames, around 52% of Brazilians have a unique name. In (46), it is shown that the uniquely-named population is representative of the overall population.

[38]For details on the criminal prosecution data, see (46).

2.3 Direct Effects

2.3.1 Empirical Strategy

To estimate PBF's long-term effects at the individual level, we use a differences-in-differences setup in which temporal variation comes from cohorts born earlier versus later relative to PBF implementation (70; 71; 72). We measure family-level exposure to PBF and compare children young enough to benefit from the program with their older siblings (first difference) in families with low versus high probability of benefiting from the program (second difference).

We use mothers' exposure to the program as an instrument for cumulative participation in PBF, measured as the number of childhood years (0-17 years old) under the program. We identify high-exposed mothers by predicting their likelihood to participate in the program in 2004 using a Machine Learning algorithm, akin to (69). The baseline model uses mothers' education, age, number and age of their children, formal employment background, and city of residence as features.[39] Our treatment group comprises children whose mothers are at the top 10% of the predicted-probabilities distribution. Appendix Table Tabela 15 shows summary statistics for both groups. High-exposed mothers are 41 p.p. more likely to be PBF beneficiaries in 2004 (66 p.p. in 2004-2013) and are less educated, have more children, and are overly-represented in the North/Northeastern regions. Children from high-exposed mothers spend on average 5.5 years under the program, while control children spend around 1 year.

As discussed in Section 2.2.2, the effects should be larger for younger children. In particular, age-specific effects should be proportional to cumulative childhood exposure to PBF. Moreover, outcomes determined before 18 years old – such as high school completion and teenage pregnancy – should not be influenced by PBF for children "treated" after that age. Hence, we estimate reduced-form event-study specifications to validate the research design and flexibly assess PBF's dynamic effects relative to age at the start of the program:

$$y_{iamt} = \sum_{t=6}^{17} \left(\boldsymbol{I}_t \times \mathit{Treat}_a\right) \phi_t \; + \; \sum_{t=19}^{24} \left(\boldsymbol{I}_t \times \mathit{Treat}_a\right) \mu_t \; + \; \vec{X}'_{imt}\Omega \; + \; \alpha_a \; + \; \lambda_t \; + \; \gamma_{s(m),t} \; + \; \varepsilon_{imt} \tag{2.1}$$

[39] All covariates are measured before the start of PBF. We use data only for the first year of the program to avoid endogeneity concerns about selection into the program and its roll-out. We provide details on the estimation and performance of the prediction exercise in Appendix B.2.1.

where y_{iamt} is an outcome of individual i with mother a born in municipality m who was t years old when PBF began.[40] The key independent variable is a interaction between the mother-level binary treatment defined above, $Treat_a$, and a set of age t indicators I_t ranging from 6 to 24 years old. The vector $\vec{X}'_{imt}$ includes individual-level characteristics such as gender and race as well as baseline municipal-level pre-determined characteristics interacted with linear cohort trends.[41] Mother (α_a) and cohort (λ_t) fixed effects characterize the differences-in-differences setup, while state-by-cohort fixed effects ($\gamma_{s(m),t}$) accounts for convergence in outcomes across Brazil's states during the period. The error term ε_{imt} is clustered at the municipality level to account for serial correlation across cohorts over time. The underlying identifying assumption is that if PBF was never implemented, within-state cross-cohort trends between exposed and non-exposed families would remain constant.

To present the effects in terms of childhood years under the program, we run IV models in which childhood cumulative exposure is instrumented by the mother's treatment status:

$$1^{\text{st}} \text{ Stage:} \quad Years^{PBF}_{iamt} = \sum_{t=6}^{17} (\boldsymbol{I}_t \times Treat_a)\, \eta_t \; + \; \sum_{t=19}^{24} (\boldsymbol{I}_t \times Treat_a)\, \psi_t + \; \varepsilon_{imt} \tag{2.2}$$

$$2^{\text{nd}} \text{ Stage:} \quad y_{iamt} = \widehat{Years}^{PBF}_{iamt}\, \beta + \; \varepsilon_{imt} \tag{2.3}$$

where $Years^{PBF}_{iamt}$ is the number of childhood years child i spent as PBF beneficiary. All fixed effects and controls (omitted) are the same as in Equation 2.1. The parameter of interest β gives the effect on y_{iamt} of one additional year of program participation.

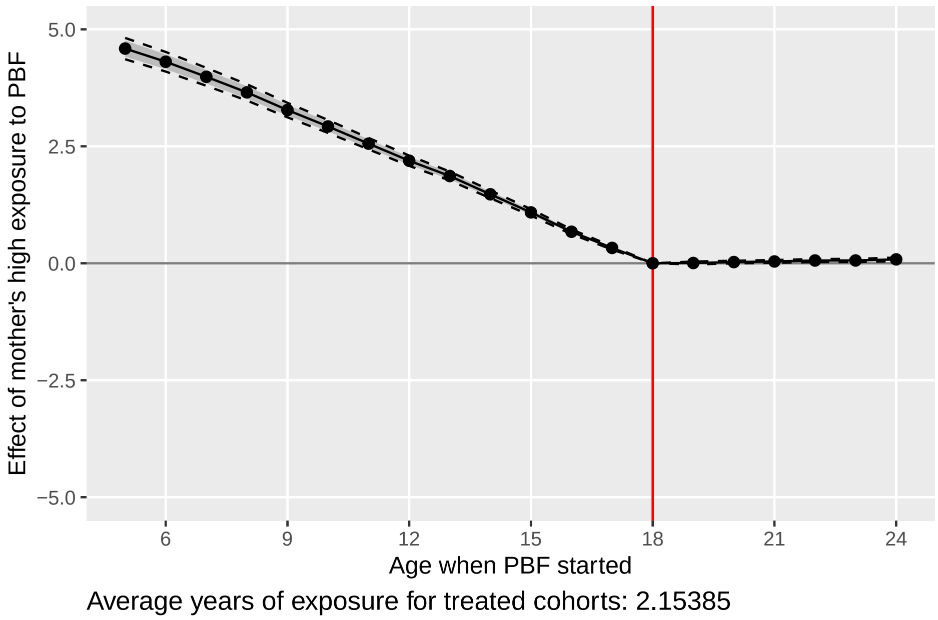
2.3.2 Results

Figure Figura 14 shows the first-stage in Equation 2.2. Mothers' predicted probability of PBF enrollment in 2004 is strongly associated with their children's exposure throughout their entire childhood. Relative to their non-exposed older siblings, children in treated families who were 5 years old when the program started spend on average 5

[40] Several municipalities were created during our period of analysis. For this reason we do not use strictly municipalities, but *"minimum comparable areas"* (MCAs) provided by the Census Bureau (IBGE). Throughout the paper, we use the terms "municipality" and "city" interchangeably to refer to MCAs.

[41] In our baseline specification, municipal characteristics are population and average income per capita measured in the 2000 Census. In Appendix B.2.2, we run robustness exercises with additional controls and include municipality-by-cohort fixed effects.

Figura 14: First-stage: Mother's Predicted PBF enrollment and Children's Cumulative Exposure



Notes: This figure plots estimates of the first-stage relationship in Equation 2.2.

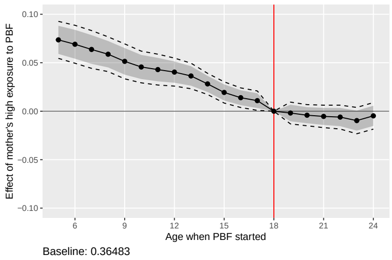
years under the program. The coefficients for older cohorts are small by construction and the decreasing pattern reflects the fact that younger cohorts had higher cumulative exposure. If the research design holds, reduced-form event-studies should have a similar pattern to Figure Figura 14.

Figure Figura 15a plots ϕ_t and μ_t coefficients obtained from Equation 2.1 for the probability of completing high school. Compared to their older siblings, children who were 6 years old at the start of PBF are 6.9 p.p. (18.9%) more likely to complete high school. The effect is decreasing on age: the effect is 1.9 p.p. (5.3%) for children who were treated when teenagers. Such pattern suggests that the research design is valid. Likewise, null effects for cohorts with no childhood exposure minimize concerns about parallel-trends assumption violation.

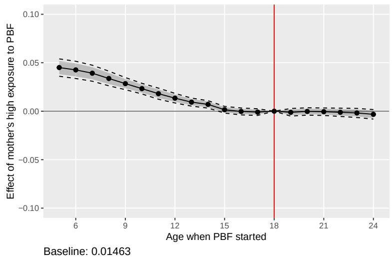
Figures Figura 15b to Figura 15f plot analogous event studies for other outcomes. Children from treated cohorts in high-exposure families have a higher probability of earning a college degree (Figura 15b), opening a firm (Figura 15f), and being employed in the formal sector when 25 years old (Figura 15e), while lower probabilities of teenage pregnancy (Figura 15d) and enrollment at PBF when adults (Figura 15c). As a benchmark, for children who were 10 years old when the program started mother's exposure is associated with increases of 2.3 p.p. (159%) in college attainment, 1.7 p.p. (79.3%) in

Figura 15: Reduced-Form Estimates of PBF Exposure Long-term Effects

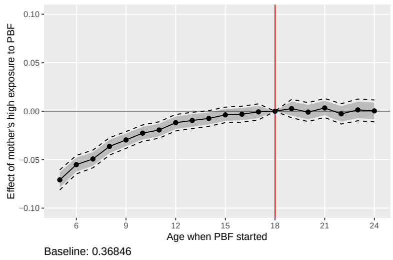
(a) High School Completion



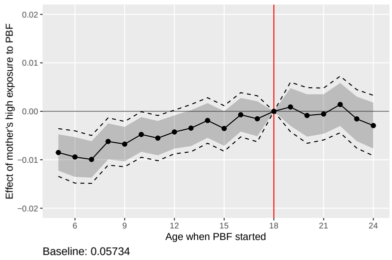
(b) College Degree Attainment



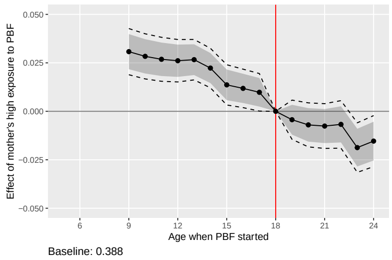
(c) PBF enrollment



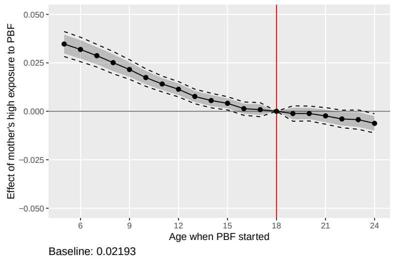
(d) Teenage Pregnancy



(e) Formal Employment at 25 yo



(f) Entrepreneurship



Notes: This figure plots reduced-form estimates of Equation 2.1.

entrepreneurship, and 2.8 p.p.(7.3%) in formal employment at 25 years old; while 0.5 p.p (8.3%) and 2.3 p.p. (6.1%) decreases in teenage pregnancy and adult-enrollment in PBF.

All plots have age-patterns similar to the first stage (Figure Figura 14), with no strong evidence of pre-trends. Some minor exceptions, however, can be justified by aspects of the program. For instance, there are slight pre-trends for labor market outcomes (Figures Figura 15e and Figura 15f). As discussed in Section 2.2.2, "children" who were older than 18 when their families entered the program could change their labor supply and entrepreneurship decisions upon the attenuation of family-level financial burden. Nevertheless, there is a clear trend-break for fully treated children, pointing to the effect of conditionalities. Moreover, the absence of pre-trends for outcomes determined before 18 years old – in particular, high school completion and teenage pregnancy – is a strong validation of the research design.

In Table Tabela 2, we present IV estimates expressing long-term effects of PBF in terms of the number of childhood years spent as a beneficiary of the program. Each additional year contributes to 1.8 and 1.1 p.p. in the likelihood of high school and college education completion, respectively (Panel A, Columns 1-2). Likewise, one year under the program is associated with decreases of 0.2 p.p. in teenage pregnancy, 1.3 p.p. in adult enrollment in PBF, and 0.1 p.p. in hospitalizations and incarceration (Panel A, Columns

3-8). Finally, we estimate per-year increases of 0.8 p.p. in entrepreneurship and 1.7 p.p. in formal employment (Columns 9-10).

Panels B and C show effects for males and females, respectively. Education gains are larger for girls, both in absolute and relative terms. One year under the program adds 2 p.p. in high school completion and 1.3 p.p. in college degree likelihoods for girls, while 1.6 p.p. and 0.7 p.p. for boys. Labor market gains are also larger for girls (Columns 9-11). Gender heterogeneity also provides some useful placebo exercises. For instance, Columns 3 and 4 show that effects on adult enrollment at PBF and teenage pregnancy are driven by females, while the decrease in incarceration is only present for males (Column 6). Race heterogeneity is more nuanced. Panels D and E show estimates for whites and non-whites, respectively. While relative education gains are somewhat mixed, labor market improvements seem larger for non-whites. However, the decrease in adult PBF enrollment is mainly driven by whites. Such pattern can relate to long-lasting determinants of racial mobility gaps (64).

To confound our estimates, any contemporaneous shock or policy other than PBF would have to be correlated with childhood exposure to the program and replicate the shape of the effects in Figure Figura 15 across cohorts and outcomes. In particular, one could be worried about other public policies around PBF implementation that targeted poor families and/or could affect children's human capital accumulation. In the Appendix B.2.2, we add controls to the vector X_{imt} to absorb further cohort-specific shocks of possible confounders and later include municipality-by-cohort fixed effects in our estimations.

Tabela 2: IV Estimates of Long-term Effects of PBF

	High School	College	PBF	Teen Preg.	Childbirth	Hosp.	Crime	Jail	Entr.	Empl.	Wage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
						A. Full Sample					
$Years_{PBF}$	0.0177***	0.0111***	-0.0129***	-0.0020***	-0.0040***	-0.0014***	0.0003	-0.0012***	0.0078***	0.0166***	46.66***
	(0.0020)	(0.0007)	(0.0005)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0003)	(0.0005)	(0.0013)	(3.271)
DV Mean	0.619	0.120	0.141	0.036	0.079	0.055	0.026	0.019	0.078	0.410	795.76
Obs	24,071,579	24,071,579	27,244,061	27,244,061	27,244,061	27,244,061	27,244,061	27,244,061	27,244,061	27,244,061	27,244,061
						B. Males					
$Years_{PBF}$	0.0163***	0.0075***	-0.0025***	0.0000	0.0000	-0.0003	0.0004	-0.0024***	0.0078***	0.0080***	39.45***
	(0.0024)	(0.0008)	(0.0006)	(0.0000)	(0.0001)	(0.0007)	(0.0005)	(0.0006)	(0.0007)	(0.0021)	(5.376)
DV Mean	0.578	0.095	0.026	0.000	0.002	0.055	0.041	0.034	0.088	0.453	928.55
Obs	12,114,165	12,114,165	13,775,005	13,775,005	13,775,005	13,775,005	13,775,005	13,775,005	13,775,005	13,775,005	13,775,005
						C. Females					
$Years_{PBF}$	0.0200***	0.0126***	-0.0220***	-0.0035***	-0.0082***	-0.0027***	0.0000	-0.0002	0.0070***	0.0252***	50.59***
	(0.0032)	(0.0012)	(0.0013)	(0.0008)	(0.0013)	(0.0008)	(0.0003)	(0.0002)	(0.0006)	(0.0018)	(4.096)
DV Mean	0.660	0.143	0.259	0.073	0.157	0.056	0.010	0.003	0.007	0.365	659.96
Obs	11,957,414	11,957,414	13,469,056	13,469,056	13,469,056	13,469,056	13,469,056	13,469,056	13,469,056	13,469,056	13,469,056
						D. Whites					
$Years_{PBF}$	0.0109***	0.0135***	-0.0126***	-0.0021***	-0.0040***	-0.0021***	-0.0004	-0.0018***	0.0090***	0.0229***	64.23***
	(0.0015)	(0.0008)	(0.0009)	(0.0006)	(0.0008)	(0.0008)	(0.0005)	(0.0004)	(0.0006)	(0.0013)	(5.042)
DV Mean	0.694	0.160	0.100	0.031	0.076	0.056	0.029	0.018	0.010	0.538	1,115.89
Obs	12,595,215	12,595,215	12,643,398	12,643,398	12,643,398	12,643,398	12,643,398	12,643,398	12,643,398	12,643,398	12,643,398
						E. Non-Whites					
$Years_{PBF}$	0.0251***	0.0063***	-0.0064***	-0.0024***	-0.0039***	-0.0009	0.0006*	-0.0012***	0.0032***	0.0258***	40.19***
	(0.0031)	(0.0005)	(0.0011)	(0.0006)	(0.0009)	(0.0008)	(0.0003)	(0.0004)	(0.0003)	(0.0019)	(3.532)
DV Mean	0.513	0.052	0.236	0.050	0.095	0.058	0.021	0.017	0.035	0.378	615.75
Obs	10,545,689	10,545,689	10,645,457	10,645,457	10,645,457	10,645,457	10,645,457	10,645,457	10,645,457	10,645,457	10,645,457

Notes: The table reports second-stage estimates from Equation 2.3. Coefficients express the effect of 1 additional childhood year spent as a PBF beneficiary on long-term outcomes. (*p<0.1; **p<0.05; ***p<0.01).

2.4 Spillovers

2.4.1 Empirical Strategy

PBF was large, and the literature suggests that could be spillovers to non-eligible families. In this section, we restrict our analysis to children in the control group in Section 2.3, that is, those with mothers with a low propensity to be PBF beneficiaries. We assess potential spillovers with a design that compares younger and older siblings in municipalities with varying treatment intensities – measured by the number of PBF beneficiaries per capita. Intuitively, we contrast siblings' differences in cities with small versus large shares of the population under the program. If spillovers exist, they should be present for younger cohorts and larger for children born in cities where the program was relatively more present.

As a proxy of cross-municipality PBF intensity, our main independent variable is the number of PBF slots per capita in 2003.[42] Appendix Figure Figura 42a maps the distribution of this measure across the country, while Appendix Figure Figura 42b shows it is a good predictor of PBF intensity over time.

We run standard differences-in-differences regressions to estimate the effects of growing up in a city that had relatively more PBF beneficiaries:

$$y_{iamt} = Young_t \times Slots_m \pi + \varepsilon_{imt} \tag{2.4}$$

where $Young_t$ is a dummy equal to one for children younger than 18 in 2004 and $Slots_m$ is the number of PBF slots allocated to city m in 2003 per capita. The specification contain the same set of fixed effects and controls (omitted) as Equation 2.1. Now, the identifying assumption is that cross-cohort trends among non-eligible children in low-versus high-exposed cities would not change absent the program.

2.4.2 Results

Panel A of Table Tabela 3 presents the long-run effects of growing up in a high-exposed city estimated for the full sample of non-eligible children. Columns 1-2 document that a 10 p.p. increase in PBF coverage is associated with 0.6 p.p. (0.8%) and 0.3 p.p. (1.5%) increases in high school completion and college degree attainment, respectively. Columns 3-8 show that we do not find significant effects for teenage pregnancy, but small reductions in adult PBF enrollment (2.9%), criminal prosecution (4%), incarceration (3.2%), overall hospitalizations (3.2%), and complications during pregnancy/childbirth (4.2%). Likewise, columns 9-11 point to gains in entrepreneurship (3.5%) and formal employment (2.5%), as well as in real wage (0.9%).

To investigate potential mechanisms, we estimate spillover effects according to mothers' education. In Table Tabela 3, the subsequent panels present estimates for children whose mothers did not complete high school (panel B), complete high school (C),

[42] We only use the first allocation of slots, computed before the start of the program. We compute slots from survey data using the formulas described in (76)

Tabela 3: Long-term Spillover Effects of PBF

	High School (1)	College (2)	PBF (3)	Teen Preg. (4)	Childbirth (5)	Hosp. (6)	Crime (7)	Jail (8)	Entr. (9)	Empl. (10)	Wage (11)
A. Full Sample (avg. exposure: 0.18 years)											
$Slots_m \times Young_t$	0.0604***	0.0299***	-0.0252***	0.0014	-0.0255***	-0.0145***	-0.0104***	-0.0054**	0.0405***	0.1146***	89.83**
	(0.0122)	(0.0085)	(0.0065)	(0.0037)	(0.0060)	(0.0050)	(0.0031)	(0.0023)	(0.0065)	(0.0127)	(36.11)
DV Mean	0.750	0.192	0.027	0.088	0.025	0.017	0.045	0.061	0.116	0.466	993.75
Obs	13,610,843	13,610,843	15,857,487	15,857,487	15,857,487	15,857,487	15,857,487	15,857,487	15,857,487	15,857,487	15,857,487
B. Mothers < high school (avg. exposure: 0.97 years)											
$Slots_m \times Young_t$	0.0714***	0.0282***	-0.0307***	0.0046	-0.0210***	-0.0173***	-0.0096***	-0.0059**	0.0442***	0.1293***	62.81*
	(0.0137)	(0.0080)	(0.0074)	(0.0045)	(0.0069)	(0.0060)	(0.0035)	(0.0025)	(0.0062)	(0.0145)	(37.71)
DV Mean	0.645	0.083	0.041	0.136	0.033	0.027	0.058	0.087	0.070	0.526	904.49
Obs	9,400,568	9,400,568	10,813,914	10,813,914	10,813,914	10,813,914	10,813,914	10,813,914	10,813,914	10,813,914	10,813,914
C. Mothers with high school (avg. exposure: 0.67 years)											
$Slots_m \times Young_t$	0.0123	0.0430**	-0.0073	-0.0080	-0.0463***	-0.0164	-0.0145*	-0.0058	0.0287*	0.0585**	157.1*
	(0.0217)	(0.0217)	(0.0153)	(0.0086)	(0.0135)	(0.0123)	(0.0078)	(0.0057)	(0.0162)	(0.0270)	(94.52)
DV Mean	0.786	0.170	0.023	0.063	0.025	0.015	0.044	0.051	0.098	0.448	965.41
Obs	2,718,035	2,718,035	3,069,292	3,069,292	3,069,292	3,069,292	3,069,292	3,069,292	3,069,292	3,069,292	3,069,292
D. Mothers with college (avg. exposure: 0.12 years)											
$Slots_m \times Young_t$	0.0374	0.0853***	-0.0103	-0.0108	-0.0347***	0.0090	-0.0067	-0.0046	0.0520**	0.0740**	238.3*
	(0.0239)	(0.0294)	(0.0126)	(0.0080)	(0.0133)	(0.0126)	(0.0090)	(0.0048)	(0.0223)	(0.0291)	(127.9)
DV Mean	0.911	0.439	0.009	0.021	0.018	0.005	0.026	0.024	0.183	0.384	1,189.29
Obs	1,492,240	1,492,240	1,974,281	1,974,281	1,974,281	1,974,281	1,974,281	1,974,281	1,974,281	1,974,281	1,974,281

Notes: The table reports estimates from Equation 2.4. Coefficients express the effect of a 100 p.p. increase in municipality-level PBF coverage for low-exposure children. (*p<0.1; **p<0.05; ***p<0.01).

and earned a college degree (D).[43] Noticeably, most of the effect documented in Panel A is driven by low-educated mothers, in particular for high school completion (+1.1%) and adult PBF enrollment (-2.7%). The fact that we find null effects for these outcomes for medium- and high-educated mothers is reassuring, as we should not expect richer families to change children's secondary-schooling attainment due to PBF or be enrolled in the program at the baseline.

Interestingly, we find meaningful long-term gains in college education attainment, entrepreneurship, and formal employment across all levels of mothers' education. In particular, for children of college-educated mothers, a 10 p.p. increment in municipal-level PBF coverage is associated with gains of 0.8 p.p. (3.6%) in college attainment, 0.5 p.p. (2.7%) in entrepreneurship rates, and 0.7 p.p. (1.8%) in formal employment. In the Appendix B.3.2 we run some robustness exercises including additional municipal characteristics interacted with linear cohort trends to the X_{imt} vector.

2.5 PBF and Intergenerational Mobility

In Section 2.3, we document that PBF increased the human capital accumulation and labor market outcomes of treated children. Since PBF improved the living conditions

[43]The sum of observations in these groups do not add up to the number of observations in Panel A as we drop mothers for which we do not observe education.

of the poor, these results suggest the program promoted (absolute) upward intergenerational mobility. At the same time, Section 2.4 shows how non-eligible (richer) children in areas in which PBF was more intense had improvements in many outcomes, in particular college-education attainment. Hence, it is not clear whether PBF improved relative intergenerational mobility, i.e. the gap in adult outcomes of children born in rich versus poor families.

To assess PBF effects on intergenerational mobility we compute two measures of mobility at the municipality-cohort level. We measure absolute mobility by the probability of children whose mothers did not complete high school succeeding to do so, and relative mobility by the gender-adjusted intergenerational correlation in years of education. We then compare cohorts in high- versus low-exposed municipalities in an event study design:

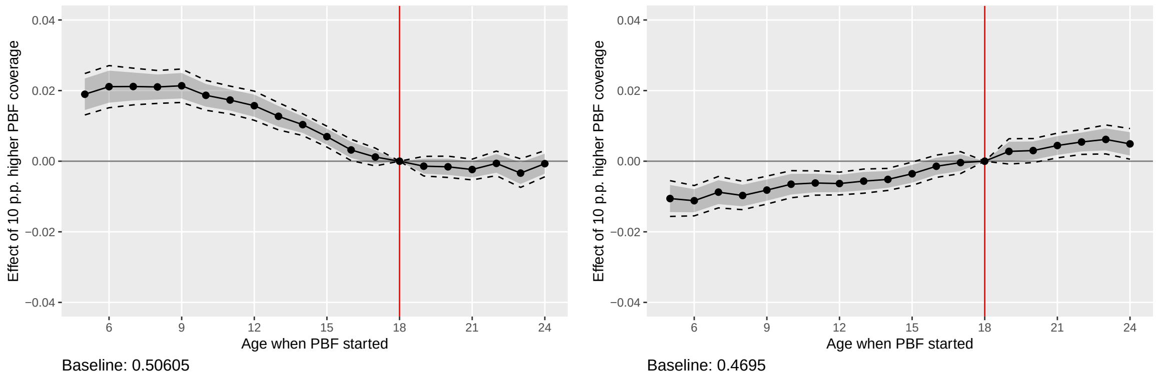
$$y_{mt} = \sum_{t=6}^{17} \left(\boldsymbol{I}_t \times Slots_m\right) \phi_t + \sum_{t=19}^{24} \left(\boldsymbol{I}_t \times Slots_m\right) \mu_t + \vec{X}'_{mt}\Omega + \lambda_t + \delta_m + \gamma_{s(m),t} + \varepsilon_{mt} \quad (2.5)$$

where y_{mt} is a mobility measure in city m for cohort t and $Slots_m$ is the same as in Equation 2.4. City (δ_m) and cohort (λ_t) fixed effects define the differences-in-differences design, while state-by-cohort fixed effects ($\gamma_{s(m),t}$) and controls (X_{mt}) are equal to previous estimations.

In Figure Figura 16, we plot event studies for the absolute (Figura 16a) and relative (Figura 16b) mobility. Cities with higher PBF coverage experienced larger improvements in upward and relative mobility. The clear shift in cohort trends similar to previous estimates reinforces the research design and indicates that a 10 p.p. increase in PBF coverage is associated with increments up to 2.0 p.p. (3.9%) in absolute mobility and 1.0 p.p. (1.8%) in relative mobility.

Figura 16: PBF Effect on Intergenerational Mobility

(a) Absolute Educational Mobility (b) Relative Educational Mobility



Notes: This figure plots estimates from Equation 2.5. Coefficients report the effect of a 10 p.p. increase in municipality-level PBF coverage on municipality-cohort measures of absolute and relative educational mobility.

Regions with larger PBF coverage thus had larger improvements in social mobility. These areas were in the less developed Northern and Northeastern regions of the country. Therefore, this suggests that PBF could have decreased the large regional gaps in upward mobility documented in (64).

2.6 Conclusion

We find evidence that PBF fulfilled its long-term goal of reducing the intergenerational persistence of poverty. Poor children exposed to the program are more educated adults and show better health and labor market outcomes. Moreover, spillovers to low-exposure children suggest that the program had larger aggregate effects that deserve more attention in future research. Finally, PBF contributed to decreasing the intergenerational gap between low- and high-income families and reducing the regional gap in opportunities in Brazil.

A APPENDIX TO CHAPTER 1

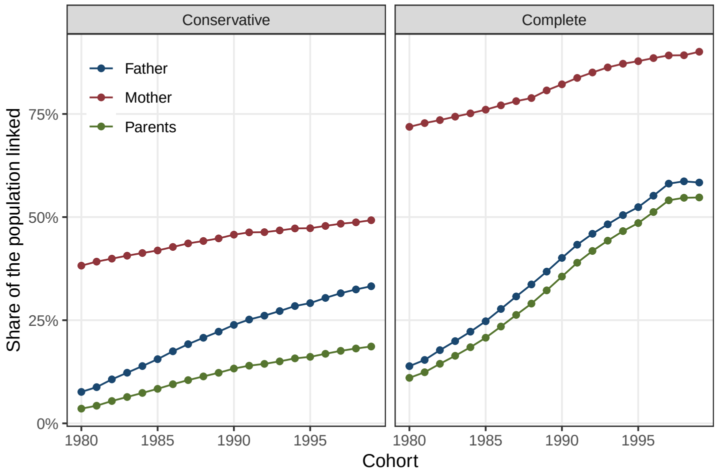
A.1 Data and Measurement

A.1.1 Family links

Our main analysis is based on a conservative family linkage procedure focused on avoiding any erroneous links. For robustness purposes, we assemble an expanded sample based on an additional, less conservative procedure dropping the restriction that the parent has an unique name. Instead, we link children to parents if they (i) share the same name, although not unique in the population; (ii) they resided at the same address in 2019; and (iii) their age difference is in a given range – 10-55 and 10-60 years for mothers and fathers, respectively. As before, the process is run separately for mothers and fathers. Although this procedure is somewhat less conservative relative to the original linkage, it remains highly accurate given that addresses provided in the person registry are extremely granular, including the apartment/house/unit in addition to the street name, number, neighborhood and postal code.

Figure Figura 17 plots the share of children from each cohort linked to their parents when using either procedure. Many more children can be linked to their parents – particularly mothers, since mothers' names are available for the entire population whereas fathers' names are available for roughly two-thirds of the population in the welfare registry. In addition, the share of successful links is increasing over time because younger cohorts can be claimed throughout more childhood years in the tax data, which start in 2006.

Figura 17: Number of parent-child links relative to the population by cohort

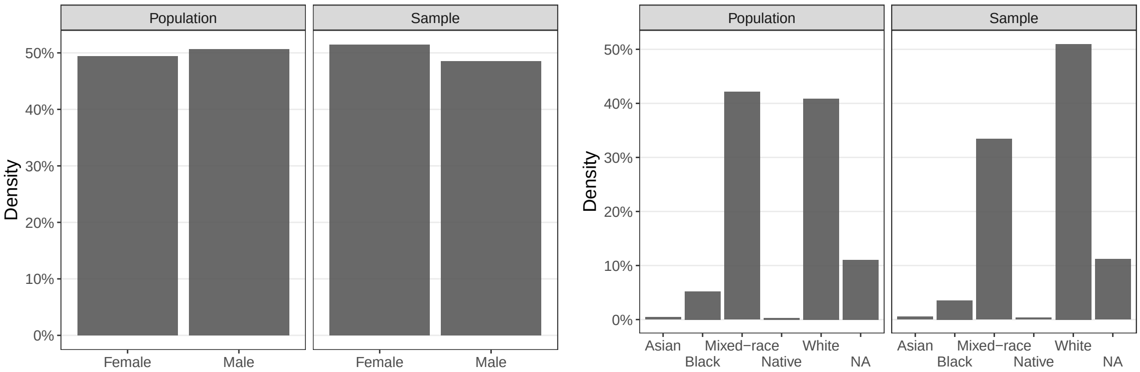


Notes: The figure plots the share of the population that can be linked to their parents by cohort following our baseline, more conservative method (left graph) and the alternative, less conservative method (right graph). The first method links children to parents using unique person codes in dependent claims tax data and using names for uniquely named parents in population and welfare registries. The second method allows for additional links using individual names and addresses.

A.1.2 Sample selection

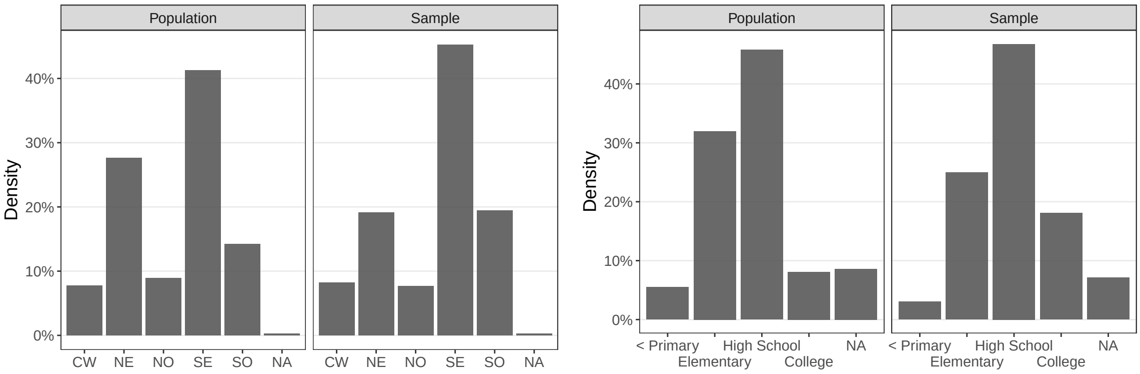
Figure Figura 18 shows that the distributions of gender, race, region of residence, and education are broadly similar between our baseline sample, composed by children in the 1988-1990 cohorts that can be linked to both parents, and the same cohorts in the total population. Whites and women are slightly overrepresented in the sample because some family links are based on unique names that slightly oversample white individuals – as white surnames are rarer – and because women are slightly overrepresented in the welfare register used to link children to their father. Likewise, our main sample is slightly more concentrated in the Center-South region of the country and has above-average education. Columns 1-2 in Table Tabela 4 provide descriptive statistics for the population and our main sample. The standardized differences in column 3 are below the critical value .2 for all but three variables that slightly exceed the cutoff (race, college education and living in the North-East), indicating only small differences in the underlying distributions (85). Nevertheless, in light of these small differences, as a robustness test we show in Section A.2.1.1 that our main findings are unaffected: (i) when substantially enlarging the sample by using the less conservative procedure to link families (see Section A.1.1 above), and (ii) re-weighting the sample to perfectly match the first and second moments of several characteristics in the population using the entropy algorithm by (86) (see Table Tabela 4, colums 4-5).

Figura 18: Distribution of sample characteristics



(a) Gender

(b) Race



(c) Region of residence

(d) Education Level

Notes: This figure plots the distribution of gender (a), race (b), region of residence (c), and educational level (d) in the population (left panels), and in our main sample (right panels). In plots (b), (c), and (d), we include missing data as an additional category.

Tabela 4: Descriptive statistics of main sample

	Population	Sample	Std. Diff.	Weighted	Std. Diff., W
Female	0.494	0.512	0.037	0.494	0.000
Non-White	0.532	0.417	0.232	0.532	0.000
Primary	0.060	0.033	0.130	0.060	0.000
Elementary	0.350	0.270	0.172	0.350	0.000
High School	0.501	0.503	0.004	0.501	0.000
College	0.089	0.194	0.305	0.089	0.000
Welfare	0.624	0.581	0.086	0.622	0.003
Formal Job	0.858	0.899	0.124	0.856	0.006
Cohort 1988	0.344	0.319	0.054	0.344	0.000
Cohort 1989	0.337	0.336	0.001	0.337	0.000
Cohort 1990	0.319	0.345	0.055	0.319	0.000
North	0.089	0.077	0.044	0.089	0.000
Northeast	0.277	0.192	0.201	0.277	0.000
Southeast	0.414	0.453	0.080	0.414	0.000
South	0.143	0.195	0.140	0.144	0.005
Center-West	0.077	0.082	0.018	0.076	0.006
State capital	0.251	0.290	0.088	0.251	0.000

Notes: The table compares the average characteristics of our main sample of children born in 1988-1990 (omitting missing values) with the average characteristics of the same cohorts in the general population. The means for each variable are presented (columns 1-2), along with the standardized difference (column 3), the mean in the main sample after re-weighting observations to match the first and second moments of population characteristics (86) (column 4), and the standardized difference between the samples in columns 1 and 4. All variables are recorded as dummy indicators.

A.1.3 Imputation method based on random forests

We use PNAD and population censuses to assemble a repeated cross-section from 1991 to 2019 of all adults aged 18-65 in any occupation – formal, informal, firm owner, or self-employed. We leverage the high-quality information contained in these surveys to train an RF model to predict informal income each year. We repeat the same process for estimating formal non-labor income when tax data are not available. We implement the generalized RF algorithm as developed by (49).

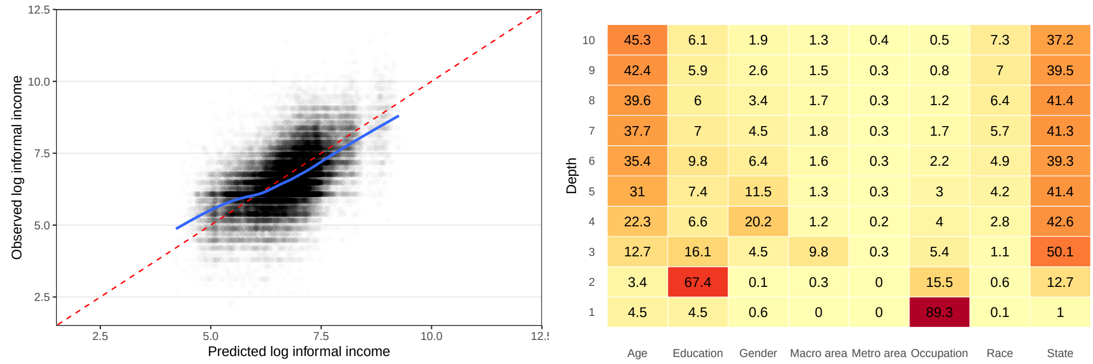
An RF is a collection of trees, each one endogenously splitting the covariate space to predict our outcomes of interest. To generate each tree, the algorithm starts by sampling without replacement from the survey dataset defining a root node. The root node is split over the space of covariates into child nodes as follows. A random subset of covariates are selected as candidates to split on, and the algorithm selects the split that maximizes heterogeneity in the prediction outcome. The sample splits take place recursively until a stop criterion met, so that overfitting is avoided.

For each outcome (informal income and formal non-labor income) and each year of data, we grow a RF with 5,000 trees, feeding the algorithm with a random subset of the original survey data. We grow *honest* forests, meaning that separate sets of data, selected at random, are used for splitting the node and estimating prediction improvements (49). We tune all parameters of the model via cross-validation – namely, the number of variables selected at each split, the minimum node size, the penalization for imbalanced splits, and the subsample fraction for honest splits. The model covariates are state of residence (27), state capital dummy, gender, age, four education dummies, a white/non-white dummy, and worker category. Following the literature on the representation of categorical variables in ML models (87), we encode state of residence dummies as a set of real-valued covariates related to demographic and socioeconomic indicators.[44]

We generate RF predictions to test prediction accuracy. Figure Figura 19a plots the predicted (x-axis) versus observed (y-axis) values for 10,000 workers not used to train the model, so that overfitting is not an issue, taking the year of 2019 as an example. The blue line is a best-fit third-order polynomial that closely tracks the 45-degree line (in red). In turn, Figure Figura 19b reports the split frequency plot of the model, indicating which variables are more relevant for income prediction. We find similar patterns for the estimated models predicting income for other years.

[44]We define formally self-employed as owners of formal firms with zero employees, and firm owners as those owning firms with at least one formal spell in the year. We consider all working-age adults who are not formal workers or firm owners as informal workers (i.e., the residual occupation category). For census years, we use more detailed information on the municipality of residence instead of the state of residence, since the former is available in the data.

Figura 19: Random forest fitting



(a) Prediction performance

(b) Variable split frequency

Notes: Panel (a) plots the relationship between predicted (horizontal axis) and observed (vertical axis) log informal income in 2019 for 10,000 individuals not used in the training process, where the blue line is a fitted third-order polynomial. Panel (b) reports the RF split frequency for up to ten levels of depth. Each cell shows the share of splits conducted at each covariate (columns) in each depth (rows).

A.1.4 Description of data sources

We use several datasets to track additional individual outcomes:

- Person Registry: The *Cadastro de Pessoa Física* is the administrative population registry maintained by *Receita Federal*, the Brazilian tax authority. It contains all individuals who have ever held a Brazilian person code (CPF) – 255 million people in total. The CPF is similar to the social security number in the United States. Every individual in the country is identified by this unique and non-exchangeable code. Besides the person code, each observation has the person's full name, date of birth, gender, and the full name of the mother. If the person is dead, it contains the death year, which we use to create mortality outcomes.
- Address Registry: The tax authority provided us with a dataset containing the history of individuals' place of residence. The tax authority updates these addresses from several administrative sources, such as electoral registries and tax declarations, and when individuals autonomously update their information in the person registry. Each observation is identified by the individual's person code, the year when the address was updated and the full residential address (street name, number, apartment/house/unit, neighborhood and postal code). Overall, there are more than 500 million addresses, which we geocoded to longitude and latitude coordinates.
- Tax Returns: The tax authority also provided us with all personal income tax returns filed during the period 2006-2020. Each observation is identified by the returnee person code, all dependents' tax codes, and reported income divided into three categories: taxable income (mainly labor earnings and rents), tax-exempted income (mainly dividends, donations, and bequests), and income subjected to withheld or definitive taxation (mainly investment earnings and capital gains from real estate transactions). These data cover the period 2015-2019 for children in our main sample (cohorts 1988-1990) and the period 2006-2010 for their parents.
- Firm Ownership: The *Cadastro Nacional de Pessoa Jurídica* (CNPJ) is maintained by the tax authority and contains the universe of (formal) firms in Brazil, which are identified by a unique code (*CNPJ*), dates of opening and (eventual) closing, tax regime, city of registry, and a list of all shareholders identified by their person codes.
- Formal Employment: The *Relação Anual de Informações Sociais* (RAIS) is a linked

employer-employee administrative dataset covering the universe of firms and workers in the formal labor market, provided by the Ministry of Labor. We use all years of RAIS available, from 1985 to 2019. Employment spells are identified by the worker's person code and the firm's unique identifier (*CNPJ*),[45] workers' full name, gender, race, date of birth, and education; and complete information on the work contract such as dates of start and (eventual) termination, hours, wages, occupation.

- Welfare Registry: The *Cadastro Único* (CadÚnico) is an administrative registry maintained and constantly updated by the Ministry of Social Development to track the socioeconomic conditions of families with per capita income below half minimum wage or with total income below three minimum wages. It also includes all individuals of every family that has ever been a beneficiary of a federal social welfare program. We construct a yearly panel of CadÚnico from 2011 to 2020 with the individual's full name, gender, year of birth, race, education, and mother's and father's full names for more than 135 million individuals identified by their person codes. Each household is also identified by a unique identifier, which allows for the recovery of marriages and family structures.
- Hospitalization records: Individual-level data on admissions to public hospitals SIH-SUS (*Sistema de Internações Hospitalares*) for the period 2002-2019. It includes information on individual characteristics such as age, sex, municipality and zip code of residence, and descriptive information on the hospital admission, including the ICD-10 diagnostic, and date of admission. We use ICD-10 codes on hospitalization due to assaults to generate a measure of crime victimization. To merge these records to other datasets, we focus on individuals who can be uniquely identified by their postal code, gender and birth date – all of which can be observed for the entire population in the person registry, maintained by the Brazilian Tax Authority.

A.2 Income mobility at the national level

A.2.1 Additional robustness exercises

We now present a series of additional robustness exercises that further support our main results. In some of these analyses, we use additional birth cohorts born in the

[45] From 1985 to 2001, workers are identified by a different (unique) code, the PIS. We retrieve PIS-CPF pairs for all workers matching individuals across RAIS waves by their full name and date of birth.

1983-1990 period, additional parent-child links, and also vary the period when income is measured. Since we were granted access to tax data on our cohort of parents only for the period 2006-2010 and on our cohort of children only for the period 2015-2019, we rely on other sources of data on formal income to run some of these tests.[46] Therefore, we preliminarily show that mobility estimates for our main sample cohorts (1988-1990) are only slightly attenuated in magnitude when we do not rely on tax data, the rank-rank declining from 0.546 to 0.453 (column 1 of Table Tabela 5, Panel A).

A.2.1.1 Sample selection

Larger samples. Our baseline sample comprises 1.3 million children born during the period 1988-1990 that we can link to both parents, comprising 15% of all children in such cohorts. We show that our main results are robust to expanding the sample along three dimensions. First, we include all children that can be linked to their father (regardless of whether they are linked to their mother), which increases the sample size by 1 million, and run the analysis solely based on the father's income. The results in columns 2-3 of Table Tabela 5, Panel A, show that the father-child rank correlations in the baseline and enlarged samples are nearly identical (.44 and .45), and they are also identical to the baseline rank-rank slope estimated without tax data, reported in column 1.

Second, we expand the sample to include all cohorts born in 1983-1990, for a total of 3.5 million children. In this case as well, the estimated rank-rank coefficient remains identical.

Finally, Panel B of Table Tabela 5 replicates the analysis using the less conservative linking procedure described in Appendix Section A.1.1. Once again, all estimated coefficients are virtually unaffected (0.44 - 0.47).

[46]Specifically, we adopt the procedure laid out in Section 1.3.2 relying on formal employment data and the survey-based prediction process to measure formal income (as already done in our main analysis when tax data is unavailable).

Tabela 5: Robustness to larger samples

	1988-1990 cohorts			1983-1990 cohorts		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Baseline links						
Parent Rank	0.453***			0.455***		
Father Rank		0.440***	0.451***		0.437***	0.445***
Observations	1,304,586	1,304,586	2,361,010	3,416,131	3,416,131	3,901,433
Only father links			Yes			Yes
Panel B. Complete links						
Parent Rank	0.470***			0.462***		
Father Rank		0.457***	0.464***		0.443***	0.451***
Observations	3,552,156	3,552,156	6,486,256	9,374,785	9,374,785	10,749,696
Only father links			Yes			Yes

Notes: The table reports the estimated slope of the rank-rank regressions in equation (1.1), i.e. our (inverse) measure of income mobility, in different samples. In Panel A, we link parents to children using our baseline, conservative procedure; in Panel B, we expand the sample using the less conservative procedure described in Section A.1.1. Columns 1-3 cover the 1988-1990 cohorts – as in our main sample – while columns 4-6 cover the 1983-1990 cohorts. Columns 1-2 and 4-5 are based on children who can be linked to both parents, while columns 3 and 6 are based on children who can be linked at least to their fathers. The dependent and explanatory variables are always the child and parental income percentile rank. For consistency, in all specifications we measure income without using tax data, which are only available for the 1988-1990 cohorts. Income for the 1988-1990 cohorts is measured between 2015-2019, at ages 25-31 (as of baseline), while income for other cohorts is measured when they are 25-29 years old (*p<0.1; **p<0.05; ***p<0.01).

Reweighting. To address any residual concern about the representativeness of our sample, we re-weight the data to match a rich set of characteristics (up to their 2$^{\text{nd}}$ moment) in the general population, using the algorithm proposed by (86). Specifically, we balance our baseline sample with respect to gender, race, month and year of birth, state of residence (27), state capital dummy, education (4), and indicators for being in social welfare registries, formal labor market participation, and having a unique name in the country.[47] In Table Tabela 6, we report estimates of the rank-rank coefficient on the raw and re-weighted data (row 1 and 2, respectively) for the full sample (column 1), males (2), females (3), whites (4), and non-whites (5). Re-weighting has no effect whatsoever on the estimated rank-rank slope.

[47] Table Tabela 4 in Section A.1.2 displays descriptive statistics of the sample before and after reweighting together with standardized differences with respect to the population.

Tabela 6: Reweighting Procedure

	Full Sample	Males	Females	Whites	Non-whites
	(1)	(2)	(3)	(4)	(5)
Baseline	0.546***	0.469***	0.613***	0.521***	0.573***
Weighted	0.566***	0.504***	0.624***	0.546***	0.574***
Observations	1,304,586	633,489	671,097	672,015	546,773

Notes: The table reports the estimated slope of the rank-rank regressions in equation (1.1), i.e. our (inverse) measure of income mobility, in the raw data (first row) and in the re-weighted data (second row). The re-weighted data match the first and second moments of the population of children in the same birth cohorts along several characteristics, using the entropy algorithm by (86) (see Table Tabela 4). All samples cover the 1988-1990 cohorts and the dependent and explanatory variables are the child and parental income percentile rank (*p<0.1; **p<0.05; ***p<0.01).

A.2.1.2 Timing of income measurement

The timing when we measure income may lead to two main types of estimation bias, which we describe next. We provide several tests showing that our results remain similar are not affected by the specific timing that we choose. We estimate income mobility without relying on tax data, as the latter are not available for some of the cohorts and years required for such tests.

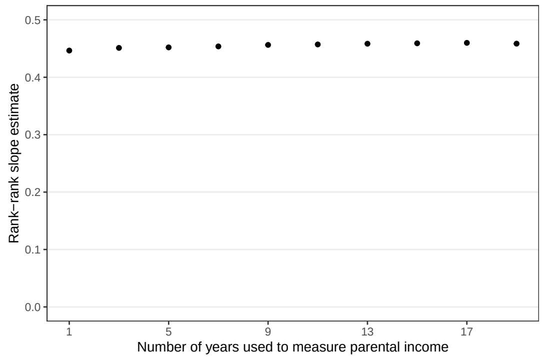
Attenuation bias. Measuring income for short time spans may attenuate estimates of income mobility, due to temporary income shocks. This is not a main concern when we measure parental income since we virtually cover children's entire childhood (from age 3 to 18). However, this could be a more relevant issue for child income in our main analysis, which uses a five-year window (age ranges 25-29, 26-30, 27-31 for the cohorts born in 1988, 1989 and 1990, respectively). In light of this, we show in Figure Figura 21a how the rank-rank slope changes as we vary the number of years used to measure parental income. The estimates are remarkably stable regardless of how many years are used in the analysis. Moreover, in Figure Figura 20b we show that estimates also remain largely stable when using 1 to 5 years to measure children's income. The estimates vary by less than 5 percentage points in both exercises, relative to the 0.453 rank-rank slope benchmark (without tax data). Overall, these results support the idea that the five-year window used to measure child income is sufficient to prevent meaningful attenuation bias in our main analysis.

Life-cycle bias. Measuring income too early may not adequately capture permanent income, possibly leading to life-cycle bias (88; 89). Again, this could be relevant when measuring children's income with a five-year window at the age range 25-31 in our main analysis. We use the fact that we can track parental income for a long period of time to study how our estimates change when measuring parental income at different ages. In particular, we focus on parents in our main sample born in 1960-1965. Figure Figura 20c shows that our estimates do not vary much when using a 3-year window to measure father's income centered from age 31 to 45. Next, in Figure Figura 20d, we show how the rank-rank slope changes as we center a three-year window around different ages for measuring children's income. We focus on the 1988 cohort, for which we can track income up to age 31. Again, the estimated rank-rank slope remains fairly constant within cohorts when income is measured at varying ages.

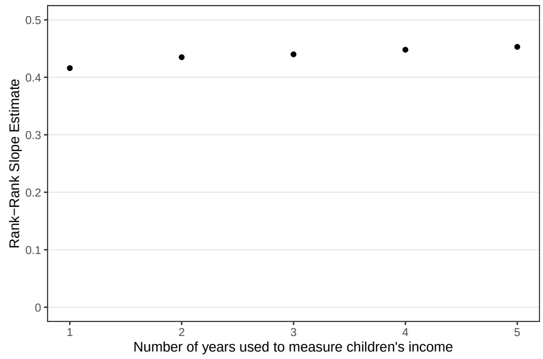
This result can be explained by the fact that there are little positional changes in annual income in Brazil from early ages, especially from the age of 24 (Figure Figura 20d); this is due, in turn, to the fact that most Brazilians enter the labor market relatively early, given that college enrollment is low. Positional changes below the age of 24 are concentrated at the very top of the income distribution, and they are driven by a large share of high-income children who attend college and delay entry into the labor market (see also Figure Figura 11a). Figure Figura 21 provides additional evidence on these aspects by showing a near parallel evolution of income by quartiles of completed education and parental income. These patterns are in contrast with the case of developed countries – as documented by (89) for the US and Sweden – where a much larger share of individual attend college.

Figura 20: Sensitivity of child and parental income to timing

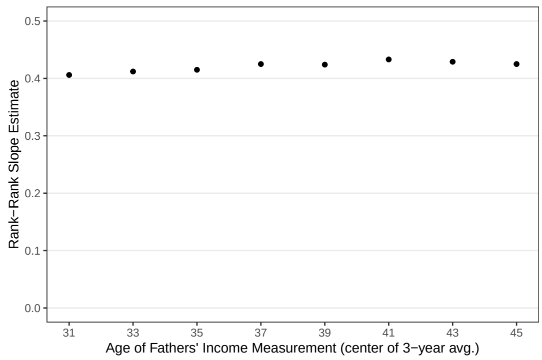
(a) Parents : Attenuation Bias



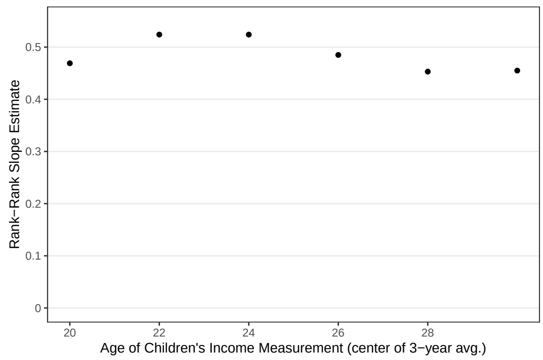
(b) Children : Attenuation Bias



(c) Fathers : Life-cycle Bias



(d) Children : Life-cycle Bias

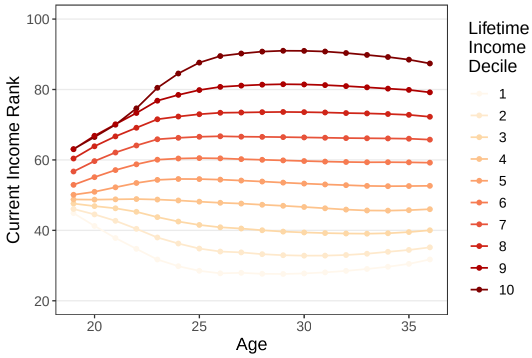


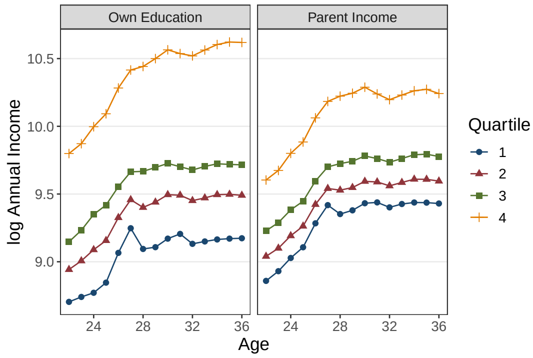
Notes: This figure plots robustness exercises for attenuation bias (a, b) and life-cycle bias (c, d). In Panels (a) and (c) child income is held constant and measured as in our baseline estimates, and we vary how parental income is measured. In Panels B and D, we measure parental income as in our baseline and vary how child income is measured. Panel A displays estimates of the rank-rank slope from separated rank-rank regressions in which we vary the number of years used to compute parental income, from 1 to 17 years, and centered at the age of 11. Panel (b) displays an analogous exercise in which we measure children's income using from 1 to 5 years, centered at age 27. In Panel (c), we run rank-rank estimates using father's income (rather than parental income) and vary the age when father's income is measured using a three-year window from ages 31 to 45. Finally, in Panel (d) we vary the age when we center the three-year window to measure children's income, from 20 to 30 years old. In Panels (a) and (b) we use our full baseline sample of the 1988-1990 cohorts. In Panel (c), we restrict the sample to children whose fathers are born between 1960-65 and focus on fathers' rather than parental income – to precisely gauge the sensitivity concerning different age windows. In Panel (d), the working sample is the 1988 cohort since income data is only available until 2019.

Figura 21: Evolution of children's position at the income distribution

(a) Evolution of income rank

(b) Evolution of log income



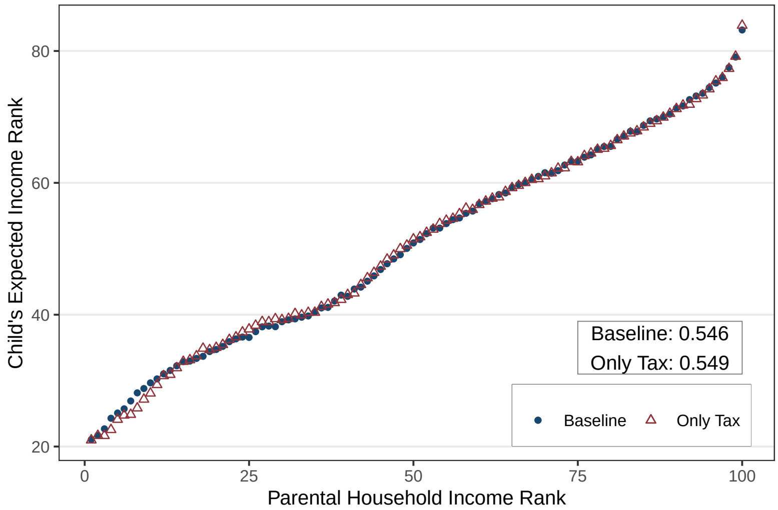


Notes: The figure plots the evolution of children's income distribution over time. Panel (a) shows the mean income percentile rank (on the vertical axis) when aged 18 to 36 (horizontal axis) for individuals in each decile of total lifetime income distribution. In turn, panel (b) shows the evolution of log incomes of the 1983 cohort when 22 to 36 years old, by quartiles of children's educational level (left) and parental income (right).

A.2.1.3 Estimates fully based on years when tax data are available

Our main analysis relies on an imputation procedure to predict the non-labor component of formal income when measuring parental income before 2006, when tax data are not available. In Figure Figura 22, we re-estimate the mobility curve while measuring parental income between 2006 and 2010, when tax data are available for parents. These results show that measuring parental (and child) income in years when tax data are available has little impact on the mobility curve, and offer additional support for our imputation procedure.

Figura 22: Mobility curve based fully on tax data



Notes: The figure plots our baseline mobility curve (blue dots) and the curve obtained using only years when tax data are available (red triangles). Both curves are based on our main sample (1988-1990 cohorts). In the baseline curve, children's income is measured between 2015-2019 and parental income is averaged over the years children are 3-18 years old. In the curve using only tax data, children's income is the same but parental income is measured in 2006-2010, when tax data are available. For each curve, the figure also reports our relative mobility measure based on Equation 1.1.

A.2.1.4 Alternative income and occupation definitions

We now show that two potentially relevant choices that we take to define total annual income have virtually no impact on our IGM estimates. Table Tabela 7 presents our baseline mobility estimate (column 1), along with alternative estimates that we describe next (columns 2-4). First, we rely on survey questions on "normal" monthly income to predict annual informal income and formal non-labor income (used when tax data are not available). In our main analysis, we extrapolated such income to the entire year, multiplying it by 12. We show that results do not change if we adopt a similar procedure for measuring formal labor income (derived from administrative employment data). Namely, we take the average monthly formal income while formally employed and multiple by 12 each year, instead of considering the sum over the year (column 2). Alternatively, we move back to our baseline but multiply predicted monthly informal income and formal non-labor income by the number of months that individuals spend out of formal employment in the year (rather than by 12) (column 3). Second, when predicting unobserved income in the main analysis, we label as informal workers in the administrative data those who do not hold any formal job in the entire year and who are not firm owners. We vary this assumption by defining informal workers as those who work formally for less than three consecutive months in the year and who are not firm owners (column 4).

Tabela 7: Alternative Income Measures

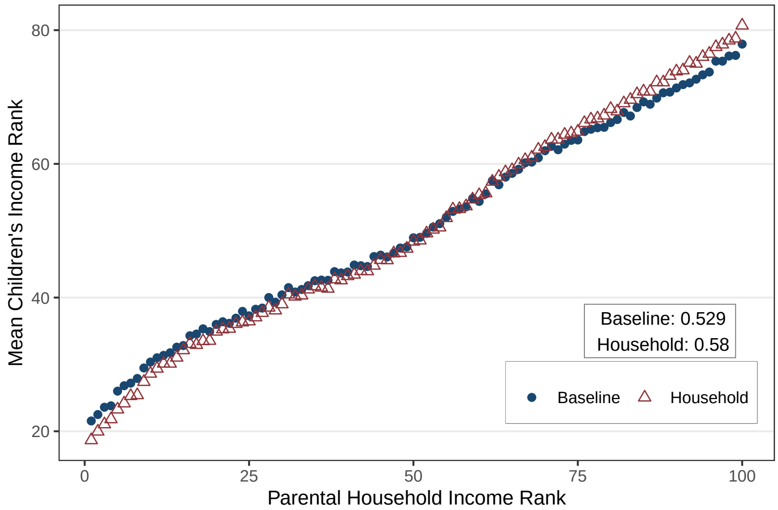
	(1)	(2)	(3)	(4)
Baseline	0.547***	0.545***	0.547***	0.547***
	(0.001)	(0.001)	(0.001)	(0.001)
Income definition	Baseline	Alternative 1	Alternative 2	Baseline
Informal workers definition	Baseline	Baseline	Baseline	Alternative
Observations	1,304,586	1,304,586	1,304,586	1,304,586

Notes: The table reports relative mobility estimates based on the slope of rank-rank regressions – as in eq. (1.1) – using different income and occupation definitions. It presents estimates when using our baseline income and occupation definitions (column 1); when measuring formal labor income by multiplying its monthly average in each year by 12 (column 2), when measuring predicted informal and formal non-labor income by multiplying the predicted monthly quantities by the number of months out of formal employment in the year (column 3); when defining informal workers as those who are formally employed for less than three months in the year and who are not firm owners. All samples cover the 1988-1990 cohorts and the dependent and independent variables are the child and parental income percentile rank (*p<0.1; **p<0.05; ***p<0.01).

A.2.2 Individual vs. household income for child ranks

We assess whether using household instead of individual income for children affects our results. We focus on the 38% of children who can be linked to their spouses in tax declarations and welfare registries (*CadÚnico*). We compute our baseline income measures for spouses in the same calendar years that their partner's income is measured starting from the first year when we observe both together. Household income is defined as the sum of both partners' individual income. Figure Figura 23 plots the mobility curves using individual (baseline) and household income for children in the married sample. Both curves perfectly overlap and are similar to the baseline mobility curve based on our main sample (Figure Figura 3). Overall, the exercise indicates that taking household or individual income has little impacts on our IGM estimates, which only become slightly larger, increasing from .529 to .58. They are also similar to our main estimates based on children's individual income, using the full sample rather than the married sample (.546).

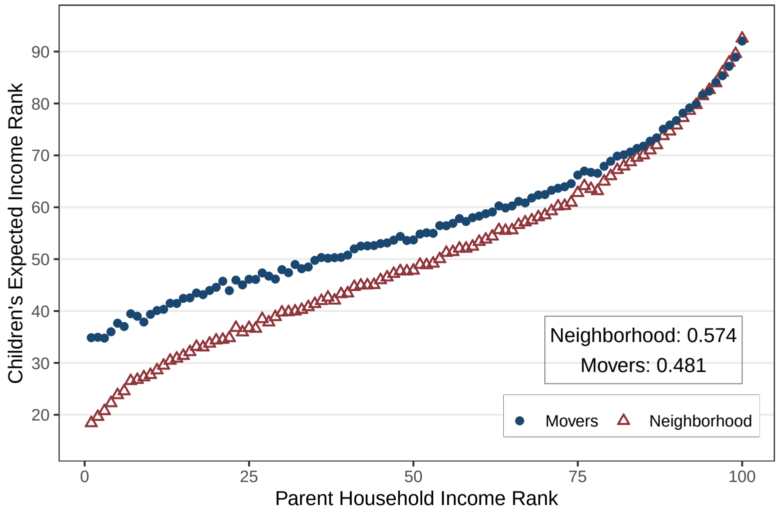
Figura 23: Mobility curve using household income for children



Notes: The figure plots the mobility curves using individual (blue dots) and household (red triangles) income for children, while parental income is the sum of father and mother's income, as in our baseline. The sample comprises the individuals in the baseline sample for which we are able to recover partners from tax declarations or *CadÚnico*. Household income is the sum of both partners' individual annual income starting from the year we observe them as a couple. For each curve, the figure also reports our relative mobility measure based on Equation 1.1.

A.2.3 Neighborhood-based measure for movers

Figura 24: Neighborhood-based measure: children changing address



Notes: This figure plots mobility curves based on the neighborhood-based measure, for our main sample (red triangles) and for a sample of movers who live in a different zip code than the place where they grew up (blue dots), both covering the 1988-1990 cohorts. For each parental rank in each curve, it plots the mean child rank. The neighborhood-based measure is given by the average formal income in the census tract where children grew up (parental rank) and where they live as adults (child rank). See Section 1.4.3 for a detailed description of these measures. For each curve, the figure also displays our relative mobility measure based on Equation (1.1).

A.2.4 Cross-country comparisons

A.2.4.1 Counterfactual exercises on the probability that children earn more than their parents

We study the probability A that children earn more than their parents and also implement a counterfactual exercise to study the factors driving cross-country differences, following (4; 61). This measure is defined as:

$$A = \int \int 1\{y(R) \geq y^P(R^P)\}\, C(R, R^P)\, dRdR^P \tag{A.1}$$

where $y(R)$ and $y^P(R^P)$ are the child and parental marginal income distributions at child income rank R and parental income rank R^P, respectively. **1** is an index function indicating whether real income at a given child rank is higher than income at a given parental income rank. The copula $C(R, R^P)$ is defined by the joint marginal distribution of child and parental income ranks.

The formula makes evident that the probability that children earn more than their parents depends both on the joint distribution of income across generations - the copula – and on the marginal distribution of income within generations. In turn, the latter can be decomposed into a component related to the average intergenerational income growth and a second component related to the distribution of income. We compute three counterfactuals for A, each of them abstracting from one of these factors to compare mobility in Brazil and in the US. First, we compute A using the US copula $C^{P,US}(R, R^P)$. Second, we use the US average income growth across generations to set the marginal income distribution for children in Brazil after decomposing the latter:

$$y(R) = s(R) \times Y^P(R) \times g^{US} \tag{A.2}$$

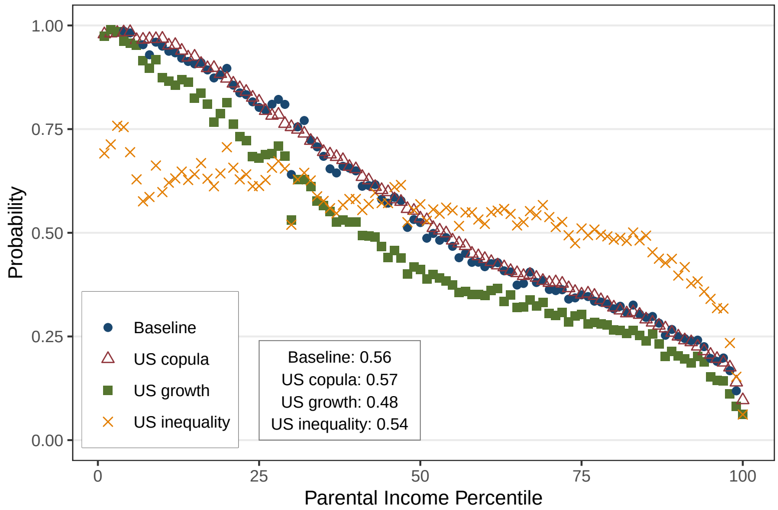
where $s(R)$ is the share of the aggregate child income accrued by children in rank R, $Y^P(R)$ is the aggregate parental income and g^{US} is the income growth rate in the US across generations. As such, we evaluate A under the counterfactual scenario where children in Brazil experience the average US intergenerational income growth. Finally, we estimate A using the US income distribution for children and parents:

$$y(R) = s^{US}(R) \times Y(R) \tag{A.3}$$

$$y^P(R^P) = s^{P,US}(R^P) \times Y^P(R) \tag{A.4}$$

We focus on father-son income, measuring their income at the same age (25-29), so to that our estimates are comparable to those available for the US (61) and Italy (4). The results are presented in Figure Figura 25, showing an average rate of .56 in Brazil, similar to the US (.55) and Italy (.53). Relative to these countries, sons at the bottom on the father's income distribution in Brazil have a higher probability of earning more than their parents, while the opposite is true for those at the top. This can be explained by the fact that (within-generation) income inequality is higher in Brazil, so that it is easier for children at the bottom to earn more than their parents, while it is more difficult for those at the top to do so. In fact, the counterfactual exercise using the US income inequality to estimate this measure for Brazil shows that mobility at the bottom would increase but, at the same time, mobility at the top would decrease. Overall, the net effect on the average probability that sons earn more than their fathers is small. In turn, we find the the joint distribution of parental and child income ranks has little impact on this measure which barely changes when running a conterfactual exercise with the US copula. Finally, when we use the US income growth across generations, the baseline probability that sons earn more than their fathers in Brazil decreases to .48. We conclude that this mobility measure for Brazil is similar to the US mainly thanks to the higher growth rate of income in Brazil.

Figura 25: Absolute Mobility, Probability That Sons Earn More than Fathers in Brazil



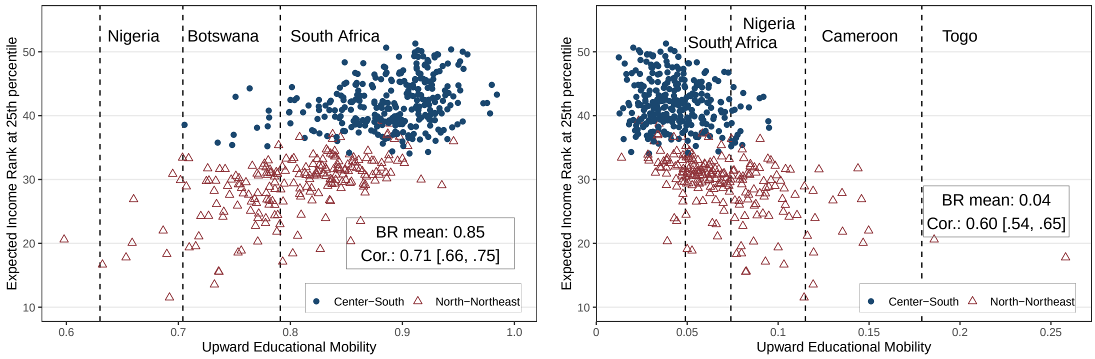
Notes: The figure shows the probability that the income of a son is at least as high as his father's income in real terms around the same age (measured at ages 25-29 for sons and 25-29 for fathers, at each fathers' income percentiles for our main sample (1988-1990 cohorts). Besides the baseline estimates (blue dots), the figures plots three counterfactuals replacing the Brazilian joint distribution of fathers' and sons' income ranks (red triangles), intergenerational income growth (green squares), and within generations income distribution (yellow X) by their US counterparts. The figure also displays the mean probability in each of these cases.

A.2.4.2 (1)

In Figure Figura 26, we analyze the relationship between our income-based upward mobility estimates (Section 1.5) and the education-based one, computed at the regional level. Three striking patterns emerge. First, the stark contrast between the North-Northeast and the Center-South of Brazil emerges for both measures. Second, educational mobility within Brazil spans a wide range of values which, overall, are comparable to the most mobile African countries (e.g., Nigeria, Botswana and South Africa). Finally, although the two measures display strong correlation, there is a large amount of variation in income mobility for given levels of educational mobility, and vice-versa. This highlights the fact that there is a lot of variation in income that is not explained by education, further motivating income-based analysis.

Figura 26: Educational mobility across Brazilian regions

(a) Upward mobility (b) Downward mobility



Notes: The figure plots estimates of upward (a) and downward (b) educational mobility across Brazilian regions (horizontal axis) versus baseline regional absolute mobility measures estimated in Section 1.5. Blue dots (red triangles) indicate regions in the Center-South (North-Northeast) region of Brazil. Vertical lines mark estimates of educational mobility for selected African countries from (1). Upward (downward) mobility is the likelihood of a child born to parents who did not (did) complete primary school succeeding (failing) to do so. The figure also reports the average upward (downward) educational mobility in Brazil and the cross-regional correlation between educational mobility and income mobility.

A.2.5 Gender, households, and assortative mating

Tabela 8: Labor market differences by gender and race

	Gender			Race		
Parent Quintile	Rank Gap	LFP Gap (pp.)	Wage Ratio	Rank Gap	LFP Gap (pp.)	Wage Ratio
1	17.7	17.6	0.84	10.4	10.0	0.95
2	18.8	13.2	0.84	9.2	8.4	0.95
3	15.2	8.2	0.84	8.6	7.2	0.96
4	10.8	4.6	0.87	6.4	6.3	1.01
5	7.1	1.9	0.89	3.4	6.5	1.12

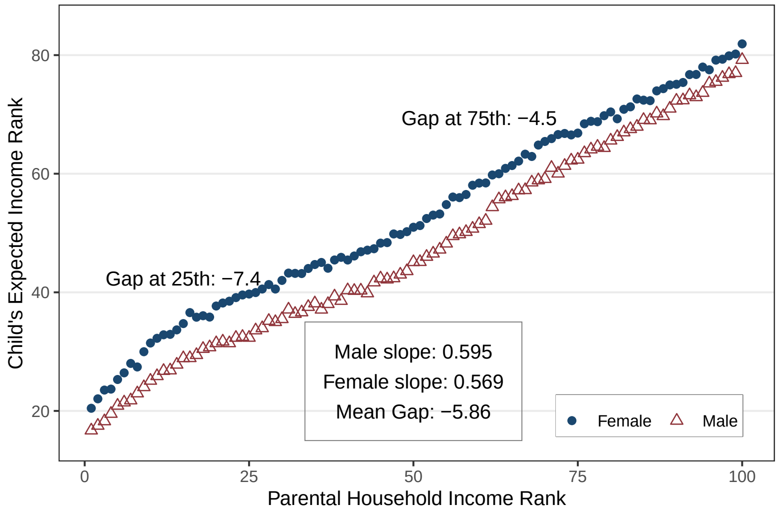
Notes: The table reports average gaps in child income ranks and labor market outcomes over gender and race, for each parental income quintile. Income rank gaps are calculated as the difference between average adult ranks for males (whites) and females (non-whites). The labor force participation (LFP) gap is the difference in average participation rate in the formal labor market between the two groups, in percentage points. Finally, the wage ratio is the ratio of the formal average wages of females (non-whites) to males (whites).

Tabela 9: Siblings comparisons by parental income quintile

Parental Quintile	Siblings Gap	Brother-Sister Gap
1	0.11	16.23
2	0.22	19.00
3	0.19	16.67
4	1.39	11.79
5	2.36	7.29

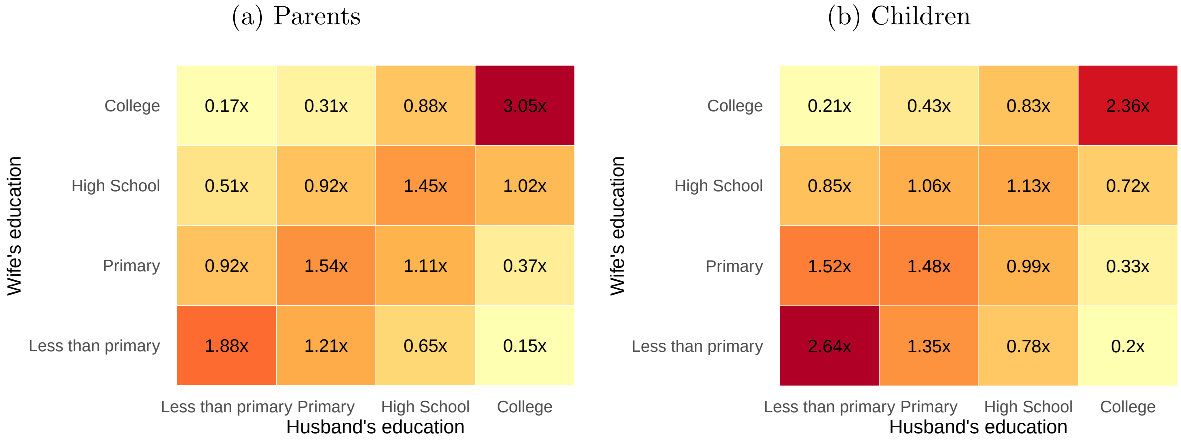
Notes: The table reports average gaps in income ranks between siblings for each parental income quintile. Siblings gaps are calculated as the difference between adult income rank of the older and younger siblings, regardless of gender. Brother-sister gaps are calculated as the difference between the male and female siblings, regardless of birth order. Both are calculated for individuals in our baseline sample of the 1988-1990 cohorts.

Figura 27: Household Income Mobility Curve, By Gender



Notes: The figure plots gender-specific mobility curves measuring parental and child income at the household level, using the married sample detailed in Appendix A.2.2 for the 1988-1990 cohorts. Females are the series in blue (dots), while males are the series in red (triangles). The ranks on both axes indicate the income positions relative to all individuals in their own cohorts (rather than each group). For each curve, the figure also reports our relative mobility measure based on Equation 1.1.

Figura 28: Educational assortative mating

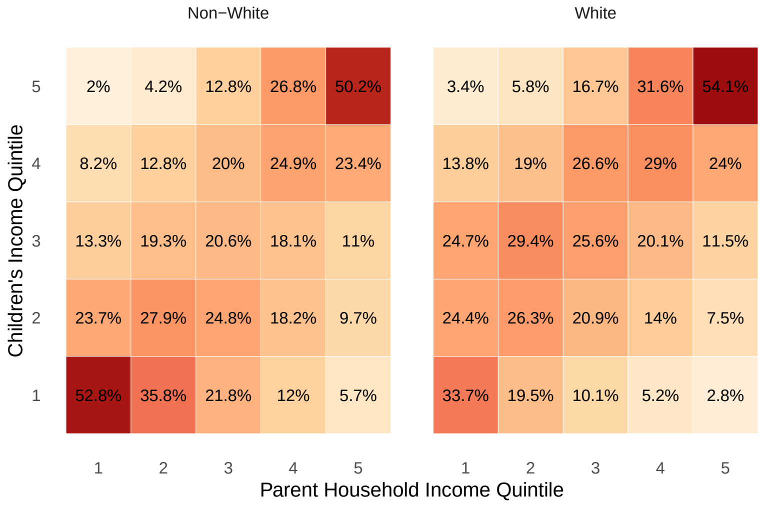


Notes: The figure plots educational sorting parameters for assortative mating between all parents in our baseline sample (panel A) and the sample of married children (panel B). We define assortative mating as a mating pattern in which individuals with similar traits mate with one another more frequently than would be expected under a random mating pattern, following the sorting parameter definition in (63). Darker colors indicate higher frequencies.

Figura 29: Assortative mating over parental income

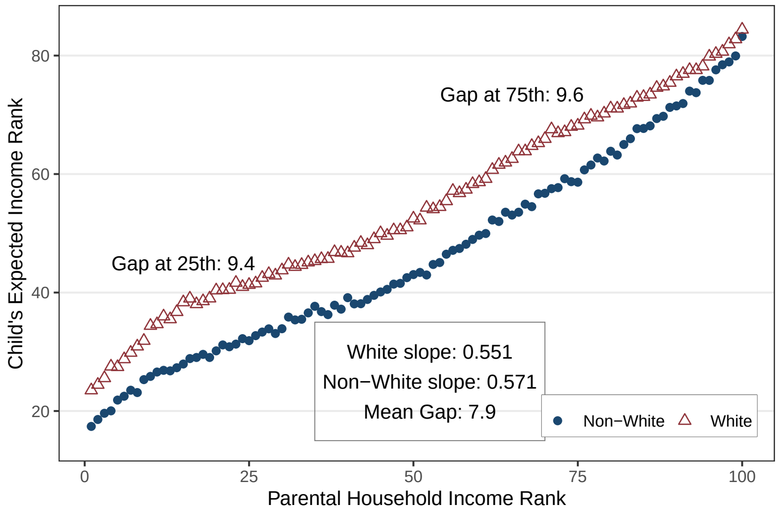
Notes: The figure plots sorting parameters for parental income assortative mating between children in the married sample for whom we can recover their partner's parental links. We define assortative mating as a mating pattern in which individuals with similar traits mate with one another more frequently than would be expected under a random mating pattern, following the sorting parameter definition in (63). Darker colors indicate higher frequencies.

Figura 30: Racial transition matrix



Notes: The figure shows the transition probability matrix by quintiles of the income distribution for the 1988-1990 cohorts separately for each race group. Each cell displays the share of children born in that parental income quintile (horizontal axis) who end up in a given income quintile in adulthood (vertical axis). Income quintiles in both axes indicate the income positions relative to all individuals in their own cohorts (rather than each group). Cells are colored according to the quintile-quintile transition probability, with darker red tones indicating higher likelihoods.

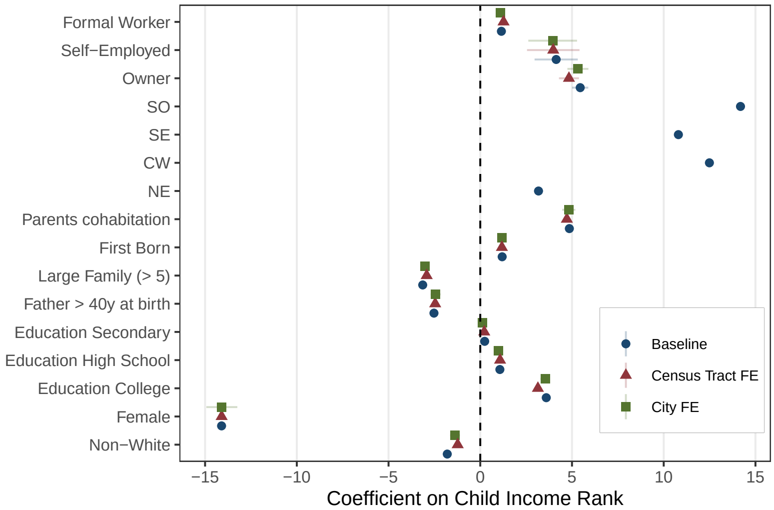
Figura 31: Household Income Mobility Curve, By Race



Notes: The figure plots race-specific mobility curves measuring parental and child income at the household level, using the married sample detailed in Appendix A.2.2 for the 1988-1990 cohorts. Non-whites are the series in blue (dots), while whites are the series in red (triangles). The ranks on both axes indicate the income positions relative to all individuals in their own cohorts (rather than each group). For each curve, the figure also reports our relative mobility measure based on Equation 1.1.

A.2.6 Factors explaining variation in child ranks beyond parental income

Figura 32: Variation in Income Mobility: Predetermined Factors

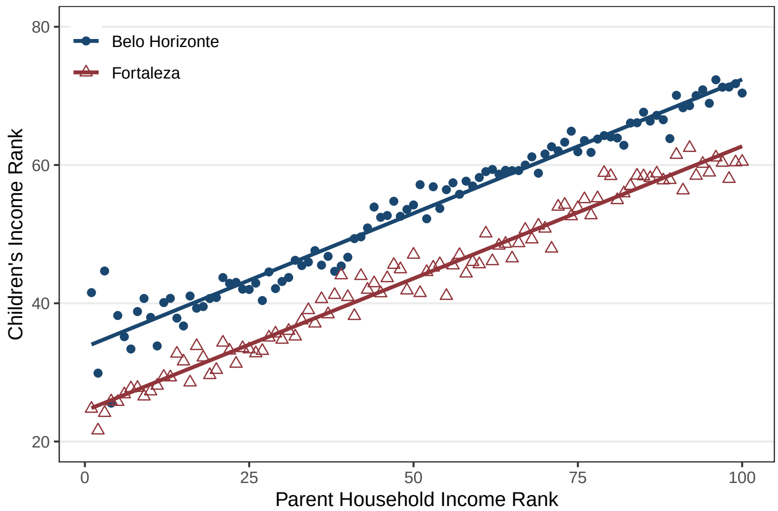


Notes: The figure reports coefficients of rank-rank regressions extended to include parental characteristics as explanatory variables in addition to parental income rank. The constant and coefficient on parental income rank are omitted. The dots in blue includes no fixed effects while the series in green and red add, respectively, parental city and census tract fixed-effects. All regressions are run in the baseline sample of the 1988-1990 cohorts.

A.3 Geographic variation in mobility

A.3.1 Individual mobility curves

Figura 33: Individual mobility curves in Fortaleza and Belo Horizonte

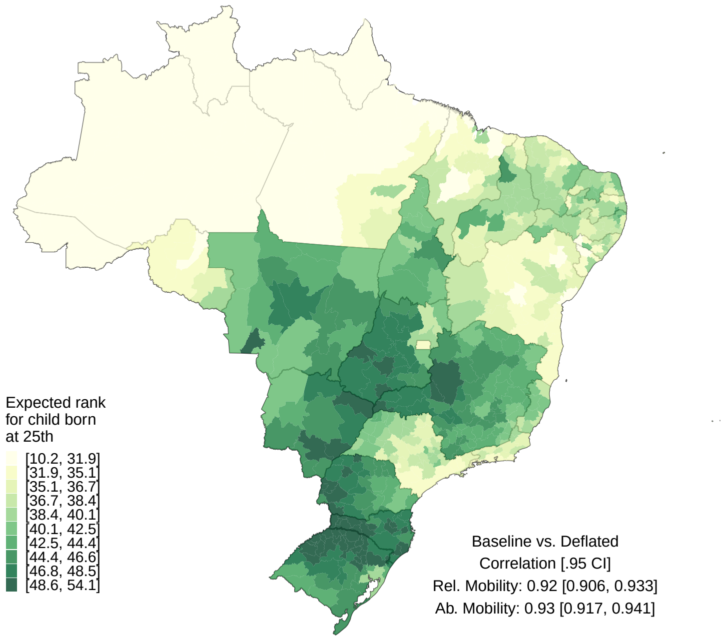


Notes: The figure plots separate mobility curves for Fortaleza (red triangles) and Belo Horizonte (blue dots). Both curves are non-parametric binscatters constructed by plotting mean child income rank for children born in each parental percentile income rank in both regions. The figure is based on our main 1988-1990 cohorts sample. Income is defined as our baseline measure and both children and parents continue to be ranked according to the respective national income distribution. Children are assigned to regions according to the location of their parents in 2000, regardless of where they live in adulthood.

A.3.2 Robustness of subnational estimates

We construct a regional price index to generate mobility maps that account for price differences across Brazil. To create the index, we use the POF (*Pesquisa de Orçamentos Familiares*), a household budget survey conducted by IBGE. POF gathers rich demographic and expenditure data at a fine geographic level. We use the 2003 and 2019 editions to calculate – across Brazilian areas – the average price of the reference basket used to compute the main Brazilian consumer inflation index (IPCA). Specifically, we compute prices at the state level, distinguishing between the (state) capital, the metropolitan areas around the capital, and the countryside. Next, we rescale parents' and children's income by the index computed for 2003 and 2019, respectively, according to their location. Finally, we re-estimate our regional mobility measures based on the price-adjusted income. Figure Figura 34 shows that such adjustment has little impact on regional mobility patterns. It shows that absolute mobility remains similar across space, and that both absolute and relative mobility are strongly correlated with our original, region-specific, mobility measures (see Figure Figura 12).

Figura 34: Price-Adjusted Absolute Mobility Map



Notes: The figure displays price-adjusted absolute mobility – scaled by deciles – in Brazil's 510 immediate geographical regions (IGRs) for our main sample (1988-1990). Parent and child incomes are deflated by regional price indexes constructed with POF survey data and ranked in the national income distribution (measured when children are aged 3-18 and 25-31, respectively). Absolute mobility indicates the expected rank for children in below-median income families, based on Equation (1.1). Darker green tones indicate higher absolute mobility. Children are assigned to IGRs according to the location of their fathers in 2000.

A.3.3 Mobility estimates for the 50 largest metropolitan areas

Tabela 10: Summary of mobility estimates for the 50 largest metropolitan areas

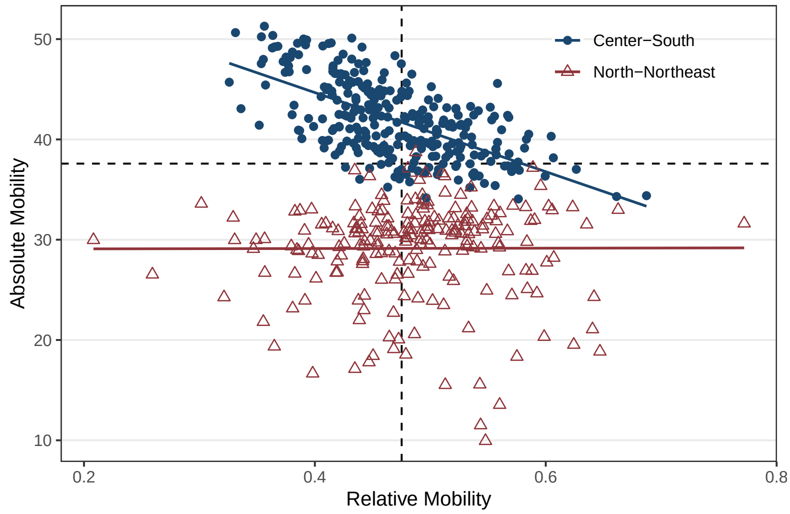
Region	2021 pop. (thousands)	Slope	$E[y\|p=25]$	$E[y\|p=75]$	Q1Q1	Q1Q5
São Paulo, SP	22,049	0.66	34.3	67.4	44.4	1.7
Rio de Janeiro, RJ	12,901	0.55	35.6	63.0	30.1	1.6
Belo Horizonte, MG	5,348	0.50	39.3	64.4	23.2	2.2
Fortaleza, CE	4,179	0.49	33.1	57.8	45.8	2.0
Recife, PE	4,108	0.54	31.8	58.6	48.5	1.7
Salvador, BA	4,065	0.54	29.1	56.3	52.6	1.5
Curitiba, PR	3,732	0.53	41.7	68.3	17.9	2.4
Porto Alegre, RS	3,267	0.50	39.5	64.5	21.8	2.4
Campinas, SP	3,201	0.53	40.4	66.9	33.1	2.3
Distrito Federal, DF	3,094	0.54	36.7	63.6	41.5	2.4
Belém, PA	2,773	0.58	25.1	54.3	63.0	0.8
Goiânia, GO	2,628	0.49	42.4	66.8	24.4	3.8
Manaus, AM	2,605	0.57	24.5	53.0	58.1	1.4
Vitória, ES	2,100	0.48	40.1	64.2	25.4	3.1
Santos, SP	1,927	0.58	34.1	62.9	44.3	2.9
Sorocaba, SP	1,840	0.53	39.3	65.6	32.3	2.8
Natal, RN	1,734	0.49	33.9	58.6	44.9	1.7
São Luís, MA	1,657	0.46	32.9	55.8	49.6	2.1
Ribeirão Preto, SP	1,534	0.60	36.3	66.3	42.8	2.7
João Pessoa, PB	1,430	0.48	33.9	57.9	46.8	2.6
Maceió, AL	1,316	0.52	30.9	56.8	52.7	0.7
Feira de Santana, BA	1,242	0.44	31.1	53.0	57.9	2.4
Aracaju, SE	1,233	0.51	31.9	57.3	49.1	2.1
Florianópolis, SC	1,181	0.42	46.5	67.5	11.2	4.0
Campo Grande, MS	1,131	0.49	41.1	65.8	31.5	3.4
São José dos Campos, SP	1,125	0.54	36.3	63.2	37.3	2.2
Teresina, PI	1,116	0.48	37.1	61.1	43.8	2.8
Londrina, PR	1,114	0.44	43.9	65.7	25.7	2.8

Region	2021 pop. (thousands)	Slope	$E[y\|p = 25]$	$E[y\|p = 75]$	Q1Q1	Q1Q5
Cuiabá, MT	1,105	0.48	40.3	64.2	24.5	2.3
Joinville, SC	1,044	0.46	46.6	69.7	15.7	2.0
Jundiaí, SP	973	0.53	41.2	67.7	37.5	9.1
Uberlândia, MG	959	0.43	43.3	64.7	30.5	4.4
São José do Rio Preto, SP	934	0.53	41.5	67.8	30.9	4.0
Novo Hamburgo - São Leopoldo, RS	908	0.48	42.0	65.9	17.8	2.3
Pelotas, RS	845	0.39	42.1	61.8	28.1	3.3
Caxias do Sul, RS	841	0.42	46.6	67.7	23.3	3.4
Maringá, PR	801	0.43	45.3	66.8	25.7	1.4
Montes Claros, MG	770	0.44	40.1	62.0	33.4	2.7
Juiz de Fora, MG	753	0.44	38.7	60.6	31.7	2.3
Macapá, AP	675	0.50	23.9	49.0	59.1	0.5
Bauru, SP	668	0.54	38.4	65.4	34.8	0.6
Volta Redonda - Barra Mansa, RJ	668	0.51	36.5	61.8	36.4	1.6
Porto Velho, RO	667	0.49	30.0	54.6	42.9	2.4
Campos dos Goytacazes, RJ	661	0.48	36.5	60.5	32.0	2.9
Ipatinga, MG	651	0.41	39.1	59.5	34.0	3.5
Ponta Grossa, PR	648	0.51	41.6	67.1	31.1	2.2
Taubaté - Pindamonhangaba, SP	637	0.48	37.1	61.3	26.1	0.8
Araraquara, SP	631	0.55	38.6	66.0	28.7	2.3
Piracicaba, SP	617	0.56	39.1	66.9	31.4	3.9
Santa Maria, RS	485	0.41	45.2	65.7	28.0	4.5

Notes: The table summarizes mobility estimates in the 50 largest metropolitan areas (IGRs) of Brazil, according to IBGE's population count in 2021. Mobility estimates are the rank-rank slope (relative), the expected income rank of below- and above-median income children (absolute), the bottom-bottom persistence probability (Q1Q1), and the bottom-to-top quintile transition probability (Q1Q5). Mobility measures are based on our baseline sample of the 1988-1990 cohorts. Children are assigned to IGRs based on the location of their parents in 2000.

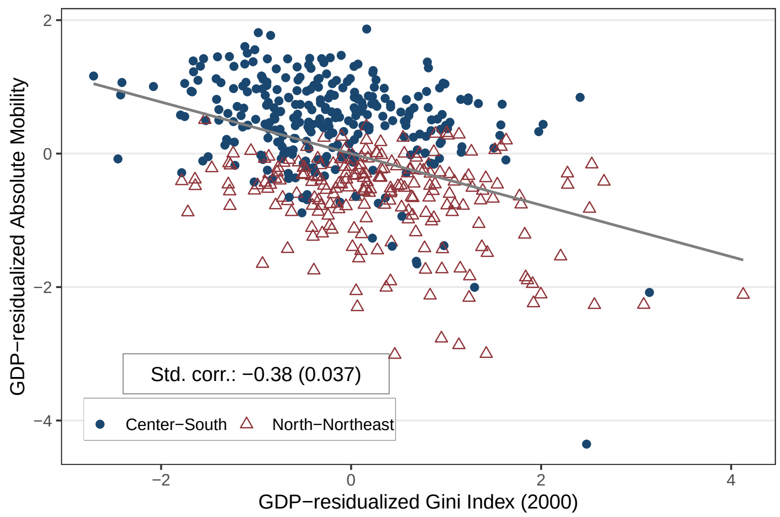
A.3.4 Additional regional mobility graphs

Figura 35: Absolute vs. relative mobility across areas



Notes: The figure is a scatterplot between relative mobility measured by the rank-rank slope (horizontal axis) and absolute mobility (vertical axis) across Brazilian regions. The series in blue (dots) displays regions in the Center-South of the country and the series in red (triangles) displays regions in the North-Northeast. Dashed lines mark the median relative and absolute mobility across all 510 regions.

Figura 36: Brazilian Great Gatsby Curve



Notes: The figure plots the relationship between income inequality measured by the Gini index in 2000 (horizontal axis) and absolute mobility (vertical axis) across Brazilian regions. Both variables are residualized with respect to GDP per capita in 2002. The series in blue (dots) displays regions in the Center-South of the country and the series in red (triangles) displays regions in the North-Northeast. The figure also reports the correlation coefficient between the two (residualized) variables.

A.3.5 Mobility correlates

We explore the correlates of social mobility by estimating univariate regressions of absolute mobility on a wide range of local indicators covering thirteen broad categories: demographics, economic structure, education, family structure, health, household, income, inequality, local infrastructure, labor market, municipal budget, public safety, and social capital. Table Tabela 11 provides a detailed description and data sourcers for the variables in each category. Figure Figura 37 plots the results of these regressions when normalizing both the dependent and explanatory variables so that coefficients can be interpreted as correlations and more easily compared with each other.[48] Overall, coefficients have the expected sign, and nearly all of them are statistically significant. Several variables related to education quality show up among the top mobility predictors – in particular, literacy rates and students' performance in standardized test scores. In line with the analysis by race in Section 1.4.5, the racial composition is also a strong mobility predictor: the share of white population displays the second highest correlation, while the share of black and mixed-race individuals yield negative coefficients. Other variables related to the number of formal firms per capita, the number of bank agencies, and labor market participation by men are also among the strongest mobility predictors. In turn, markers of socioeconomic struggle such as large or high density households, and the share of individuals without earnings are among the top predictors of low mobility, followed by the GDP share of the public sector.

One difficulty when interpreting these results is the strong correlation between the indicators considered. Thus, we reduce the dimensionality of the problem in two steps. First, we create a single index for each category based on the principal components of the initial variables, similarly to (4).[49] Figure Figura 38 report the results of multivariate regressions of absolute mobility on such indexes. Education quality yields the largest correlation with absolute mobility by far, with a positive sign (blue coefficients). Other categories showing strong correlation with mobility are the indexes related to the family structure, demographics (including the racial composition), household characteristics and local infrastructure. Once we control for region fixed effects (5 categories), the education index continues to stand out relative to other factors, being by far the strongest mobility predictor within regions (light blue coefficients).

Nonetheless, the regressors in Figure Figura 38 still have a high degree of multi-

[48]Specifically, we recenter them around the mean and rescale them so that their standard deviation is equal to one.

[49]Specifically, for each group of variables, we keep the number of principal components needed to explain 90% of the variation in them. Subsequently, we compute the index as an average weighted by the amount of variation each component absorbs.

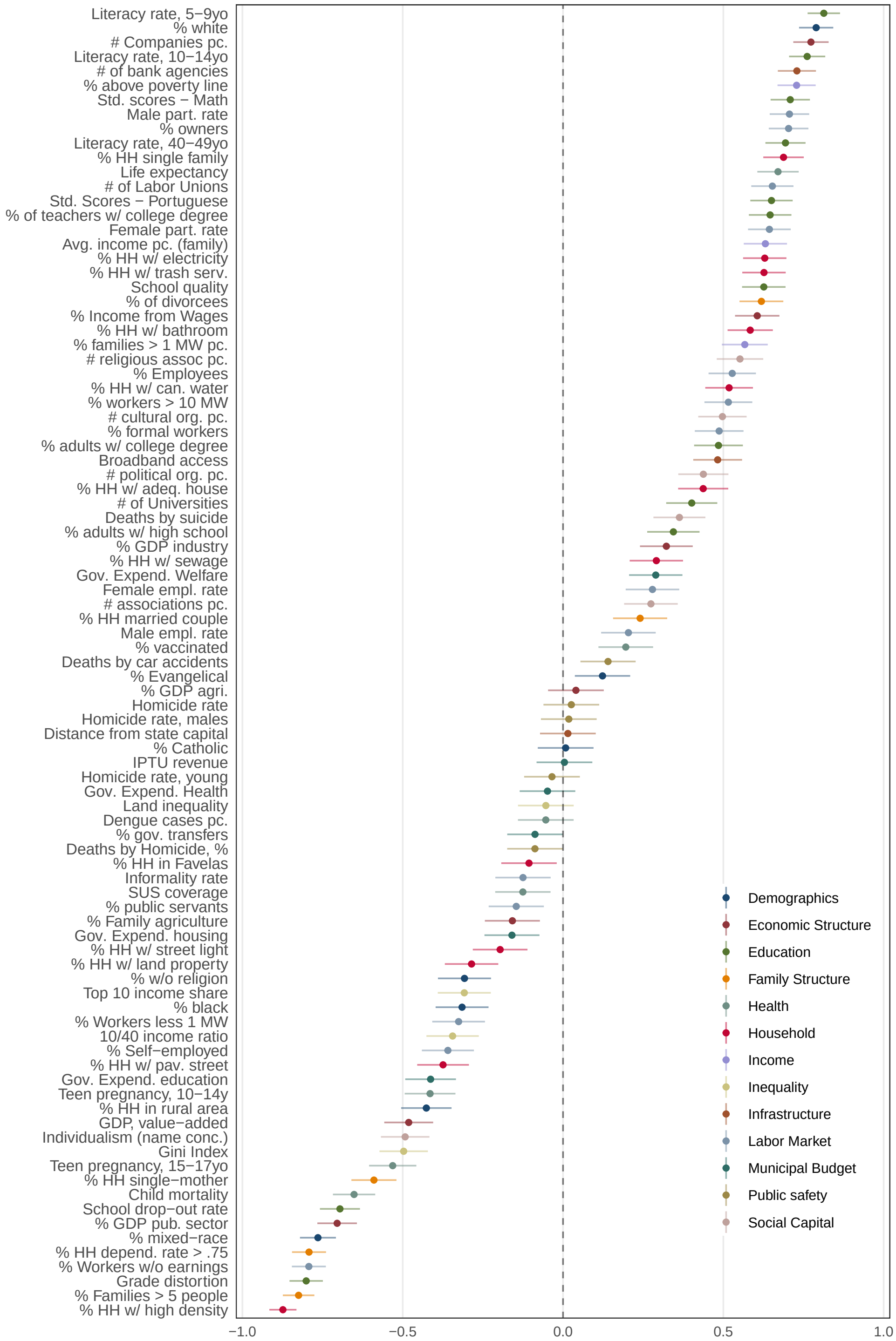
collinearity, which could harm the interpretation of the results. Hence, we employ a standard LASSO regularization procedure to select robust predictors of mobility. We validate the choice of parameters for the LASSO regression via ten-fold cross-validation. Figure Figura 39 summarizes the results of such an exercise, displaying the regressors' coefficients (y-axis) against increasing values of the regularization parameter (x-axis). As we increase the penalization for the number of regressors, coefficients are shrunk toward zero and variables leave the model. Again, the education index dominates the other variables.

Tabela 11: List of municipal socioeconomic indicators and data sources

Group	Indicator	Year	Source
Demographics	% evangelical, % catholic, % w/o religion, % black, % mixed-race, % white, % HH in rural area	2000	IBGE
Economic Structure	% GDP agriculture, % GDP pub. sector, % GDP industry, (value-added), companies pc, % income from wages, % family agriculture	2000	IBGE
Education	Test scores (portuguese and math), % teachers w/ college degree, school quality, drop-out rate, grade distortion, literacy rate (5-9yo, 10-14yo, and 40-49yo), % of adults w/ high school	2000/2005	IBGE/Inep
Family Structure	% families > 5 people, % HH single mother, % HH married couple, % HH dependency rate > .75, % of divorcees	2000	IBGE
Health	Dengue cases pc, teen pregnancy (10-14yo and 15-17yo), child mortality, life expectancy, % of vaccinated, SUS coverage	2000	DataSUS/IBGE
Household	% HH in favelas, % HH single family, % HH w/ land property, % HH w/ trash service, % HH w/ pavimented street, % HH w/ bathroom, % HH w/ piped water, % HH w/ electricity, % HH w/ street light, % HH w/ adequate housing, % HH w/ sewage, % HH w/ high people/room density	2000	IBGE
Income	Average family income pc, % families above the poverty line, % families earning more than 1 MW	2000	IBGE
Inequality	Gini Index, top 10/bottom 40 income ratio, top 10 income share, land inequality	2000	IBGE
Infrastructure	Broadband access, distance from state capital, of bank agencies	2000/2007	ANATEL/IBGE
Labor Market	% of workers ¡ 1 MW, % of workers w/o earnings, % of workers > 10 MW, female/male participation rate, female/male employment rate, % public servants, % firm owners, % self-employed, % formal workers, % employees, informality rate, of labor unions	2000	IBGE
Municipal Budget	Government spending (health, welfare, education, housing), % of federal government transfers, IPTU revenue (property-tax)	2000	FINBRA
Public Safety	Homicide rate (total, males, young), % of deaths by homicide, % of deaths by car accident	2007	Ipeadata
Social Capital	Religious associations pc, cultural organizations pc, political organizations pc, civil associations pc, % of deaths by suicide	2000	IBGE

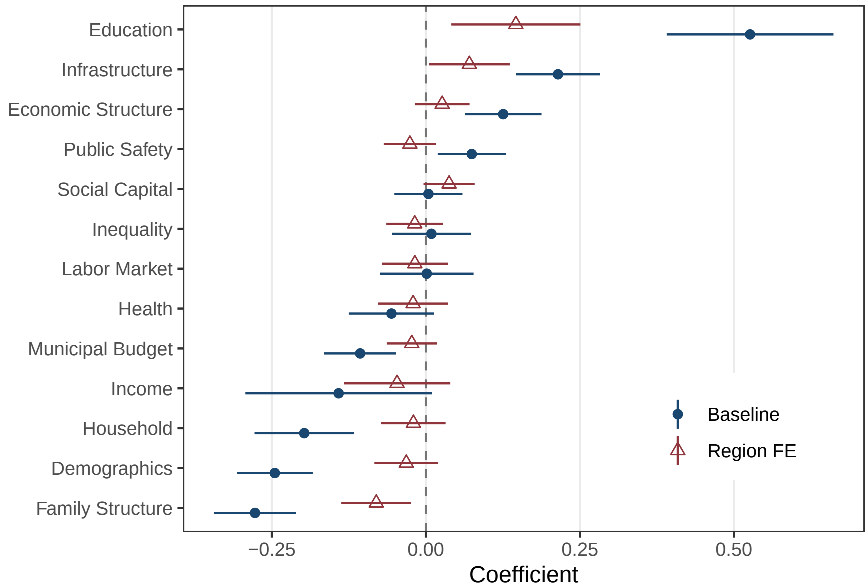
Notes: The table list all indicators used in the mobility correlates analysis, along with their source, year, and category group (used in the principal components analysis). All of them are obtained at the municipal level and then aggregated to immediate region level by population-weighted averages.

Figura 37: Mobility Correlates



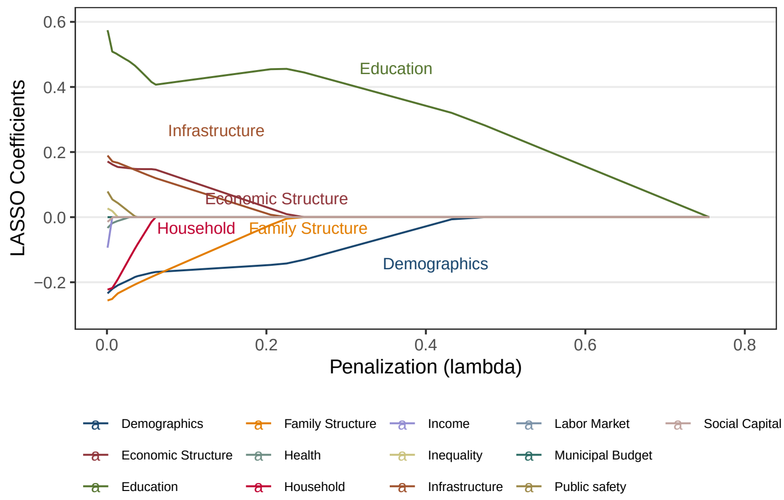
Notes: The figure summarizes a series of cross-regional univariate regressions of absolute mobility on a series of demographic, political, and socioeconomic indicators. The horizontal axis marks coefficients and .95 confidence intervals for each indicator, which are labelled in the vertical axis. Both dependent and independent variables are normalize so coefficients can be interpreted as straightforward correlations. Indicators are colored according to broad categories.

Figura 38: Mobility Correlates: Principal Components



Notes: The figure reports the results of two multivariate regressions between absolute mobility and principal components of regional characteristics. The horizontal axis marks coefficients and .95 confidence intervals for each indicator, which are labelled in the vertical axis. The blue dots includes only the principal components and the red triangles include region fixed-effects (North, Northeast, Center-West, Southeast, and South).

Figura 39: Mobility Correlates: LASSO Regularization



Notes: The figure plots the results of LASSO regularization for the correlation between absolute mobility and indexes constructed from socioeconomic indicators. The horizontal axis plots the regularization parameter lambda, while coefficients of each index are represented in the vertical axis. As lambda grows, the penalization for the number of regressors grow and coefficients are shrunk towards zero.

A.4 Causal place effects

A.4.1 Sample construction

For the analysis on causal place effects, we focus on children born in the 1983-1992 period who we can link to their fathers (following the baseline, conservative method described in Section 2.2). For every father in this sample, we retrieve all regions in which they worked during the 1992-2019 period using (formal) employment data (RAIS). We focus on the latter to track moves since address updates in the Brazilian person registry are largely incomplete before 2000. Although this choice implies that our sample is more representative of formal workers, doing so allows us to track migration more precisely, which is crucial for this analysis.

We define a mover as someone leaving a job in region a and taking a job in a new region b for at least two years. To increase precision, we focus on fathers showing up in employment data for at least five years in the 1993-2019 period.[50] Fathers who never move are defined as permanent residents of their regions. In turn, movers are those who move at least once. To simplify the analysis, we focus on families moving only once. Our final sample comprises 3,172,145 children and 2,260,645 fathers, with around 18% of them being movers.

A.4.2 Defining the predicted outcomes of permanent residents

We closely follow the research design and specifications in (24) and (38). First, we characterize outcomes of permanent residents of each region m and cohort c by running several rank-rank regressions of the type, for each region and cohort:

$$y_{imc} = \alpha_{mc} + \beta_{mc} p_{imc} + \epsilon_{imc} \tag{A.5}$$

where y_{imc} denotes the income percentile rank at the age of 24 of a child from cohort c who spent her entire childhood in region m. We focus on income at the age of 24 that we can measure for all cohorts (1983-1992) in our sample. To ensure precision, we keep only region-cohort pairs for which we have at least 400 observations. We then calculate

[50]Our results remain similar when varying this threshold to ten or fifteen years. We use the five-year cut-off to enlarge the final sample and enhance precision.

the predicted income rank of residents for every parental income rank p, region m, and cohort c: $\hat{y}_{pmc} = \hat{\alpha}_{mc} + \hat{\beta}_{mc} \times p$.

A.4.3 Parametric specification and family fixed effects

Our baseline specification (1.2) includes nearly 180 thousands of fixed effects α_{ocpa}. While they ensure that we exclusively compare very similar children (with the same origin, cohort, parental income decile and age at move) to estimate place effects, they also strongly restrict the variation used in the analysis. Consequently, they leave little space for adding family fixed effects, which also strongly restricts the variation in the analysis. Hence, we follow (20) and rely on a less restrictive, parametric specification to assess the robustness of our findings to family fixed effects:

$$\begin{aligned} y_i = \sum_{a=1}^{33} b_a I_a(a_i = a)\Delta_{odpc} + \sum_{c=1983}^{1991} \kappa_c I_c(c_i = c)\Delta_{odpc} + \\ \sum_{a=1983}^{1992} I_c(c_i = c)(\eta_c^1 + \eta_c^2 \hat{y}_{poc}) + \sum_{a=1}^{33} I_a(a_i = a)(\zeta_a^1 + \zeta_a^2 p_i) + \lambda_f + \epsilon_i \end{aligned} \tag{A.6}$$

Rather than controlling for fixed effects α_{ocpa}, this specification linearly controls for the quality of origin – which is allowed to vary by parental income and cohort – and age at move by parental income, accounting for the disruption effects of moving at different ages. Specifically, the first term in the second line is defined by cohort fixed effects η_c^1 and an interaction between the cohort dummies η_c^2 and the quality of origin $\hat{y}_{poc}$, modeled as the predicted income of permanent residents at origin o. In turn, the second term is defined by age at move fixed effects ζ_a^1 and age at move dummies ζ_a^2 interacted with parental percentile rank, p_i. Finally, the specification controls for family fixed effects λ_f, ensuring that causal place effects solely rely on variation across siblings.

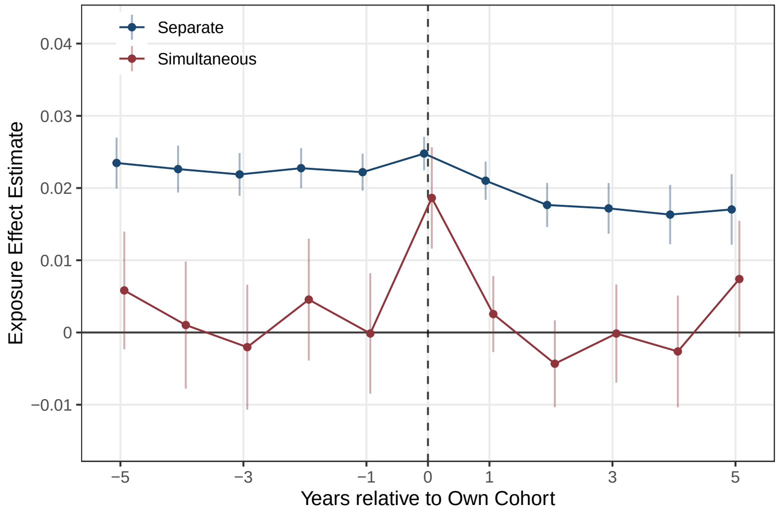
A.4.4 Overidentification tests

We start by showing that place effects are cohort-specific. The blue line in Figure Figura 40 displays the estimates for the exposure rate obtained from eleven separate regressions. In each of them, we replace the main independent variable – the predicted difference in outcomes for children of the *same cohort* – by the predicted difference for *other* cohorts, born from five years before to five years after. The coefficients obtained using adjacent cohorts are quite similar to the baseline, as regions with better opportunities for a given cohort usually are also good for other cohorts. In turn, the red line in Figure Figura 40 plots estimates when all cohort-specific predictions are simultaneously included in a single regression. Conditional on the predicted outcomes of their *own* cohort, all other cohorts' predictions are statistically insignificant, while the true cohort coefficient approaches the baseline estimate. Hence, children's outcomes converge to the outcomes of permanent residents of their *own* cohort and other cohorts' outcomes have little explanatory power. Thus, any omitted variable possibly driving our results would have to precisely emulate cohort-specific place effects.

In Table Tabela 12, we conduct an analogous exercise for gender and race. For this purpose, we construct gender-specific predictions and estimate exposure rates in three different ways: using the predicted outcomes of the child's own gender (column 1), the opposite gender's prediction (column 2), and both together (column 3). Like the blue line in the previous exercise, both regressors yield statistically significant estimates, since there is a considerable correlation between outcomes for boys and girls within regions. Nonetheless, the one based on the child's own gender has higher explanatory power. Column 3 replicates the red line in Figure Figura 40: controlling for the own gender's prediction, the coefficient on the opposite gender (placebo) is negligible. Columns 4-6 conduct the same exercise for race, showing similar results and supporting our main analysis.

All estimates up to now have been based on the predicted differences in *mean* outcomes across locations. Now we show that place effects also replicate permanent residents' outcomes along the income distribution. We construct permanent residents' predictions for the probability of being in the top and bottom deciles of the national income distribution in adulthood. Subsequently, we estimate exposure rates contrasting the distributional predictions with the mean prediction (placebo). Columns 1-2 in Table Tabela 13 show that the top ten probability is better explained by the distributional prediction than by the mean prediction when running separate regressions. In column 3, a simultaneous regression yields a significant coefficient for the distributional prediction, while the coefficient on the placebo is zero. Columns 4-6 replicate 1-3 but for the bottom ten probability, with equivalent results. Thus, the distribution of children's incomes converges to the distribution of incomes in the destination in proportion to exposure time.

Figura 40: Placebo tests: Cohort-specific convergence



Notes: This figure presents estimates of the annual childhood exposure effect on children's income ranks in adulthood using permanent resident predictions for the child's own birth cohort and surrounding "placebo" birth cohorts. The series in blue plots estimates of the exposure effect γ_t from nine separate regressions, using permanent resident predictions from cohort $c+t$ (where t ranges between -5 and 5) as the key independent variables and the outcomes of children in birth cohort c as the dependent variable. The series in red plots estimates from a single multivariate regression that simultaneously includes all nine permanent resident predictions $t = 5, ..., 5$.

Tabela 12: Placebo test: Gender- and race-specific convergence

	Exposure effect γ					
	Gender			Race		
	(1)	(2)	(3)	(4)	(5)	(6)
Own Group	0.025***		0.025***	0.022***		0.022***
	(0.001)		(0.002)	(0.001)		(0.002)
Opposite Group		0.020***	-0.000		0.017***	0.000
		(0.001)	(0.001)		(0.001)	(0.002)
Observations	285,912	285,912	285,912	267,614	267,614	267,614

Notes: The table reports estimates of annual childhood exposure effects γ using gender- (Columns 1-3) and race-specific (Columns 4-6) permanent resident predictions. In all columns, the dependent variable is the child's family income rank at the age of 24. In both panels, column 1 (4) replaces the predicted outcomes based on all permanent residents in the origin and destination with predictions based on the outcomes of children who have the same gender (race) as the child who moves. Column 2 (5) replicates column 1, replacing the own-gender (race) predicted outcomes with the predicted outcomes of the opposite gender (race). Column 3 (6) combines the variables in columns 1 and 2, including both the own-gender (race) and placebo other-gender (race) predictions (*p<0.1; **p<0.05; ***p<0.01).

Tabela 13: Placebo test: Distributional convergence

	Exposure effect γ					
	Upper Tail			Lower Tail		
	(1)	(2)	(3)	(4)	(5)	(6)
Distributional prediction	0.031***		0.034***	0.021***		0.022***
	(0.002)		(0.002)	(0.003)		(0.003)
Mean Rank prediction		0.000***	0.000		0.000***	0.000**
		(0.000)	(0.000)		(0.000)	(0.000)
Observations	285,912	285,912	285,912	285,912	285,912	285,912

Notes: This table reports estimates of annual childhood exposure effects γ for upper- and lower-tail outcomes: being in the top or bottom 10% of the cohort-specific income distribution at the age of 24. Column 1 reports estimates from a regression of an indicator for being in the top 10% on the difference between permanent residents' predicted probabilities of being in the upper tail in the destination vs. the origin. Column 2 replicates column 1 but uses the difference between permanent residents' predicted *mean* ranks on the right-hand side of the regression. Column 3 includes both the (distributional) and the mean rank prediction. Columns 4-6 replicate columns 1-3 using an indicator for being at the bottom 10% at the age of 24 as the outcome. In all columns, the sample comprises all children in the primary analysis sample of one-time movers (*p<0.1; **p<0.05; ***p<0.01).

B APPENDIX TO CHAPTER 2

B.1 Background and Data

B.1.1 Definition of outcome variables

Tabela 14: Description of outcome variable

Variable name	Definition	Source
High School	Complete high school up to 2019	RAIS/CadÚnico
College	Obtain college degree up to 2019	RAIS/CadÚnico
PBF	Enroll in PBF as head of HH up to 2019	PBF payments
Teen Preg.	Become a parent before 18 years old	Person Registry
Childbirth	Health complications in childbirth	SIH-SUS
Hosp.	Any hospitalization	SIH-SUS
Crime	Being criminally prosecuted up to 2019	Kurier
Jail	Being incarcerated up to 2019	DEPEN
Entr.	Owning a firm up to 2019	Firm Registry
Empl.	Being formally employed at age 25	RAIS
Wage	Formal real wage at age 25	RAIS

B.2 Direct Effects

B.2.1 ML Prediction and treatment assignment

In order to identify mothers with high exposure to PBF, we use a gradient tree boosting algorithm (XGBoost). That is the same approach used by (69) to identify which workers are more exposed to minimum-wage increases in the US. The algorithm builds decision trees sequentially, each trained on the residuals of the former. To avoid over-fitting, we choose several regularization parameters by 5-fold cross-validation.

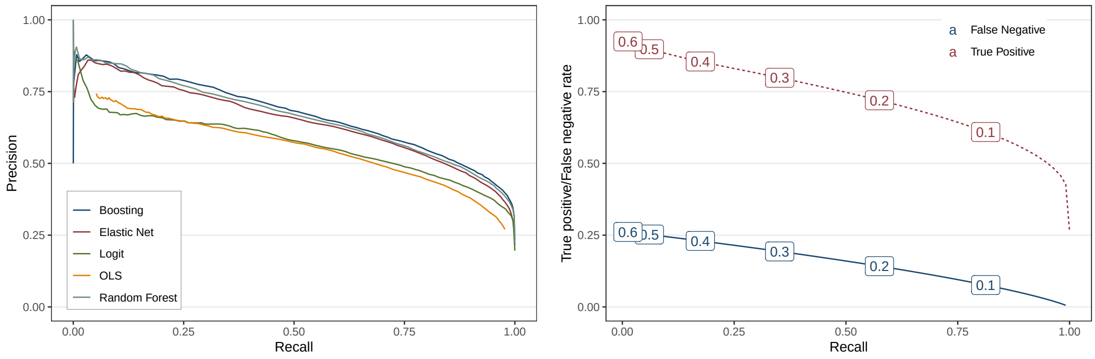
We train the model on a random set of 5% of the nearly 15 million mothers in our full sample. We predict the probability of being enrolled in PBF in 2004 based on city, age, education (4 levels), race (white/non-white dummy), number and age of children, and presence in the formal labor market between 2001-2003.

With the boosting model's predictions at hand, we need to validate it and choose a cut-off to define treatment and control groups. To do so, we plot a precision-recall curve

Figura 41: ML Prediction, Precision-Recall curves

(a) Comparison of Algorithms

(b) Precision-Recall curve, Boosting



for the boosting model and compare it with analogous curves for other algorithms (OLS, Logit, Elastic Net, and Random Forest). The horizontal axis in Figure Figura 41a plots the models' recall, i.e., the share of treated (enrolled in PBF in 2004) mothers included in the treatment group under varying treatment cut-offs. In turn, the vertical axis plots their precision, that is, the share of mothers in the treatment group that were indeed treated. Classification algorithms pose a trade-off between precision and recall, and we wish to have as much as possible of both.

In Figure Figura 41a, it is clear that the boosting algorithm outperforms all other models. Moreover, it is possible to obtain a treatment group with very high precision and still cover a good part of the universe of beneficiaries. In Figure Figura 41b, we plot a modified precision-recall curve for the boosting model where the blue and red lines indicate respectively the share of false negatives and true positives in the treatment group for a given cut-off (indicated in the labels). Based on it, we set our treatment cut-off at 0.4, or nearly the top 11% of predictions. Table Tabela 15 describes the characteristics of mothers in treatment and control groups according to this threshold.

Tabela 15: Descriptive characteristics of mothers in treatment and control groups

	Treatment	Control
Years of schooling	4.85	9.30
Black or Mixed	0.71	0.29
Number of children	3.59	2.03
Age in 2004	36.0	41.6
Age first child	21.6	25.0
North/Northeast	0.74	0.25
PBF 2004 (enrolled)	0.49	0.08
PBF 2004 (value)	43.06	379.92
PBF 2004-2013 (enrolled)	0.91	0.25
PBF 2004-2013 (value)	1,405.79	8,842.29
Avg. cum. child exposure (years)	5.49	1.07
Number of mothers (obs)	848,650	13,359,069
Number of children	3,047,341	27,188,775

B.2.2 Robustness

In our baseline specifications, we control for differential trends across municipalities by including cities' population and average income in 2000 interacted with cohort trends in the vector of controls X_{imt}. We now extend this approach by controlling for more municipal characteristics (Column 2 of Table Tabela 16), specifically the racial composition, share of rural households, share of poor families, informality rate, and child enrollment in school. Moreover, to rule out any possible city-level confounder of PBF's long term effects, in Column 3 we include a full set of city-by-cohort fixed effects.

Tabela 16: IV Estimates of Long-term Effects of PBF, additional controls

	(1)	(2)	(3)
High School	0.0177***	0.0141***	0.0109***
	(0.0020)	(0.0018)	(0.0013)
College	0.0111***	0.0123***	0.0134***
	(0.0007)	(0.0007)	(0.0007)
PBF	-0.0129***	-0.0133***	-0.0151***
	(0.0005)	(0.0006)	(0.0005)
Teen Preg.	-0.0020***	-0.0020***	-0.0020***
	(0.0003)	(0.0003)	(0.0003)
Crime	0.0003	0.0004*	0.0004**
	(0.0002)	(0.0002)	(0.0002)
Incarceration	-0.0012***	-0.0011***	-0.0008***
	(0.0003)	(0.0003)	(0.0002)
Baseline	Yes	Yes	Yes
Additional Controls		Yes	Yes
City-by-Cohort FE			Yes

Notes: The table reports estimates from Equation 2.3. Column 1 replicates the results in Panel A of Table Tabela 2 Coefficients express the effect of 1 additional childhood year spent as PBF beneficiary. (*p<0.1; **p<0.05; ***p<0.01).

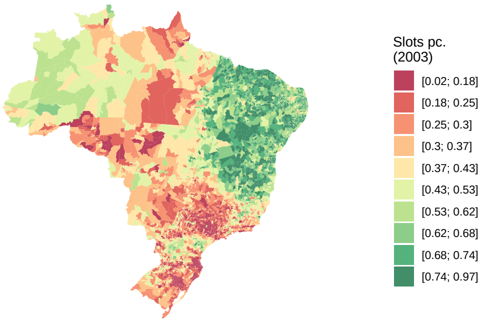
B.3 Spillovers

B.3.1 PBF slots

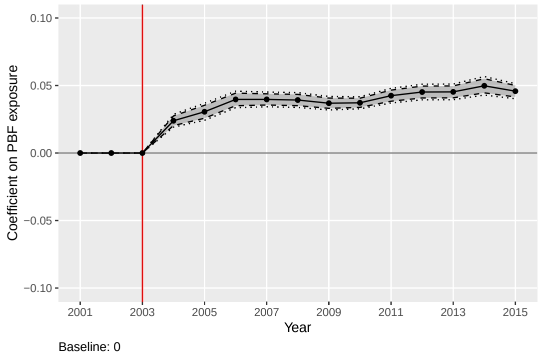
Figure Figura 42a maps PBF slots per capita in 2003 ($Slots_m$) across the country. In Figure Figura 42b we show that this variable is a good predictor of exposure to PBF by regressing the share of population of the city m in year t that is beneficiary of PBF on $Slots_m$ interacted with year indicators with city and year fixed effects.

Figura 42: PBF Slots

(a) PBF slots in 2003



(b) PBF Slots: first-stage



B.3.2 Robustness

In our baseline specifications, we control for differential trends across municipalities by including cities' population and average income in 2000 interacted with cohort trends in the vector of controls X_{imt}. We now extend this approach by controlling for more municipal characteristics (Column 2 of Table Tabela 17), specifically the racial composition, share of rural households, share of poor families, informality rate, and child enrollment in school.

Tabela 17: Long-term Spillover Effects of PBF, additional controls

	(1)	(2)
High School	0.0604***	0.0574***
	(0.0122)	(0.0122)
College	0.0299***	0.0331***
	(0.0085)	(0.0083)
PBF	-0.0252***	-0.0278***
	(0.0065)	(0.0064)
Teen Preg.	0.0014	0.0012
	(0.0038)	(0.0037)
Crime	-0.0104***	-0.0101***
	(0.0030)	(0.0031)
Incarceration	-0.0054**	-0.0053**
	(0.0023)	(0.0023)
Baseline	Yes	Yes
Additional Controls		Yes

Notes: The table reports estimates from Equation 2.4. Column 1 replicates the results in Panel A of Table Tabela 3 Coefficients express the effect of a 100 p.p. increase in municipality-level PBF coverage for low-exposure children. (*p<0.1; **p<0.05; ***p<0.01).

REFERENCES

[1] ALESINA, A. et al. Intergenerational mobility in africa. *Econometrica*, v. 89, n. 1, p. 1–35, 2021. ISSN 1468-0262. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17018>.

[2] ALESINA, A.; STANTCHEVA, S.; TESO, E. Intergenerational mobility and preferences for redistribution. *American Economic Review*, v. 108, n. 2, p. 521–554, 2018. ISSN 0002-8282. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20162015>.

[3] CHETTY, R. et al. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, v. 129, n. 4, p. 1553–1623, 2014. ISSN 0033-5533. Available at: <https://doi.org/10.1093/qje/qju022>.

[4] ACCIARI, P.; POLO, A.; VIOLANTE, G. 'and yet, it moves': Intergenerational mobility in italy. *American Economic Journal: Applied Economics*, 2021.

[5] BRATBERG, E. et al. A comparison of intergenerational mobility curves in germany, norway, sweden, and the US. *The Scandinavian Journal of Economics*, v. 119, n. 1, p. 72–101, 2017. ISSN 1467-9442. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/sjoe.12197>.

[6] CONNOLLY, M.; CORAK, M.; HAECK, C. Intergenerational mobility between and within canada and the united states. *Journal of Labor Economics*, v. 37, p. S595–S641, 2019. ISSN 0734-306X. Available at: <https://www.journals.uchicago.edu/doi/abs/10.1086/703465>.

[7] DEUTSCHER, N.; MAZUMDER, B. Intergenerational mobility across australia and the stability of regional estimates. *Labour Economics*, v. 66, 2020. Available at: <https://ideas.repec.org/a/eee/labeco/v66y2020ics0927537120300658.html>.

[8] HEIDRICH, S. Intergenerational mobility in sweden: a regional perspective. *Journal of Population Economics*, v. 30, n. 4, p. 1241–1280, 2017. ISSN 1432-1475. Available at: <https://doi.org/10.1007/s00148-017-0648-x>.

[9] HELSØ, A.-L. Intergenerational income mobility in denmark and the united states. *The Scandinavian Journal of Economics*, Wiley Online Library, v. 123, n. 2, p. 508–531, 2021.

[10] KRUEGER, A. B. The rise and consequences of inequality in the united states. *Speech at the Center for American Progress*, v. 12, p. 3, 2012.

[11] LEME, L. G. D. S. *Genealogia paulistana*. [S.l.]: Duprat & Comp., 1905.

[12] VASCONCELOS, R. S. *Arquivo Nobiliárquico Brasileiro*. [S.l.: s.n.], 1918.

[13] IBGE. *Pesquisa Nacional por Amostra de Domicílios (PNAD)*. 2019.

[14] ULYSSEA, G. Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, v. 108, n. 8, p. 2015–47, August 2018. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20141745>.

[15] ULYSSEA, G. Informality: Causes and consequences for development. *Annual Review of Economics*, v. 12, n. 1, p. 525–546, 2020. Available at: <https://doi.org/10.1146/annurev-economics-082119-121914>.

[16] HECKMAN, J. J.; LANDERSø, R. *Lessons from Denmark about Inequality and Social Mobility*. 2021. Available at: <https://www.nber.org/papers/w28543>.

[17] CORAK, M. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, v. 27, n. 3, p. 79–102, 2013. ISSN 0895-3309. Available at: <https://www.aeaweb.org/articles?id=10.1257/jep.27.3.79>.

[18] DAVIS, J. M.; MAZUMDER, B. Racial and ethnic differences in the geography of intergenerational mobility. *Unpublished Working Paper*, 2018.

[19] CHETTY, R. et al. Race and economic opportunity in the united states: an intergenerational perspective. *The Quarterly Journal of Economics*, v. 135, n. 2, p. 711–783, 2020. ISSN 0033-5533. Available at: <https://doi.org/10.1093/qje/qjz042>.

[20] CHETTY, R.; HENDREN, N. The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, v. 133, n. 3, p. 1163–1228, 2018. ISSN 0033-5533. Available at: <https://doi.org/10.1093/qje/qjy006>.

[21] BIASI, B. School finance equalization increases intergenerational mobility. *Journal of Labor Economics*, forthcoming, 2023.

[22] DERENONCOURT, E. Can you move to opportunity? evidence from the great migration. *American Economic Review*, v. 112, n. 2, p. 369–408, 2022.

[23] MOUNTJOY, J. *Community colleges and upward mobility*. [S.l.], 2021.

[24] CHETTY, R.; HENDREN, N. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, v. 133, n. 3, p. 1107–1162, 2018. ISSN 0033-5533. Available at: <https://doi.org/10.1093/qje/qjy007>.

[25] SOLON, G. Intergenerational mobility in the labor market. In: *Handbook of labor economics*. [S.l.]: Elsevier, 1999. v. 3, p. 1761–1800.

[26] BLACK, S. E.; DEVEREUX, P. J. et al. Recent developments in intergenerational mobility. *Handbook of Labor Economics*, v. 4, p. 1487–1541, 2011.

[27] BLANDEN, J. Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. *Journal of Economic Surveys*, Wiley Online Library, v. 27, n. 1, p. 38–73, 2013.

[28] BJÖRKLUND, A.; JÄNTTI, M. Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: a comparison of four approaches. *Research in Social Stratification and Mobility*, Elsevier, v. 70, p. 100455, 2020.

[29] ASHER, S.; NOVOSAD, P.; RAFKIN, C. Intergenerational mobility in india: Estimates from new methods and administrative data. 2021.

[30] FERREIRA, S.; VELOSO, F. Mobilidade intergeracional de educação no brasil. *Pesquisa e Planejamento Econômico*, v. 33, n. 3, 2003.

[31] MAHLMEISTER, R. et al. Revisitando a mobilidade intergeracional de educação no brasil. *Insper Policy Paper*, n. 26, 2017.

[32] DUNN, C. The intergenerational transmission of lifetime earnings: Evidence from brazil. *The B.E. Journal of Economic Analysis & Policy*, v. 7, n. 2, p. 1–42, 2007. Available at: <https://ideas.repec.org/a/bpj/bejeap/v7y2007i2n2.html>.

[33] LEONE, T. The geography of intergenerational mobility: Evidence of educational persistence and the "great gatsby curve" in brazil. *Review of Development Economics*, Wiley Online Library, 2018.

[34] MENESES, F. Intergenerational mobility in chile: A year-to-year analysis of a national cohort of students (rr). *Unpublished Working Paper*, 2020.

[35] NARAYAN, A. et al. *Fair progress?: Economic mobility across generations around the world*. [S.l.]: World Bank Publications, 2018.

[36] PORTA, R. L.; SHLEIFER, A. Informality and development. *Journal of Economic Perspectives*, v. 28, n. 3, p. 109–26, 2014.

[37] MEDINA, L.; SCHNEIDER, M. F. *Shadow economies around the world: what did we learn over the last 20 years?* [S.l.]: International Monetary Fund, 2018.

[38] DEUTSCHER, N. Place, peers, and the teenage years: Long-run neighborhood effects in australia. v. 12, n. 2, p. 220–249, 2020. ISSN 1945-7782. Available at: <https://www.aeaweb.org/articles?id=10.1257/app.20180329>.

[39] CHETTY, R.; HENDREN, N.; KATZ, L. F. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, v. 106, n. 4, p. 855–902, 2016.

[40] CHYN, E. Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, v. 108, n. 10, p. 3028–56, 2018.

[41] DAMM, A. P.; DUSTMANN, C. Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, v. 104, n. 6, p. 1806–32, 2014.

[42] CHYN, E.; KATZ, L. F. Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives*, v. 35, n. 4, p. 197–222, 2021.

[43] CARRILLO, B.; CHARRIS, C.; IGLESIAS, W. Moved to poverty? a legacy of the apartheid experiment in south africa. *Unpublished Working Paper*, 2022.

[44] World Bank. *Poverty and Inequality Platform*. 2021.

[45] DAHL, M.; DELEIRE, T. The association between children's earnings and fathers' lifetime earnings: Estimates using administrative data. *Institute for Research on Poverty, University of Wisconsin*, v. 1342, 2008.

[46] BRITTO, D. G.; PINOTTI, P.; SAMPAIO, B. The effect of job loss and unemployment insurance on crime in brazil. *Econometrica*, Wiley Online Library, v. 90, n. 4, p. 1393–1423, 2022.

[47] FERRAZ, C.; FINAN, F.; SZERMAN, D. *Procuring firm growth: the effects of government purchases on firm dynamics*. [S.l.], 2015.

[48] GERARD, F.; GONZAGA, G. Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in brazil. *American Economic Journal: Economic Policy*, 2021.

[49] ATHEY, S.; TIBSHIRANI, J.; WAGER, S. Generalized random forests. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 47, n. 2, p. 1148 – 1178, 2019. Available at: <https://doi.org/10.1214/18-AOS1709>.

[50] FRIEDBERG, R. et al. Local linear forests. *Journal of Computational and Graphical Statistics*, Taylor Francis, v. 30, n. 2, p. 503–517, 2021. Available at: <https://doi.org/10.1080/10618600.2020.1831930>.

[51] BILAL, A.; ROSSI-HANSBERG, E. Location as an asset. *Econometrica*, v. 89, n. 5, p. 2459–2495, 2021.

[52] CARD, D.; ROTHSTEIN, J.; YI, M. *Location, Location, Location*. [S.l.], 2021. Available at: <https://EconPapers.repec.org/RePEc:cen:wpaper:21-32>.

[53] CORAK, M.; HEISZ, A. The intergenerational earnings and income mobility of canadian men: Evidence from longitudinal income tax data. *The Journal of Human Resources*, v. 34, n. 3, p. 504, 1999. ISSN 0022166X. Available at: <https://www.jstor.org/stable/146378?origin=crossref>.

[54] LEE, C.-I.; SOLON, G. Trends in intergenerational income mobility. *The Review of Economics and Statistics*, v. 91, n. 4, p. 766–772, 2009.

[55] NYBOM, M.; STUHLER, J. Biases in standard measures of intergenerational income dependence. *Journal of Human Resources*, v. 52, n. 3, p. 800–825, 2017. Available at: <https://ideas.repec.org/a/uwp/jhriss/v52y2017i3p800-825.html>.

[56] CORAK, M. The canadian geography of intergenerational income mobility. *The Economic Journal*, Oxford University Press, v. 130, n. 631, p. 2134–2174, 2020.

[57] LINDAHL, M. et al. Long-term intergenerational persistence of human capital an empirical analysis of four generations. *Journal of Human Resources*, v. 50, n. 1, p. 1–33, 2015.

[58] BRAUN, S. T.; STUHLER, J. The transmission of inequality across multiple generations: testing recent theories with evidence from germany. *The Economic Journal*, v. 128, n. 609, p. 576–611, 2018.

[59] BOUND, J.; KRUEGER, A. B. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of labor economics*, University of Chicago Press, v. 9, n. 1, p. 1–24, 1991.

[60] BOUND, J. et al. Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, University of Chicago Press, v. 12, n. 3, p. 345–368, 1994.

[61] CHETTY, R. et al. The fading american dream: Trends in absolute income mobility since 1940. *Science*, v. 356, n. 6336, p. 398–406, 2017. ISSN 0036-8075, 1095-9203. Available at: <https://www.sciencemag.org/lookup/doi/10.1126/science.aal4617>.

[62] BERMAN, Y. The long-run evolution of absolute intergenerational mobility. *American Economic Journal: Applied Economics*, v. 14, n. 3, p. 61–83, 2022.

[63] EIKA, L.; MOGSTAD, M.; ZAFAR, B. Educational assortative mating and household income inequality. *Journal of Political Economy*, v. 127, n. 6, p. 2795–2835, 2019. Available at: <https://doi.org/10.1086/702018>.

[64] BRITTO, D. G. et al. Intergenerational mobility in the land of inequality. *Unpublished Working Paper*, 2022.

[65] VIANA, I. *Bolsa Família 15 anos (2003-2018)*. [S.l.: s.n.], 2018.

[66] FISZBEIN, A.; SCHADY, N. *Conditional Cash Transfers : Reducing Present and Future Poverty.* [S.l.: s.n.], 2009.

[67] BAIRD, S. et al. Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: A systematic review. *Campbell Syst. Rev.*, Wiley, v. 9, n. 1, p. 1–124, jan. 2013.

[68] MILLÁN, T. M. et al. Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *J. Dev. Econ.*, Elsevier BV, v. 143, n. 102385, p. 102385, mar. 2020.

[69] CENGIZ, D. et al. Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes. *Journal of Labor Economics*, University of Chicago Press, v. 40, n. S1, p. S203–S247, abr. 2022. Available at: <https://doi.org/10.1086/718497>.

[70] BAILEY, M. et al. Is the social safety net a long-term investment? large-scale evidence from the food stamps program. *Review of Economic Studies*, Forthcoming 2022.

[71] HOYNES, H.; SCHANZENBACH, D. W.; ALMOND, D. Long-run impacts of childhood access to the safety net. *American Economic Review*, v. 106, n. 4, p. 903–34, April 2016. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20130375>.

[72] GOODMAN-BACON, A. The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. *American Economic Review*, v. 111, n. 8, p. 2550–93, August 2021. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.20171671>.

[73] PARKER, S. W.; VOGL, T. S. Do conditional cash transfers improve economic outcomes in the next generation? evidence from mexico. *Working Paper*, September 2021.

[74] ATTANASIO, O. et al. Long term effects of cash transfer programs in colombia. *NBER Working Paper*, 2021.

[75] BOBONIS, G. J.; FINAN, F. Neighborhood peer effects in secondary school enrollment decisions. *Review of Economics and Statistics*, MIT Press - Journals, v. 91, n. 4, p. 695–716, nov. 2009. Available at: <https://doi.org/10.1162/rest.91.4.695>.

[76] GERARD, F.; NARITOMI, J.; SILVA, J. Cash transfers and formal labor markets: Evidence from brazil. *Working Paper*, 2021.

[77] EGGER, D. et al. General equilibrium effects of cash transfers: Experimental evidence from kenya. *Econometrica*, The Econometric Society, v. 90, n. 6, p. 2603–2643, 2022. Available at: <https://doi.org/10.3982/ecta17945>.

[78] BROLLO, F.; MARIA KAUFMANN, K.; LA FERRARA, E. Learning Spillovers in Conditional Welfare Programmes: Evidence from Brazil. *The Economic Journal*, v. 130, n. 628, p. 853–879, 05 2020. ISSN 0013-0133. Available at: <https://doi.org/10.1093/ej/ueaa032>.

[79] ANGELUCCI, M.; GIORGI, G. D. Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, v. 99, n. 1, p. 486–508, March 2009. Available at: <https://www.aeaweb.org/articles?id=10.1257/aer.99.1.486>.

[80] DUFLO, E. The medium run effects of educational expansion: evidence from a large school construction program in indonesia. *Journal of Development Economics*, v. 74, n. 1, p. 163–197, 2004. ISSN 0304-3878. New Research on Education in Developing Economies. Available at: <https://www.sciencedirect.com/science/article/pii/S0304387803001846>.

[81] KHANNA, G. Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *Journal of Political Economics*, Forthcoming 2022.

[82] BAIRD, S.; MCKENZIE, D.; ÖZLER, B. The effects of cash transfers on adult labor market outcomes. *IZA J. dev. migr.*, Walter de Gruyter GmbH, v. 8, n. 1, dez. 2018.

[83] BARR, A.; EGGLESTON, J.; SMITH, A. A. Investing in infants: The lasting effects of cash transfers to new families. *Q. J. Econ.*, Oxford University Press (OUP), v. 137, n. 4, p. 2539–2583, set. 2022.

[84] IMBERT, C.; PAPP, J. Labor market effects of social programs: Evidence from india's employment guarantee. *American Economic Journal: Applied Economics*, v. 7, n. 2, p. 233–63, April 2015. Available at: <https://www.aeaweb.org/articles?id=10.1257/app.20130401>.

[85] COHEN, J. *Statistical power analysis for the behavioral sciences.* [S.l.]: Routledge, 2013.

[86] HAINMUELLER, J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, v. 20, n. 1, p. 25–46, 2012.

[87] JOHANNEMANN, J. et al. Sufficient representations for categorical variables. *Unpublished Working Paper.*, 2019.

[88] HAIDER, S.; SOLON, G. Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, v. 96, n. 4, p. 1308–1320, 2006. Available at: <https://ideas.repec.org/a/aea/aecrev/v96y2006i4p1308-1320.html>.

[89] MELLO, U.; NYBOM, M.; STUHLER, J. A lifecycle estimator of intergenerational earnings mobility. *Unpublished Working Paper*, 2021.