Departament d’Economia Aplicada
Universitat Autònoma de Barcelona

WORKING PAPER 26-03

Intergenerational mobility in Uruguay using income-tax administrative data

Martín Leites
Xavier Ramos
Cecilia Rodríguez
Joan Vilá

Departament d’Economia Aplicada
Edifici B
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain.
d.econ.aplicada@uab.cat
www.uab.cat/en/applied-economics
@ecapuab.bsky.social

Intergenerational mobility in Uruguay using income-tax administrative data*

Martín Leites Xavier Ramos Cecilia Rodríguez Joan Vilá

Abstract

We contribute to the very incipient literature that estimates the intergenerational mobility of income from large-scale administrative data using high-quality income data and provide novel evidence of intergenerational income mobility in a middle-income country, Uruguay. Our estimates address the important role of informal labor markets, one of the features of low- and middle-income countries, and a major challenge to obtain unbiased estimates of intergenerational mobility in these countries. We estimate an IRA of 0.292, indicating that persistence is higher in Uruguay than in high-income countries, but lower than in the US. Our results show that (i) informal income increases intergenerational persistence, (ii) intergenerational persistence is higher at the upper half of the distribution, especially at the richest decile, and (iii) intergenerational income persistence is largest among parents and children of the same sex.

Keywords: Intergenerational income mobility, Informal labor markets, Uruguay, Non-linearities.

JEL Classification: D31 J62 E26

*This research was funded by CSIC I+D 2018 at Universidad de la República. Xavier Ramos and Martín Leites acknowledge financial support from project PID2022-137352NB-C41 (Spanish Ministerio de Ciencia e Innovación). Martín Leites: IECON-UDELAR and EQUALITAS, email: martin.leites@fcea.edu.uy. Xavier Ramos (corresponding author): Universitat Autònoma de Barcelona, IZA, and EQUALITAS, email: xavi.ramos@uab.cat. Cecilia Rodríguez: IECON-UDELAR, email: cecilia.rodriguez@fcea.edu.uy. J. Vilá: IECON-UDELAR, email: joan.vila@fcea.edu.uy.

1 Introduction

The transmission of economic advantage from parents to children is a long-standing concern in the social sciences for reasons of equity and efficiency, which has recently gained social and political importance as a result of the very large inequalities of modern societies and the positive relationship between intergenerational income persistence and cross-sectional income inequality, a relationship dubbed the Great Gatsby curve by Krueger (2012). This transmission and the ensuing persistence is especially relevant in low- and middle-income countries, which typically feature high economic inequality and poverty coupled with economic and political instability that often hampers long-term economic growth (Aghion et al., 1999; Breen and García-Peñalosa, 2005; Bourguignon et al., 2007; Alvaredo and Gasparini, 2015).

The increasing use of income tax administrative and social security data in a handful of high-income countries represents a leap forward in the study of intergenerational mobility, yielding more consistent and credible estimates than those derived from cross-sectional survey data, as they help mitigate the classical problems of measurement error, attenuation, and life-cycle biases (Nybom and Stuhler, 2017; Mitnik et al., 2015; Chetty et al., 2014a; Jantti and Jenkins, 2015). Estimates from administrative records typically reveal higher mobility than previous estimates from self-reported cross-sectional data (e.g. Acciari et al., 2022).[1] Furthermore, the large samples of administrative datasets allow for a more complete description of mobility patterns. Previous studies, for instance, have shown that mobility differs substantially across regions within a country or that income transmission differs by the sex of the child. We also know that persistence is typically higher among the rich (Nybom and Stuhler, 2017).

Many of the mechanisms behind the intergenerational transmission of economic status, such as the education system, the structure and institutions of the labor market, income support policies, or the degree of residential segregation, are very different in low- and middle-income countries relative to high-income countries.[2] Therefore, intergenerational income mobility is also likely to differ between these types of countries. However, due to the unavailability of high-quality administrative data in low- and middle-income countries, we lack mobility esti-

[1] Acciari et al. (2022) contend that this happens because in cross-sectional studies, the instruments used to impute parental income are correlated with child income.

[2] See Stuhler (2018) for a review of intergenerational mobility drivers and Becker et al. (2018) for a theory that unfolds some mechanisms driving intergenerational mobility.

mates comparable to those we have for high-income countries.[3] This study contributes to the very incipient literature that documents intergenerational mobility of income from large-scale administrative data using high-quality income data to provide novel evidence of intergenerational income mobility in a middle-income country, Uruguay.[4]

Informal (labor) markets are prominent in low- and middle income countries.[5] Participation in the informal market is not random,[6] informal labor increases earnings inequality, and is likely to affect intergenerational earnings (and income) mobility. The presence of informality is a major and distinct challenge that studies in poor- and middle-income countries face, as administrative data only provide information on the formal sector of the economy. A distinctive feature of this paper is that it pays particular attention to the implications of including the informal market in the analysis. Informality has two important implications. The first and most obvious one is that informal incomes are missing. We follow Britto et al. (2022), the first study on intergenerational income mobility in a middle-income country, and impute informal earnings using the generalized random forest method (Athey et al., 2019). Not imputing informal incomes leads to a downward bias of 7% in income persistence. The second implication is that the distribution of formal income is not representative as it does not include individuals who work systematically in the informal sector. Because of this, income ranks are mis-estimated and rank-based mobility measures are biased. Previous studies have ignored this issue. We estimate income ranks in a distribution that combines formal and informal incomes. To build this reference distribution, we combine survey and administrative data, and find a downward bias of 8% if we do not include informality in the reference distribution. The downward bias that results from not taking into account the informal sector in the estimation of intergenerational income mobility is sizable for Uruguay. However, they are much smaller than the downward bias

[3]Muñoz and van der Weide (2025) have recently presented a new intergenerational income mobility data base for 87 countries, covering 84% of the world's population. All estimates for low- and middle-income countries are based on self-reported income from survey data and are estimated using two-sample two-stage least squares.

[4]Only Britto et al. (2022) use income tax administrative data to study intergenerational mobility of income in Brazil. Cortés Orihuela et al. (2024) use administrative records from the unemployment insurance program to study intergenerational mobility of earnings in Chile.

[5]In Uruguay, for instance, the informal labor market accounts for roughly one-fourth of total employment. Between 2009 and 2016, formalization increased: in 2009, the formal sector represented 67.8% of all workers and 80.6% of salaried workers, rising to 74.7% and 87.9%, respectively, by 2016 (Amarante and Gómez, 2016).

[6]Informality affects mostly younger workers and women (Leites et al., 2018). Moreover, individuals who enter the informal sector at a young age tend to remain in that sector (Carrasco, 2012), and individuals with weaker attachment to the formal labor market have higher chances to persist in low-paid jobs, as they are in the informal sector or inactive (Carrasco, 2021).

reported in Britto et al. (2022) for Brazil, which is larger than a third. Two issues may explain these large differences: First, Britto et al. (2022) impute formal capital income in addition to informal labor income, and second, the incidence of informal work and the turnover of workers between the formal and informal sectors are much larger in Brazil than in Uruguay.

To estimate intergenerational income mobility, we match large-scale income tax data and social security registers for the period 2009-2016, and combine it with representative household survey data. Our baseline estimation sample comprises nearly 100,000 children born between 1973–1992 and their parents. In line with recent comparable studies for high-income countries, our main mobility measure is the intergenerational ranks association (IRA), i.e. the slope of a rank-rank regression. We estimate an IRA of 0.292 for Uruguay. This implies that it would take about three generations for a child from a family at the 25th percentile to reach the same rank as that of a child from a family at the 75th percentile. This suggests that intergenerational income persistence in Uruguay is larger than in many rich countries (it is 0.17 in Canada, 0.18 in Denmark, 0.20 in Australia, 0.21 in Sweden, 0.22 in Norway, 0.22 in Italy, 0.24 in Germany), similar to other Latin American countries, such as Chile (0.27), and significantly lower than in the US (0.34) and Brazil (0.55) (Bratberg et al., 2017; Mazumder, 2016; Deutscher and Mazumder, 2020; Acciari et al., 2022; Díaz et al., 2021; Britto et al., 2022; Chetty et al., 2014a). Unlike previous studies that find that mobility is constant for the middle 80% of the distribution and is lower at the tails, our results suggest that mobility in Uruguay is lower at the upper half of the distribution than at the lower half, and it is especially low in the richest decile. This non-linear relationship between the ranks of parents and the ranks of children along the income distribution is robust to various estimation methods.

In line with previous findings, our results suggest that capital income contributes to increasing income persistence across generations and that, in line with the life-cycle bias, estimated mobility is higher when the income of the generation of children is measured at an earlier age. We also find that the sex of children and parents matters. We find persistence to be larger for daughters than for sons. This is consistent with the findings of Britto et al. (2022) for Brazil, but at odds with the findings of Chetty et al. (2014a) for the US. A novelty of our study is that we examine persistence by the sex of parents and their children, and find that persistence is higher when parents and children are of the same sex.

Our results are robust to the main weaknesses of our data and to the most important methodological decisions. First, we examine the possible implications of the gender-imbalanced composition of the sample of parents of children aged 30-39. Second, we consider a less restrictive sample selection criteria, which enlarges our sample by including individuals with a weaker attachment to the formal labor market. Third, we address the possible income under-report at the upper tail of the survey income distribution and the possible under-report at the lower tail of the income distribution of the administrative data. Fourth, we enlarge the sample period over which we compute permanent income. Fifth, we use two different indicators of economic status, the logarithm of income and the z-score of income, and thus measure mobility with the intergenerational income elasticity (IGE) and the intergenerational z-score association (IZA). Finally, we report IRA estimates when parental income is the average of the income of both parents for the subsample where we observe both parents.

Our contribution to the literature on intergenerational income mobility is twofold. First, and importantly, we provide precise estimates of intergenerational income mobility for a middle-income country using large-scale and high quality administrative data combined with survey data. Our findings document gender differences, as well as greater persistence both in the upper half of the distribution and when total income is considered rather than earnings. This is an important contribution, as the empirical literature for low- and middle-income countries is very thin. Our second contribution is methodological. We thoroughly address one of the main challenges in the estimation of mobility in poor- and middle-income countries with administrative data, i.e. the presence of informal markets, and show for the first time that not correcting for the direct effect that informality has on the income rank of formal workers biases the intergenerational mobility estimates that are rank-based.

The remainder of the paper is structured as follows. Section 2 describes the three data sets we use in the analysis, the income variable, and examines how the income distribution changes when we take into account the informal sector. Section 3 describes the empirical strategy, and section 4 presents our main results on intergenerational income mobility, showing that mobility differs if parents are poorer or richer, that it also depends on the sex of parents and children, and finally reports the robustness of our findings. Finally, section 5 concludes.

2 Data

To estimate intergenerational mobility from administrative records in Uruguay, we build a novel database that results from matching two main sources of administrative data: (i) a sample of parental linkages (parents/offspring) from social security records and (ii) the universe of income records from Uruguay's tax agency (*Dirección General Impositiva*). Both data sources are matched using a unique identifier especially created for this project.[7] Since the income tax data does not include informal incomes, we use national representative survey data, the National Household Survey (*Encuesta Contínua de Hogares*), to estimate informal labor income and to build an income distribution that includes formal and informal incomes, which we shall refer to as *global distribution*. Since our measure of intergenerational mobility is rank-based, it is important to place individuals in a distribution comprising both formal and informal incomes.

We start this section with a brief description of the data sources we use: the family links database (2.1) and the income tax records (2.2). Then, we discuss our definition of permanent income (2.3) and finally discuss how we include the informal labor market in the analysis (2.4).

2.1 Family links database

The first data set we use, which we call the Family Links database, allows us to know which individuals belong to the same family, and thus to match parents and offspring. This data set includes all individuals who are beneficiaries of one of the policies administered by the Social Security Administration (*Banco de Previsión Social*, BPS), which include the provision of health care to formal workers, the National conditional cash transfer program, and other social benefits for low-income households. As Tables A.1 and A.2 show, we obtain a very representative sample, as the range of policies used to include individuals in the sample is very wide.

Tables A.1 and A.2 report the number of observations of the three data sets used to construct our estimation samples by birth cohort, for the generation of offspring and parents, respectively. Column (2) shows the number of observations of the family links sample, column (3) shows the size of the projected population based on household survey data,[8] and column (6) shows

[7]We had access to an anonymized personal identifier from social security records.

[8]We use the population weights given by the Statistical Office to project the population form the Survey

the number of individuals with positive income in the universe of income tax records. Column (4) illustrates how our sample of family links represents an increasing share of the projected population as cohorts get younger, for the offspring generation. While this share is only 40% for cohorts born before 1980, it increases rapidly for cohorts born in the 1980s, and it is greater than 90% for cohorts born after 1990. The increasing coverage of the national health system could explain part of this growing trend in the representativeness of our sample. Column (7) indicates that the fraction of the universe of tax records that can be matched to offspring who belong to the family link sample and who have positive formal income also increases as cohorts get younger. The last four columns of the Table show that the share of women is similar —about half in most birth cohorts— across datasets and that the share of women with positive income is stable across cohorts —but for the younger cohorts, that include women younger than 20 years old.

Table A.2 presents the same information for the generation of parents. Again, the fraction of the universe of tax records that the sample of family linkages represents is larger for recent cohorts, being less than 50% for the generations prior to 1950, but approaching 80% for the cohorts born after 1960. In addition to the reduced number of observations, the 1940s cohorts show that women are overrepresented in the sample of family links relative to both the population projected on the base of household surveys and the universe of income tax records. The lower representativeness and the higher proportion of mothers represent a potential threat for the estimates that include these cohorts in the estimation sample. However, since we define permanent income for only one of the two parents, the presence of assortative mating in terms of earnings and educational achievement in Uruguay mitigates these potential concerns (Online Appendix Tables A.3 and A.4 summarize the association of years of education and earnings for a representative sample of couples). To further minimize those concerns, Section 4.4.1 reports a set of exercises that examine the sensitivity of our findings to the greater presence of mothers than fathers in our sample.

sample.

2.2 Income tax records

Information about individuals' incomes is obtained from income tax data for the period 2009-2016. This data set includes income from the main formal sources for the entire population —i.e. earnings, capital income, and pensions—, and also information on socio-economic characteristics of the individuals and of the firms where they worked. These same tax micro-data have been previously used to estimate top income shares in Uruguay (Burdín et al., 2022). As the latter paper shows, the data set includes about 70% of the adult population aged 20 and older, each year, and allows accurate estimation of formal income, particularly at the upper tail of the distribution. It is worth noting that despite the potential evasion and elusion problems present in all administrative records, our formal income measure captures a larger proportion of total income for top income groups than any other dataset available for Uruguay. Compared to other data sources, administrative tax records particularly capture the different sources of capital income for the top income groups.[9]

To link the income of parents and their children, we merge the tax records with our sample of family links by means of a unique identifier. As noted above, the sample of parents/offspring links comes from a register of social security benefits, which includes more than one member of the household. This allows us to recover family ties with fathers, mothers, or both. One potential concern is that these policies target certain groups of the population, which could result in a selected sample. To examine whether our sample is selected in terms of formal incomes, we compare it with the universe of tax records.

Table 1 shows descriptive statistics of our sample of matched parents and children, the universe of tax records, and the household survey for the generation of children aged 30 to 39 years in 2012.[10] Table A.5 in the appendix presents the same descriptive statistics for children aged 20 to 29 years.[11] Our sample of matched parents/children sample is over one third of the universe of tax records for children aged 30-39 in 2012. Panels A and B of Table 1 show that important features of the distribution of income of the matched parents/children sample are

[9] Unlike previous studies, our estimations only include the set of incomes reported either directly by each individual or by third-party report, and do not make any imputation of non-nominative sources of income.

[10] The main conclusions of this analysis are robust to choosing other sample years.

[11] We split the sample of offspring into two age groups: 20-29 and 30-39 years old. The sample of individuals aged 20 to 29 years represents about 75% of the same group in the universe of tax records, while this share falls to less than 40% for the 30-39 age group.

consistent with the distribution of income of the universe of tax records. Namely, average labor income, the main component of our variable of interest, the fraction of people with no labor income, average age, and the gender composition are very similar between the two distributions. All these features are also very similar in the distributions of labor income that come from the matched parents/children sample and from the household survey, but for the fraction of people without labor income, which is larger in the latter distribution.

Appendix Table A.6 shows the features of the distribution of labor earnings that come from the three datasets for the generation of parents. In this case, our sample of matched parents/children pairs includes over half the individuals aged between 45 and 65 years from the universe of tax records. The share of women in our sample of matched pairs is larger than that in the universe of tax records (54% vs 49%), and this overrepresentation of mothers is explained almost exclusively by the older cohorts (born before 1950). Average labor income is again slightly lower in our sample of matched pairs, but the difference is to be expected, as women are overrepresented in the sample of matched pairs compared to the universe of tax records. As expected, the average earnings are lower and the share of individuals without labor income is larger in the household survey-based projection than in the other two data sets.

Table 1: Summary Statistics of Three Samples by Age Group: Offspring Aged 30–39, 2012

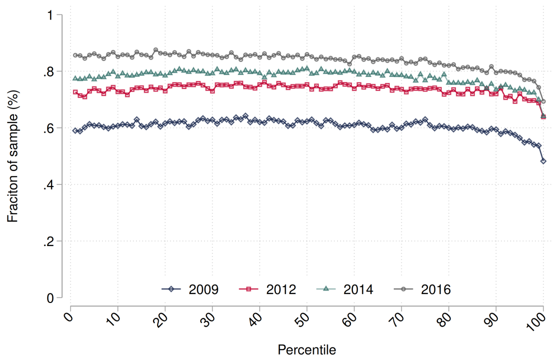
	Distribution			Mean test vs. Panel A	
	Mean	Median	SD	Diff.	F (*p*-value)
Panel A: Parents/children sample					
Labour income	254,852	187,886	284,235	–	–
Fraction no labour income	7.7%	–	–	–	–
Fraction female	48.5%	–	–	–	–
Age	34.2	34.0	2.9	–	–
Observations		125,770		–	–
Panel B: Tax records					
Labour income	253,759	181,994	301,172	1,093	1.23 (0.267)
Fraction no labour income	6.9%	–	–	0.8	169.93 (0.000)
Fraction female	46.7%	–	–	-1.8	265.79 (0.000)
Age	34.4	34.0	2.8	-0.2	11349.87 (0.000)
Observations		333,974			
Panel C: Household survey					
Labour income	265,567	214,527	261,588	-10,745	14.83 (0.000)
Fraction no labour income	16.78%	–	–	-9.08	541.82 (0.000)
Fraction female	47.2%	–	–	1.3	9.25 (0.002)
Age	34.5	35.0	2.8	-0.32	116.33 (0.000)
Observations		10,428			
Weighted observations		307,400			

Notes: The table reports descriptive statistics for individuals aged 30–39 in 2012 using three alternative data sources: the matched parents/children sample (Panel A), tax records (Panel B), and the household survey (Panel C). Differences and F-tests compare Panels B and C with Panel A. The household survey sample includes only individuals with positive total income in 2012. *Source:* Based on social security records (BPS), tax records (DGI), and household survey data (INE).

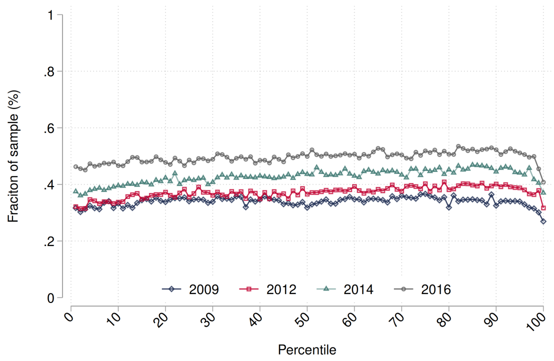
To explore how representative our sample is across the distribution of formal income, Figure 1 shows the share of individuals from the universe of tax records that is included in the sample of parents/children pairs by percentile of the distribution of formal labor income derived from tax records. Panels (a) and (b) show the generation of children aged 20-29 and 30-39 years, respectively, while panel (c) shows the generation of parents.

The three samples displayed in panels (a) to (c) are distributed fairly homogeneously across the percentiles of the formal income distribution. This means that the rank of individuals is very similar in both formal income distributions, of matched pairs and of the universe of tax records. All shares are usually larger for more recent years (e.g. 2016) than for more distant years (e.g. 2009). The share of the sample only falls at the very top income percentiles of the two offspring samples. In sum, the three panels in Figure 1 indicate that all percentiles of the formal income distribution are adequately represented.

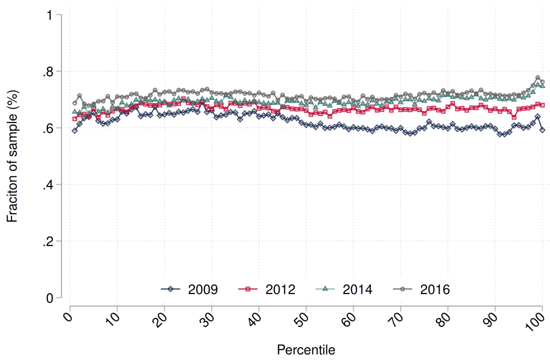
Figure 1: Fraction of individuals from the universe of tax records included in the sample of parents/offsrping links, by percentile of the distribution of labor income and year



(a) Offspring aged 20-29



(b) Offspring aged 30-39



(c) Parents

Notes: The figure shows the percentage of observations included in the sample of matched parents/children pairs by percentile of the distribution of labor income derived from tax records. Panels (a) and (b) show this fraction for the generation of offspring, by age group, and panel (c) show if for the generation of parents.
Source: Social security records (BPS) and tax records (DGI).

2.3 Definition of income variables

Our main income variable is total income, measured as the sum of four income sources: formal labor income, informal labor income —which we will impute—, capital income, and pensions. As it is common in the literature, we will also examine mobility of earnings, which include formal wages and self-employed income. Both concepts are measured before taxes and only incorporate taxable income, which excludes, for example, income from owner-occupied housing and non-contributory public transfers.[12] When we observe both parents, we use the maximum

[12] Since we use income ranks in our main analysis, using pre- or post-tax income should yield very similar results.

income of both parents. Since previous studies typically use the average income of both parents, in Section 4.4.6, we provide evidence suggesting that the main results would not substantially change if we employed average parental income.[13]

To avoid temporary income fluctuations, as it is customary in previous studies, we average yearly incomes over 5 consecutive years, including zero incomes.[14] The time periods used for parents and children are 2009-2013 and 2011-2016, respectively. This is our baseline permanent income definition.

In addition to the potential biases outlined above, which are generated by measurement error, one of the main concerns in the mobility literature is the presence of life-cycle bias due to the observation of incomes at early ages (Haider and Solon, 2006; Nybom and Stuhler, 2017). Previous studies show that intergenerational mobility estimates tend to stabilize when offspring are about 30 years old (Nybom and Stuhler, 2017; Chetty et al., 2014a,b; Böhlmark and Lindquist, 2006). Thus, to minimize life-cycle bias in our estimates, we employ income records of offspring aged 30 to 39 and when possible take averages over the sample period when offspring are closer to 30 years old to compute permanent income. Parents of these children are aged 45 to 65. We also use weights to replicate the age structure of the universe of tax records in our sample.

Previous work also indicates that rank-rank estimates tend to be less sensitive to life-cycle bias, with less attenuation bias and more stability over age (Nybom and Stuhler, 2017; Chetty et al., 2014a; Mitnik et al., 2015). Based on these recommendations, we use the percentile rank as our preferred measure of permanent income.[15] The income distribution from which we estimate the ranks is thus of great importance. The literature typically uses the sample of matched parents and children that stems directly from the administrative database as a reference distribution to compute percentiles. We term his reference distribution based solely on formal income *sample distribution*.

[13]While the Family Links data ensures a match between the child and at least one of the parents, information for both parents is not always available. As mentioned above, for older generations of children, mothers are overrepresented in the sample of matched parent-child pairs. In Section 4.4.1, we thoroughly examine whether this overrepresentation of mothers introduces biases in the mobility estimates.

[14]We show in Section 4.1 that results are robust to excluding these years with zero income when computing average income.

[15]Percentile ranks assign the same rank to incomes that fall in the same income percentile. Our main results are robust to computing ranks as the relative position of the individual divided by the total number of individuals. Results are available upon request.

2.4 Including the informal labor market into the analysis

Informality accounted for roughly one quarter of total employment in Uruguay during the period under analysis (Amarante and Gómez, 2016; Instituto Nacional de Estadística, 2026). Thus, omitting informal earnings is likely to yield biased estimates of intergenerational mobility. To address this source of bias, we estimate and impute informal earnings to the individuals in our sample. Our preferred definition of income includes informal labor income estimated with the Generalised Random Forest method put forth by Athey et al. (2019), which endogenously splits the space of covariates to generate predictions for informal earnings.[16] We grow separate Random Forests to predict informal earnings in each year separately for parents and children using data from the National Household Survey as training datasets. The set of predictors includes dummies for sex, age, education (3), region (19), industry (8), and formal earnings deciles (10).[17] To explore the bias that we may introduce when more parsimonious methods that require less information are used, in Section 4.4 we report IRA estimates that use different ways of estimating informal incomes.

When the correlation of income ranks (instead of the correlation of log-incomes) is used as a measure of mobility, an additional source of bias comes from misestimating the income rank of individuals. This occurs when the distribution of income we use to compute the ranks, which we shall call the reference distribution, is not representative of the population. Previous studies compute income ranks in the sample income distribution. However, in the presence of a significant informal sector, the income ranks of the individuals with formal earnings that are observed in the administrative tax records are bound to be biased. In line with the methods used in the top income studies (Atkinson, 2007), to address this issue, we add individuals with informal labor incomes from the household survey dataset (ECH) to the sample income distribution of formal labor incomes that come from tax records.[18] This procedure increases the reference income distribution by about 25% with individuals who have no formal labor

[16]The IRA estimate remains virtually unchanged when we use OLS regressions instead of Random Forest methods to impute informal labor income –see Section 4.1.

[17]The National Household Survey informs about formal and informal labor income separately. Previous studies have shown that informal labor income is measured accurately (Amarante et al., 2016; Burdín et al., 2022). We are taking due account of the probability of informality as the dependent variable in our prediction exercise includes zero as well as positive informal labor incomes.

[18]An additional concern of using the parent/offspring sample itself for the analysis could be that our sample of family linkages were not representative of the population in the formal sector. However, as we show in Section 4.4.2, this is not the case.

income but do have labor income earned in the informal market. Note that these additional observations are added with the only purpose of estimating the income rank of the individuals observed in the administrative tax records. They do not directly contribute to the estimation of the IRA as we cannot match them with their relevant relatives.

The distribution that results from this combined (tax records and household survey) data is our best approximation to the income distribution in Uruguay for each generation and year. Henceforth, we refer to this distribution as the *global* reference distribution. It is important to emphasize that the *global* reference distribution cannot be recovered without using household survey data. The most comprehensive distribution that can be obtained if we solely rely on tax records data excludes individuals with no formal earnings who are not observed in the eight-year sample period. As a result of this exclusion, the income ranks of the matched observations are estimated with bias, which, in turn, biases the IRA estimates. We find that the estimated IRA is downward biased[19] when income ranks are computed from the largest and most comprehensive sample we can build using only the administrative records.[20]

As noted above, this is the first intergenerational mobility study based on administrative records that estimates income ranks from the global reference distribution instead of using the sample reference distribution. The advantage of using the income ranks from the global reference distribution is that the estimated ranks do not change when we examine different samples of the distribution, as we do in Figure 3 or in Section 4.4.2.

To estimate the income ranks that we use in the intergenerational mobility estimations, we order incomes in the combined global reference distribution from poorest to richest and compute the relative position of individuals with formal labor income in the reference income distribution. Hence, the relative position of the individuals with formal income in the distribution of their generation also depends on the percentage of informal workers and their incomes.

Panel (a) of Figure 3 shows how the observations in our sample are distributed across the percentiles of the global reference distribution (in blue) and the percentiles of the sample

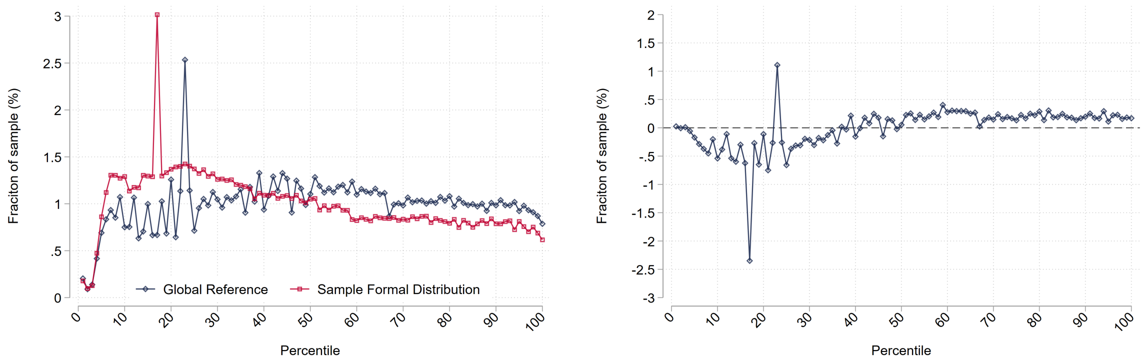
[19] As we report in Table 3, IRA is 0.246 if income ranks are computed directly from the sample distribution (i.e. we use the sample as a reference distribution) instead of 0.263, which is obtained when ranks are computed from the global reference distribution. The difference between the two IRA estimates is statistically significant (p-value $= 0.001$).

[20] This sample includes individuals with no formal earnings over the sample period who are matched with a parent/offspring who reports positive formal labor income for at least one year, in addition to individuals reporting positive formal earnings.

reference distribution of formal income (in red). Note that if the observations in our sample were uniformly distributed throughout the global distribution or the distribution of formal income, they should represent 1% of each percentile. Our sample is overrepresented in the bottom half of the income distribution of formal income, most notably in the first three deciles. However, our sample observations are only overrepresented between the percentile 40 to the percentile 65 of the global distribution because the inclusion of the informal market in the global distribution shifts upward the observations of our sample. In other words, the rank of a given observation is likely to be higher in the *global* reference distribution than in the *reference* distribution of formal income. Panel (b) displays this shift in mass from the bottom half to the upper half of the distribution.

Appendix Figures A.1 and A.2 show this same exercise for the generation of offspring aged 20-29 and for the generation of parents. The pattern followed by offspring aged 20 to 29 and by the generation of parents is similar to that shown in Figure 3.

Figure 2: Fraction of individuals in each percentile of formal income distribution and of combined formal and informal income distribution, children aged 30-39



(a) Formal income distribution

(b) Difference

Notes: Panel (a) shows how our observations are distributed across the *global* reference distribution (blue) and the *sample* reference distribution (red). Panel (b) plots the difference between these distributions. The *global* reference distribution combines formal labor income from tax records and informal labor income from survey data, while the *sample* reference distribution includes only formal labor income from tax records. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

3 Empirical Strategy

The study of intergenerational income mobility seeks to obtain a characterization of the joint distribution of children's and parents' incomes, $f(y^p, y^{ch})$, where y^p and y^{ch} are the income vectors of parents and children, respectively. This distribution can decompose into two components: (i) the joint distribution of parents and children ranks, formally known as the copula of the distribution, and (ii) the marginal income distributions of parents and children, $f(y^p)$ and $f(y^{ch})$. While the marginal distributions determine the level of income inequality of each generation, the copula is a key determinant of mobility across generations, as it is the transformation that links the marginal distributions of both generations (Jantti and Jenkins, 2015).

Our baseline model estimates the average intergenerational ranking association (IRA).

$$P_i^{ch} = \beta' P_i^p + f(\gamma,\, age^{ch}, age^p) + v_i \tag{1}$$

where P_i^{ch} and P_i^p identify the relative position of children and parents of family i in the income distribution of their respective generations. Previous studies since Dahl and DeLeire (2008) argue that income ranks within each generation provide a good proxy of the long-run economic status of both parents and their offspring.

We characterize mobility with the slope (β') of this rank-rank relationship, which provides an average measure of the strength of the association in the copula of the joint distribution (Mitnik et al., 2015). The assumption of linearity ensures that (β') is both locally and globally informative. According to Chetty et al. (2014b), β' can be interpreted as the average difference in the mean percentile/rank of children from the richest families vs. children from the poorest families.

In contrast to the widely used intergenerational income elasticity (IGE) estimates from log-log income regressions, the size of the IRA is insensitive to changes in the income inequality within each generation.[21] In addition, IRA estimates avoid or minimize some empirical problems that IGE estimates face. First, attenuation bias and life-cycle bias are considerably weaker in rank-based measures than in IGE estimates (Nybom and Stuhler, 2017). Second, IRA estimates

[21]The intergenerational elasticity corresponds to the coefficient β from an OLS regression of children's log income on parents' log income. This coefficient can be decomposed into two components: the correlation between parents' and children's incomes and the ratio of the standard deviations of the income distributions across generations. When the regression is estimated using income ranks instead of income levels, and ranks are defined relative to each generation's own sample distribution, the rank distributions of parents and children are identical by construction. In that case, the coefficient reduces to the rank correlation and is therefore independent of differences in marginal income distributions across generations. By contrast, when ranks are computed using a common global reference distribution, the dispersion of ranks may differ between parents and children. As a result, the coefficient also reflects differences in the standard deviations of the two rank distributions.

are comparatively more stable, i.e less sensitive to the samples (and the presence of outliers in the tails) and to specification choices (e.g. the way in which earnings/income are defined and to the treatment of zero incomes in particular, Dahl and DeLeire (2008); Nybom and Stuhler (2017); Chetty et al. (2014b); Mazumder (2005). Third, as Nybom and Stuhler (2017) noted, classical measurement error attenuates log-linear measures through its effect on the variance of observed incomes, but the variances of observed and true ranks are equal by definition. However, a drawback of IRA is that the error term v_i follows a non-classical error distribution as top (bottom) ranks cannot be overstated (understated). Furthermore, since IRA looks at rank movements across generations, similar IRA estimates could have different welfare implications as income distances between ranks may differ. This weakness could be problematic when making comparisons across countries with different distributions of income.

Following Björklund et al. (2012), to assess the IRA at different points of the distribution, we extend equation (1) using non-linear regressions by means of a spline function with pre-defined knots, which identify the position in the distribution of parental incomes at which the slope is allowed to change. Following previous work, we employ six knots. Namely, the four knots standard in the literature (P25, P50, P75, and P90), a knot to examine persistence at the very top 1% of the distribution (P99), and another one at the bottom decile (P10), where the incidence of informal work is highest.

$$P_i^{ch} = \beta' P_i^p + \delta_r \sum_{r=10}^{r=99} (P_i^p - P_r) + u_i \tag{2}$$

As Hertz (2009) noted, the interpretation of the coefficients δ_r in equation (2) is different than the average IRA coefficient from equation (1). In this case, the comparison of local slopes alone does not provide information about the intergenerational persistence or about the presence of differences in expected incomes of offspring from poor, middle, or rich households. For instance, we can not conclude anything about the differences in intergenerational persistence between both groups when the local slope is steeper for one of the groups. The coefficients provide information about the local relationship between offspring's and parents' ranks. For instance, it allows assessing whether the transmission of a given increase in the permanent income of parents to the expected permanent income of their offspring is equally large for rich and for poor parents. Thus, the heterogeneity in slopes helps both unpack the average IRA and also provides local marginal effects of parental permanent income for parents at different percentile groups.

The splines model imposes continuity in the relationship between parents' and children's income. To relax the continuity assumption, we use a set of linear specifications that allows us to change the slope and the intercept for each of the income fractiles defined by the knots we use in the spline regression. Finally, we also use local polynomial regressions to explore non-linearities (Cleveland et al., 1988). Our model can be written as:

$$E(P_i^{ch}|P_i^p) = F(P_i^p) \tag{3}$$

where F is the smoothing function that determines the expected rank of the offspring conditional on parental rank and E is the expectations operator.

4 Intergenerational income mobility in Uruguay

This section discusses our main findings. Section 4.1 presents the baseline average IRA estimate and shows the contribution of the informal sector to intergenerational mobility. Section 4.2 discusses how sensitive the baseline estimate is to alternative methods of imputing informal labor income, while Section 4.3 documents the importance of choosing the proper reference distribution to compute income ranks. Section 4.4 presents six exercises to show that the baseline IRA estimate is robust. The last two Sections, 4.5 and 4.6, explore heterogeneity of the intergenerational mobility by gender of children and parents, and across the income distribution.

4.1 Average mobility

Table 2 shows estimates of average intergenerational income mobility between parents and children aged 30-39, and children aged 20-29. We report estimates for total income and also for labor income to gauge the importance of capital income in shaping mobility of total income. All regressions include controls for the sex of children and the age and sex of parents and use weights to take into account the age composition of our samples.[22]

Table 2: Average IRA for Total and Labor Income: Children Aged 30–39 and 20–29

	Total income		Labor income	
	Children 30–39	Children 20–29	Children 30–39	Children 20–29
IRA	0.292***	0.140***	0.261***	0.134***
	(0.003)	(0.002)	(0.003)	(0.002)
Observations	98,977	226,258	98,977	226,258

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. All reported IRA estimates are statistically significant ($p < 0.001$).

The average Intergenerational Rank Association (IRA) estimate (from equation (1)) of total income for children aged 30-39, our preferred age group, is 0.29. According to these estimates, it would take about three generations for a child from a family at the 25th percentile to reach the same rank as that of a child from a family at the 75th percentile.[23] Although comparisons across countries must be interpreted with caution, this suggests that intergenerational income persistence in Uruguay is larger than in many rich countries (it is 0.17 in Canada, 0.18 in Denmark, 0.20 in Australia, 0.21 in Sweden, 0.22 in Norway, 0.22 in Italy, 0.24 in Germany), similar to other Latin American countries, such as Chile (0.27),[24] and significantly lower than in the US (0.34) and Brazil (0.55) (Bratberg et al., 2017; Mazumder, 2016; Deutscher and Mazumder, 2020; Acciari et al., 2022; Cortés Orihuela et al., 2024; Britto et al., 2022; Chetty et al., 2014a).

How do results change when we look at children aged 20-29? In line with previous findings for high-income countries, persistence is significantly lower (about half) for this younger group of children than for children aged 30-39 (see columns (2) and (4)). This reflects the life-cycle

[22] As we noted in Section 2, younger offspring are over-represented. We weight each individual by the inverse of the number of individuals in the same age range.

[23] Assuming that the permanent income of a family follows an AR(1) process, the number of generations that it takes for two children from parents Δ percentiles apart is the value N that solves the equation $\beta^N \Delta = 1$. In our case, $N = 3.18$. However, as Acciari et al. (2022) note, the correlation between outcomes of three generations found in previous studies for Germany and Sweden is higher that what would obtain under the AR(1) assumption (Braun and Stuhler, 2017; Lindahl et al., 2015). Thus, this calculation should be considered a lower bound. Britto et al. (2022) report that in Brazil it takes 7 generations for a child from a family at the 25th percentile to reach the same rank as that of a child from a family at the 75th percentile —their IRA estimate is 0.546.

[24] The estimate for Chile refers to intergenerational mobility of earnings, which is typically lower than the intergenerational mobility of income due to the higher persistence of capital income.

bias introduced when the age of parents and children when their incomes are measured is not similar. The lower IRA for younger children is driven by a smaller correlation between parental and children's ranks —i.e. inequality of the marginal distributions of the two generations is similar.[25][26]

How do income and earnings persistence compare? Columns (3) and (4) show that persistence is about 10% higher for income than for earnings.[27] This result is also consistent with previous findings and reflects the impact of capital income on intergenerational persistence. This impact seems to be stronger for older cohorts than for younger ones, which is consistent with the fact that capital tends to accumulate with age.

What is the impact of the informal market? Ignoring informal labor income, i.e. assuming that informal labor incomes are zero, moderately reduces the estimated IRA. The difference between the IRA estimate when we impute informal labor income (0.292) and when we assume that informal labor income is zero (0.271) is 0.021, which represents a 7.2% decrease –see the IRA-RF and IRA-0% estimates in Panel A of Table 3.[28] In contrast to the moderate effect we find, the only other study that imputes informal earnings to estimate intergenerational mobility for Brazil finds that IRA estimates increase substantially (from 0.357 to 0.546) when informal earnings and formal capital income are imputed using Random Forests methods similar to the method we employ (Britto et al., 2022). Two issues may explain why imputations have a much larger impact in Brazil than in Uruguay. First, Britto et al. (2022) impute formal capital income in addition to informal labor income, which substantially increases the rank-rank slope. Second, the incidence of informal work and the turnover of workers between the formal and informal sectors are much larger in Brazil than in Uruguay.

[25]The correlation of parental and children's ranks is 0.12 for children aged 20-29 while it is 0.23 for children aged 30-39. The relationship between standard deviations of the rank distribution of children and parents is similar for both age groups —1.08 for children aged 20-29 and 1.03 for children aged 30-39.

[26]Table A.7 shows that the share of parent/child dyads for which we observe both parents is substantially larger for the younger cohort of children. However, this different composition is unlikely to account for the difference in IRA between the younger and the older children because the difference in IRA is very similar for the subsample of dyads for which we observe both parents and for the subsample of dyads for which we observe only one parent. We examine the implications of the gender-imbalanced sample of parent of children aged 30-39 in Section 4.4.1.

[27]The null hypothesis that the income-based and the earnings-based IRA estimates are equal to each other is rejected in all cases (p-value< 0.001 for all tests).

[28]As indicated above, the IRA estimate (0.290) remains virtually unchanged when we use OLS regressions instead of Random Forest methods to impute informal labor income.

4.2 Alternative approaches to computing total income with limited data

Using Random Forest (RF) methods to impute informal income, as we do above, is very demanding on data. In addition to the administrative data, which must include the relevant individual characteristics that are used in the Random Forest, we need an ancillary dataset, which should include informal income as well as the same relevant individual characteristics included in the administrative dataset. Ideally, the ancillary dataset should be representative of the population. As explained above, we use the nationally representative household survey data, ECH, as ancillary dataset, which has information on informal income. Now, if information on informal income is not available, this can be imputed by assuming that the productivity of individuals in the informal market is systematically related to their productivity in the formal market (Britto et al., 2022). Panel A in Table 3 reports IRA estimates under alternative assumptions about earnings in years with no observed formal-sector income. Specifically, for those years, we impute informal earnings as a fixed proportion of the individual's observed formal income, using three scenarios: zero income (IRA-0%), 50% of formal income (IRA-50%), and 100% of formal income (IRA-100%). The resulting estimates are very stable across assumptions and remain lower than the IRA obtained when informal income is imputed using Random Forest methods (IRA-RF).[29] This happens because the correlation between parental and children's ranks is higher when we impute informal earnings using RF methods.[30]

What is the bias we introduce when we use an estimation sample composed of individuals with a high attachment to the formal market? Most household surveys in high-income countries do not provide information on informal income. In the absence of informal income, one may opt to

[29] This amounts to considering that individuals may not be equally productive in the informal and in the formal labor markets. According to the National Household Survey data (ECH), informal labor income is on average 39% of formal labor income. We calculate the share of informal earnings using only positive formal earnings observations over the 5-year period we use to compute permanent income. Since administrative records include monthly earnings information, as a robustness check, we have also calculated the share of informal earnings using the average monthly formal earnings reports within the same year, for the years when earnings are strictly positive for some months. IRA estimates from this alternative way of calculating informal earnings are nearly identical to our baseline estimates. They are not reported, but are available upon request.

[30] The correlation of parental and children's ranks is 0.27 when we impute informal earnings using RF methods while it is 0.25 when we impute informal earnings as a percentage of formal earnings. The relationship between standard deviations of the rank distribution of children and parents is similar for all imputing methods —it is 1.05 with RF and between 1 and 1.03 when we impute informal earnings as a percentage of formal earnings.

restrict the estimation sample to individuals who are observed to have positive formal income in each year of the sample period. However, this practice is likely to lead to a significantly smaller and selected sample and thus bias mobility estimates. This is the result we obtain with the data for Uruguay. If we estimate the IRA on a sample of parents and children for whom we have positive formal income in the five consecutive years we use to define permanent income, our estimation sample reduces to 30,192 observations —which is 25% of the baseline estimation sample— and obtain a downward biased IRA estimate of 0.232 (sd= 0.006) for children aged 30-39 years. That is, estimating the IRA on a selected sample of parents and children with high attachment to the formal market reduces the IRA estimate by 14%.[31]

4.3 Importance of the reference distribution

As outlined in Section 2.4, our baseline estimates are based on ranks in the total income distribution, which comprises both formal and informal incomes, as we argue that estimating ranks in the income distribution that only comprises formal incomes provides a biased estimation of income ranks. This subsection documents the difference in intergenerational mobility that arises from using different reference distributions to estimate income ranks. The IRA-RF estimates in Table 3 show that estimating ranks in the *Sample* income distribution, which only includes formal incomes, decreases intergenerational mobility relative to estimating income ranks in the *Global* income distribution, which also includes informal incomes. In particular, IRA-RF decreases by 8.2% when ranks are estimated in the *Sample* income distribution. It is worth noting that the impact of the reference distribution on the IRA is as important as the impact of other methodological decisions, such as imputing or not informal incomes, which is 7.2%. The percentage difference that results from using different reference distributions is also sizeable when we compare it with IRA differences across countries. For instance, it is half the difference between the IRA for a Nordic country, such as Denmark (0.18), and a Southern-European country, such as Italy (0.22). Similarly, it is one fifth of the percentage difference between the IRA in the US (0.34) and the IRA in Germany (0.24).

As we have shown, the choice of reference distribution to estimate income ranks is key. Yet, this issue has received little attention in previous studies that estimate intergenerational mobil-

[31]The IRA estimate of 0.232 we obtain with the reduced and selected sample must be compared with the IRA estimate of 0.271 obtained with the baseline sample, reported in Panel A of Table 3.

ity in countries where the shadow economy is non-negligible (most high-income economies share this feature, especially the Southern-European ones) and that use administrative records.[32] Our results suggest that the IRA estimates in high-income countries could be downward biased, implying lower intergenerational mobility than what previous studies report. Of all previous studies for high-income countries, only Acciari et al. (2022) attempts to correct for tax evasion. Their exercise inflates some sources of income that are easier to hide to tax authorities, such as self-employment and rental income, but does not include individuals that only work in the shadow economy and are, therefore, not present in their sample of tax records. Still, they find that the IRA estimate increases 18% (from 0.22 to 0.26). Importantly, thought, note that since they work with the same sample of individuals, the increase in the estimated IRA does not result from a reshuffling of percentile ranks. The evidence we provide suggests that recomputing the percentile rank of the individuals in their original sample to take due account of the individuals who systematically work in the shadow economy (and who are not observed in their administrative records), would increase the IRA even further.

4.4 Robustness analyses

To check the robustness of our findings, we next present six sensitivity analyses. First, we examine the possible implications of the gender-imbalanced composition of the sample of parents of children aged 30-39. Second, we consider a less restrictive sample selection criteria, which enlarges our sample by including individuals with a weaker attachment to the formal labor market. Third, we address the possible income under-report at the upper tail of the survey income distribution and the possible under-report at the lower tail of the income distribution of the administrative data. Fourth, we enlarge the sample period over which we compute permanent income. Fifth, we use two different indicators of economic status, the logarithm of income and the z-score of income, and thus measure mobility with the intergenerational income elasticity (IGE) and the intergenerational z-score association (IZA). Finally, we report IRA estimates when parental income is the average of the income of both parents for the subsample where we observe both parents.

[32] On average, between 2015 and 2025, the shadow economy represents between 8% and 20% of GDP in most OECD countries. It is largest in the Southern-European countries and smaller than 8% in the US, Switzerland, Austria, and New Zealand (Schneider and Boockmann, 2025).

Table 3: Average IRA for Total Income: Alternative Informal Income Imputations, Reference Distributions, and Samples (Children Aged 30–39)

	Global reference					Sample reference	
Panel A. In sample if at least one positive formal labour income							
	RF	100%	50%	0%	Under-report	RF	Under-report
IRA	0.292***	0.266***	0.271***	0.271***	0.273***	0.268***	0.262***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	98,977	98,977	98,977	98,977	98,977	98,977	98,977
Panel B. In sample even with no positive formal labour income (universal)							
	RF	100%	50%	0%	Under-report	RF	Under-report
IRA	0.263***	0.238***	0.241***	0.240***	0.262***	0.246***	0.245***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	191,194	191,194	191,194	191,194	191,194	191,194	191,194

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. RF imputes informal income using Random Forest methods. The 100%, 50%, and 0% specifications impute informal income as fixed proportions of observed formal income in years without formal earnings. Under-report adjusts for income under-reporting. All relevant IRA comparisons are statistically different ($p < 0.001$), except: in Panel A, 50% vs. 0%; and in Panel B, 50% vs. 0%, and RF vs. Under-report under both reference distributions.

4.4.1 Possible implications of the gender-imbalanced composition of the sample of parents of children aged 30-39

As outlined in section 2.1, a possible concern about the results presented so far is the unbalanced gender composition of the parents of children aged 30-39, which we do not have for the younger sample of children aged 20-29 — in 74% of parent/child dyads parental income is the income of the mother if children are 30-39, while it is only 45% if children are 20-29.[33] This gender imbalance occurs because 87% of the times we only observe one parent, the parent we observe is the mother.

To explore the possible implications of this gender-imbalanced composition of the sample of parents, we split the sample into four subgroups according to whether we observe (the income of) both parents and which parental income is highest. The four groups are the mothers whose spouse is not observed, the fathers whose spouse is not observed, the mothers whose spouse

[33] This imbalance results from the larger share of women among the beneficiaries of the policies administered by the Social Security Administration before the 1980s. The sample of parents of children aged 20-29 is gender balanced because of the increasing coverage of the national health system that took place in the 1980s.

is observed but his income is lower than hers, and the fathers whose spouse is observed but her income is lower than his. If we observed the income of both parents for the entire sample of dyads (and estimated the IRA using the highest income of the parents), the IRA estimates for the subgroup of dyads for which we observe both parents suggest that the IRA could be slightly larger, i.e. 0.31 instead of 0.29 —see the upper panel of Table A.7. However, it could be argued that the IRA estimates we obtain from the samples that include only the dyads for which we observe the income of both parents are not a good counterfactual of the situation where, instead of observing only the income of the mother, we observed the income of both parents. This would be the case if the observations of the four groups mentioned above were not uniformly distributed across the income percentiles. Figure A.3 shows that the observations in all groups of parents and children are distributed in a way that is very much uniform across the income percentiles. The group of mothers whose spouse is not observed shows a slightly higher frequency in the bottom quintile, which should not raise serious concerns. In sum, the evidence we provide suggests that our baseline IRA estimates would not be too different if we observed the income of both parents for all children aged 30-39 in our sample.

4.4.2 Less restrictive sample selection criteria

Our sample so far consists of children and parents with positive formal labor income at least in one of the five years we use to average incomes. This allows the identification of children and parents who maintained an active attachment to the formal labor market during the period under consideration. Observing positive formal earnings for at least one year improves the prediction of informal earnings.[34] However, it also reduces the sample size (to nearly half) and excludes parents and children with weaker attachment to the formal market from our analysis. To enlarge our sample and include individuals with a weaker attachment to the formal labor market, we build a sample that includes all paired children and parents, irrespective of their participation in the formal labor market. As the comparison of the IRA estimates in Panels A and B of Table 3 shows, intergenerational mobility is greater in the larger sample: average IRA is now between 7 and 10% smaller, depending on the method we use to predict informal

[34]Recall that the decile of formal earnings is one of the predicting variables in the RF model we use to predict informal earnings. Since the share of the sample for each year and generation whose formal income is zero is larger than one decile (it is 18% for the children aged 30-39 and 19% for the parents), the prediction of informal earnings at the bottom of the distribution becomes more imprecise.

incomes. The reduced IRA estimates for the larger sample are driven by a smaller rank-rank correlation among the new pairs of parent/child included in the extended sample. Intuitively, since one of the individuals of the pair does not work in the formal sector, the rank-rank correlation with the other individual who may have a stronger attachment to the formal market is weaker. The reduced IRA is also consistent with the attenuation bias that results from the larger measurement error introduced by a less precise estimation of informal labor income of the individuals who are not observed in the formal labor market during the 5-year sample period.

4.4.3 Addressing income misreport

Our estimates are based on a vector of formal income drawn from administrative records and a vector of informal income that is imputed using a Random Forest process and the combination of survey and administrative data. This allows us to construct a total income vector that aims to incorporate informal earnings or income not reported by workers (due to evasion, coverage issues, or noncompliance). These estimates assume that there are no information problems in the income vectors from administrative records and surveys. However, administrative records can suffer from measurement problems due to reporting errors, tax evasion, incomplete coverage, and inconsistencies in definitions across data sources. While, household surveys face representation, non-response, and measurement errors, which are especially problematic at the top of the income distribution, where high-income individuals are harder to reach, more likely to refuse to respond, or tend to under-report their income. Acciari et al. (2022) address problems in administrative record data related to tax evasion. They assume that survey data are closer to the true income and adjust upward reported income for categories with higher expected tax evasion in Italy, mainly the self-employed. This strategy helps mitigate tax evasion problems associated with specific categories. It is worth noting that the informal income prediction model we use in our baseline strategy could capture part of this mechanism. However, to address information problems linked to both surveys and administrative records, we adopt a different approach.

Using a matched sample linking administrative records and survey data, Flachaire et al. (2023) show that inconsistencies in income vectors depend on the relative position of individuals in the income distribution. At the bottom two deciles, administrative records tend to under-capture income relative to surveys. In contrast, survey under-reporting is concentrated at the top two deciles of the tax-records income distribution, and it increases with income. This implies that the relationship between the survey (ECH) income vector and the administrative records is not uniform across the distribution.

To account for this, we proceed as follows. First, we apply decile-specific misreporting ratios to adjust the ECH income according to the position of the individual in the income distribution. Using this corrected income vector, we re-estimate the Random Forest model and impute the recovered income for both parents and children. We finally estimate the IRA using these new income vectors for parents and their children.

The effect of taking due account of income misreport is the same as that of disregarding informal income. As the IRA estimates of the global reference distribution in Panel A of Table 3 show, the IRA decreases by 7.2% when we correct for income misreport in both the survey and the administrative datasets. This suggests that IRA estimates that do not correct for income misreports are upward biased. However, this effect disappears when we use the larger sample to estimate the IRA (see Panel B). Two elements are likely to explain this result. First, the income changes that result from the decile-specific adjustment are likely to be more relevant at the bottom than at the upper end of the income distribution, since persistence is larger at the upper end, which means that the adjustment modifies the income of the parent and the child in a similar way. Second, the income adjustments made to correct for misreporting are likely not to affect the new sample members with no formal earnings introduced in the larger sample.

4.4.4 Longer sample period

We estimate the average IRA using the longest period of tax records we have, i.e. the eight-year window from 2009 to 2016. It is important to note that we are not changing the definition of permanent income (average income over 5 years), the reference distribution to compute the income ranks, or the sample selection criteria, e.g. having at least one positive formal income. However, by increasing the time window, we incorporate in the estimation sample new parent/child pairs who did not report any positive formal income over the original 5-year window, but who have at least one positive formal income in the other three years of the enlarged time window. As a result, the size of the estimation samples increases, but the inequality in the marginal distributions of the two generations remains unchanged.[35] The estimated average IRA with the longer sample period is lower (0.272) than the baseline IRA, suggesting once again that average persistence decreases when we include individuals with weaker attachment to the formal sector (see Table 4).[36]

Table 4: Average IRA for Total Income: Longer Sample Period (8 Years), Children Aged 30–39

	Baseline RF	100%	50%	0%
IRA	0.272***	0.269***	0.267***	0.259***
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	165,826	165,826	165,826	165,826

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. Baseline RF imputes informal income using Random Forest methods. The 100%, 50%, and 0% specifications impute informal income as fixed proportions of observed formal income in years without formal earnings. All IRA comparisons are statistically different ($p < 0.001$).

4.4.5 Alternative indicators of economic status

Previous studies use the log of income as the indicator of economic status, which leads to measuring mobility with the intergenerational income elasticity (IGE). In Panel A of Table 5, we report IGE estimates using log incomes and the log-log specification typically used in the literature.[37] Consistent with previous results (Chetty et al., 2014a), we find that persistence is lower in the log-log specification than in the rank-rank specification. This is possibly due to the sensitivity of this indicator to the definition of permanent income and measurement error. Chetty et al. (2014a) find that log-log specifications yield more unstable mobility estimates because the relationship between log child income and log parent income is nonlinear and the estimate is also more sensitive to the treatment of zero incomes.[38]

The IRA and IGE reported above capture different aspects of intergenerational mobility and use different measures of the economic status of individuals. The IRA captures exchange

[35]The number of observations increases from 98,977 (5-year window) to 165,826 (8-year window). Table A.8 summarizes the distribution of years with positive earnings for parents and children.

[36]The reduction in the IRA in this sample holds for labor income. This notwithstanding, for the youngest age group, expanding the time-window slightly increases the IRA, suggesting that we are capturing better their permanent income –see Table A.9 in the Appendix.

[37]The IGE captures regression to the mean rather than positional mobility.

[38]We assign an income of one to zero incomes.

mobility and is driven by the rank correlation between children and their parents, since the distribution of percentile ranks has the same standard deviation in both generations of children and parents.[39] In contrast, the IGE captures both exchange and structural mobility, as the (marginal) distributions of log-income of the generations of children and parents differ. Regarding the measure of economic status, ranks fail to account for the distance, in terms of income, between positions in the income distribution.

We measure the economic status of an individual by the z-score of her income level and measure intergenerational mobility by the intergenerational association of z-scores (IZA).[40] Since z-scores normalize the standard deviation of the marginal distributions, the IZA captures exchange mobility, like the IRA, but takes due account of the distance between individuals, like the IGE, as the measure of economic status now indicates how many standard deviations an observation lies away from the mean of the distribution. Panel B in Table 5 shows that the IZA estimates lie between the IRA and the IGE estimates.

Table 5: Average Intergenerational Mobility: Alternative Measures and Economic Status Indicators (Children Aged 30–39)

	Baseline RF	100%	50%	0%
IGE (log-income)	0.262***	0.224***	0.218***	0.198***
	(0.004)	(0.004)	(0.004)	(0.003)
Observations	98,977	98,977	98,977	98,977
IZA (z-score)	0.280***	0.280***	0.279***	0.273***
	(0.036)	(0.034)	(0.034)	(0.034)
Observations	98,977	98,977	98,977	98,977

Notes: The table reports coefficients from OLS regressions of the child's economic status on the parent's economic status. Economic status is measured as log income for the Intergenerational Earnings Elasticity (IGE) and as the standardized income score for the Intergenerational Z-score Association (IZA). Controls include child sex and parent age and sex. Standard errors are reported in parentheses. Baseline RF imputes informal income using Random Forest methods. The 100%, 50%, and 0% specifications impute informal income as fixed proportions of observed formal income in years without formal earnings. All relevant comparisons are statistically different ($p < 0.05$), except IZA-RF vs. IZA-100%, and IZA-100% vs. IZA-50%.

[39]Note that this is what happens in most studies, but this is not the case when we use the global reference distribution to estimate the percentile ranks, as explained above.

[40]The income z-score of a child/parent is her income minus mean income of the distribution of her generation divided by the standard deviation of the distribution. Compared to percentile ranks, this measure incorporates greater variability, but unlike the logarithm of income, it is a relative measure, which could mitigate measurement error issues.

4.4.6 Parental income as the average income of both parents

Previous studies tend to define parental income as the average income of both parents. Since we observe both parents in 28% (43%) of parent/child dyads for the cohort of children aged 30-39 (20-29) years, in the baseline analysis, we prefer to define parental income as the maximum income of both parents.[41] To gain insight into how the baseline IRA estimate would change if we used average parental income, we report the results of two exercises. In a first exercise, we estimate the IRA replacing the maximum income of the parents with the average income, when both parents are observed —and leaving the income of the parent we observe when only one parent is observed. The IRA estimates that we obtain from this exercise are very similar to the baseline estimates, which use the maximum income of the parents. For children aged 30-39, the IRA is 0.289 (sd= 0.003) while for children aged 20-29, the IRA is 0.125 (sd= 0.002). In a second exercise, we compare the IRA from the subsample of parent/child pairs when parental income is the average income of the parents and when parental income is the maximum income of both parents, and find that the IRA are similar. For children aged 30-39, the IRA is slightly larger if average parental income is used (0.358 vs. 0.306), while for children aged 20-29, it is slightly smaller and not statistically different (0.137 vs. 0.141).

4.5 The gender of parents and children matter

Gender has been found to be a source of heterogeneity in the transmission of economic advantage between parents and children. Previous literature suggests differences in parental decisions about investments in human and social capital, as well as the transmission of norms, preferences, and attitudes across generations, which could determine differences in persistence by gender (Lundberg, 2005; Bütikofer, 2013; Farré and Vella, 2013; Morrill and Morrill, 2013; Fernandez and Fogli, 2009; Fernández et al., 2004). In the presence of assortative mating, women from better-off backgrounds tend to marry richer partners and are more likely to work fewer hours or not to work at all. If this is the case, rank-rank slope estimates should be lower for daughters than for sons. In addition, the earnings and income distribution is typically more compressed for women than for men. While inequality in the marginal income distributions of parents and children do not have any bearing on IRA estimates, this implies lower intergenerational income

[41] Parental income when we only observe one parent is of course the income of the parent we observe.

elasticities for women, ceteris paribus.[42]

Previous studies based on administrative records show that gender differences in intergenerational mobility are sensitive to the way mobility is measured (IGE or IRA) and to the definition of the outcome (Mitnik et al., 2015; Chetty et al., 2014a; Mazumder, 2005; Dahl and DeLeire, 2008). Using the same methodological choices as ours, Chetty et al. (2014a) finds a smaller IRA for daughters than for sons in the US. However, Britto et al. (2022) find the opposite for Brazil. Given that the disparity in intergenerational mobility by gender found in previous work refers mostly to rich countries, in this section we examine whether this is also the case in a poorer country with different institutions and potential cultural differences about the role of gender in society. The early demographic transition and the increasing participation of women in the labor market make Uruguay an interesting case. We analyse and examine whether intergenerational mobility differs by the gender of children, the gender of parents, and both.

The IRA estimates reported in Table 6 indicate that, on average, intergenerational persistence is about 16% higher for daughters than for sons. The difference in IRA between genders is very similar (14%) for the younger cohort of children, aged 20-29. Our estimates for Uruguay, then, are not in line with previous estimates for the US and challenge standard arguments to understand gender differences in intergenerational mobility.[43]

Table 6: Average IRA for Total Income by Child Gender and Age Group

	Age 30–39		Age 20–29	
	Sons	Daughters	Sons	Daughters
IRA	0.261***	0.303***	0.126***	0.144***
	(0.005)	(0.005)	(0.003)	(0.003)
Observations	50,328	48,649	119,608	106,650

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. All reported IRA estimates are statistically significant ($p < 0.001$).

[42]Since the estimation sample we use is a subset of the reference distribution we use to calculate the relative position of individuals, inequality in the distribution of ranks of children and of parents are relevant to determine the IRA.

[43]Bernuy and Esteve (2019) provide evidence that there is less educational homogamy in young couples in Uruguay than in the US. The labor market behavior of young women in Uruguay is very similar to men's, and it is not associated with the prototypical role of "secondary worker" (Espino et al., 2017). For instance, the labor market participation rate of women and men with tertiary education in Uruguay are very similar.

Table 7 shows the levels of persistence by gender of the parent with the highest income in the household, a dimension not explored in previous studies.[44] As the columns "Average" show, average persistence does not change by parental gender. However, this apparent similarity conceals substantial heterogeneity in the IRA when we consider the gender of children. Income persistence is higher when parents and children have the same gender. That is, the father-son IRA is higher than the father-daughter IRA, while the mother-daughter IRA is higher than the mother-son IRA. The higher IRA between parents and children of the same gender is consistent with the intergenerational transmission of role models from mothers to daughters and from fathers to sons reported in some studies (Perales et al., 2021; Weiss, 2023). The mother-daughter IRA is much higher than the father-son IRA. This is what drives the higher IRA estimate of the daughters reported in Table 6. Our results by gender from Table 7 also hold for earnings (see Table A.10), for the youngest age group (see Tables A.11 and A.12), and when we use the eight-year sample period (see Table A.13).

Table 7: Average IRA for Total Income by Parent Gender, Children Aged 30–39

	Fathers (maximum income)			Mothers (maximum income)		
	Average	Sons	Daughters	Average	Sons	Daughters
IRA	0.295***	0.308***	0.281***	0.294***	0.261***	0.327***
	(0.007)	(0.010)	(0.009)	(0.004)	(0.006)	(0.005)
Observations	26,094	13,517	12,577	72,883	36,811	36,072

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. All relevant IRA comparisons are statistically different ($p < 0.05$), except the average IRA for fathers vs. mothers ($p = 0.87$), and sons vs. daughters when the parent is the father ($p = 0.06$).

4.6 Your mobility depends on how rich your parents are

Rich and poor parents do not have the same resources and mechanisms to transmit economic advantage to their children. At the lower tail of the distribution, credit constraints are a plausible mechanisms to explain the concave relationship between child and parent incomes.[45] Informality may also be a source of persistence among poorer parents if informality is correlated

[44] Recall that our sample comprises households for which only one parent is observed and households for which both parents are observed. If only one parent is observed, we use his or her income in the analysis.

[45] The early contribution by Becker and Tomes (1986) explores this mechanism. More recent work by Grawe (2004), however, calls into question the relevance of credit constraints in producing non-linear intergenerational transmission of incomes.

across generations and wages in the informal and formal sectors are systematically different. The larger inheritance of wealth and employers, together with the segregation across neighbourhoods can generate larger intergenerational persistence at the top of the income distribution (see for example, Björklund et al. (2012); Corak and Piraino (2010); Durlauf and Seshadri (2018); Chetty and Hendren (2018)).

Consistent with these arguments, previous studies, especially for some European countries, report higher rank-rank slopes at both tails of the distribution (Acciari et al., 2022; Bratberg et al., 2017; Nybom and Stuhler, 2017). Evidence for other countries, such as US, Australia, and Brazil, however, shows rather homogeneous rank-rank slopes across the income distribution (Chetty et al., 2014a; Deutscher and Mazumder, 2020; Britto et al., 2022). Given the disparate previous evidence, we examine whether intergenerational income rank persistence is linear in Uruguay.

We use three complementary empirical strategies to explore non-linearities. As outlined in Section 3, we first estimate non-linear parametric spline regressions with pre-defined knots at percentiles P10, P25, P50, P75, P90 and P99 (equation (2)). Then, we use a more flexible (but still parametric) strategy and fit a separate standard IRA regression model (equation (1)) for each fractile defined by the knots we use in the spline regressions.[46] Finally, we analyse the expected income rank of children conditional on their parents' rank, which results from flexible non-parametric kernel-weighted polynomial regressions (equation (3)).

Panel (a) in Figure 3 plots the expected rank of children conditional on parent's rank that result from the linear model and from three flexible models described above.[47] The results from these three specifications consistently show a non-linear relationship between the income of both generations. Persistence (the slope) is higher at the upper half of the distribution than at the bottom half and is especially pronounced at the top decile. Thus, these results for Uruguay differ from previous findings for the US, Australia, and Brazil, where the rank-rank slope is found to be linear, but also differ from the results for Italy, Germany, or Sweden, where persistence is larger both at the upper and lower tails. We find larger persistence at the upper

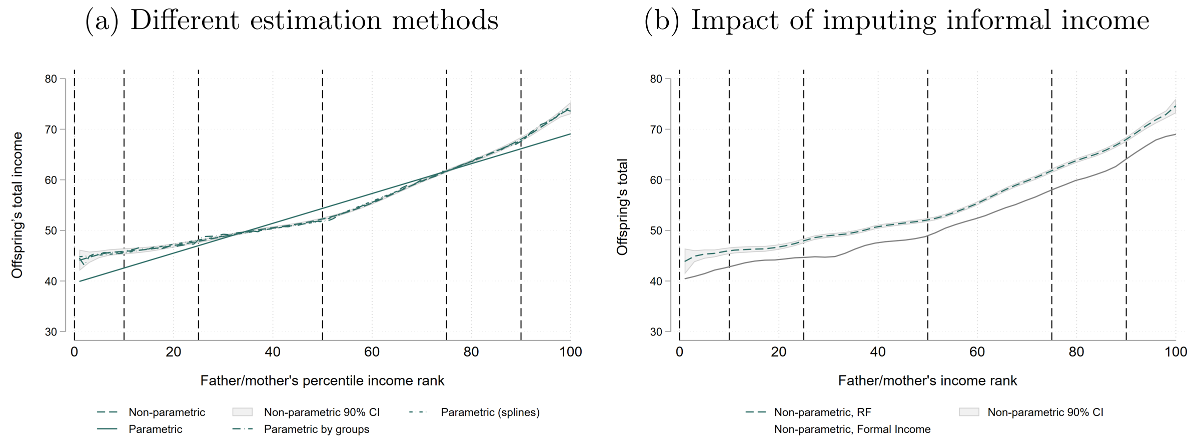
[46]This strategy lifts the assumption of a common intercept for all fractiles imposed by spline regressions.

[47]The prediction of the splines model and the parametric model with multiple intercepts are based on the estimated coefficients shown in Figure A.6, while the prediction of the linear model is based on the estimated coefficients of Table 2.

but not at the lower tail of the distribution.[48]

Panel (b) illustrates the effect of imputing informal income. The predicted rank-rank relationship when we impute informal income is roughly parallel to the rank-rank relationship for formal income. This means that imputing informal income shifts the intercept upward, but does not substantially change the slope, i.e. the IRA estimates.[49][50] This is what Figure A.6 shows, which displays the point estimates for the slopes and intercepts for the two parametric methods. This figure also shows that IRA estimates across the income distribution are robust to the estimation method.

Figure 3: Income Rank persistence across the income distribution. Children aged 30-39



Notes: Panel (a) compares alternative estimation methods for the relationship between parental income rank and offspring income rank among children aged 30–39. We report non-parametric estimates, linear parametric estimates, estimates by percentile groups, and spline specifications. Panel (b) compares non-parametric estimates using only observed formal labor income with estimates that additionally incorporate imputed informal income using Random Forest methods. Shaded areas represent 90% confidence intervals. Vertical dashed lines indicate selected parental income rank percentiles. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

[48]Earnings show a similar relationship between children and parents' income ranks across the income distribution —see Figure A.5.

[49]Note, however, that the point estimate of the IRA is smaller (though imprecisely estimated) in the poorest decile when we impute informal income. This happens because in the first decile we find the largest number (and also share) of child/parent pairs in which both do not have formal income in at least one year. Imputing informal income increases the variability and reduces the rank correlation between parents and children.

[50]Note that despite the IRA of the different fractiles and of the splines not changing, the average IRA reported in Table 3 increases when we impute informal income. This highlights the importance of the intercept in intergenerational mobility regressions. The average IRA increases because the intercept decreases. The point estimates of the intercept with and without informal income are 6.4 and 12.2, respectively.

5 Final comments

We contribute to the very incipient literature that estimates intergenerational mobility of income from large-scale administrative data using high-quality income data and provide novel evidence of intergenerational income mobility in a middle-income country, Uruguay. Using income tax administrative data, social security records, and household survey data, we estimate a rank-rank slope (IRA) of 0.292. This suggests that intergenerational income persistence is higher in Uruguay than in high-income countries, but lower than in the US.

Informal labor markets are an important feature of low- and middle-income countries, which increase earnings inequality and are likely to contribute to intergenerational income mobility. The presence of informality is a major and distinct challenge that studies in poor- and middle-income countries face, as administrative data only provide information on the formal sector of the economy. We pay particular attention to the implications of including the informal market in the estimation of intergenerational income mobility and document that not taking due account of the informal sector introduces a substantial downward bias in the estimation of intergenerational income mobility.

The large-scale data we use allow us to obtain precise estimates of intergenerational income mobility at different parts of the income distribution and to examine whether parental income conditions intergenerational income persistence. We find persistence to be higher at the upper half of the income distribution, especially in the richest decile. This pattern differs from the pattern reported in high-income countries, as well as Brazil, where intergenerational income mobility is constant for the middle 80% of the distribution and is lower at the tails. The greater intergenerational persistence we find for the very rich can undermine the quality of democratic institutions and contribute to the political inequality between citizens. This implications may call for policies to reduce the transmission of economic advantage among the rich (Robeyns, 2019).

Our analysis by sex suggests that persistence is higher for daughters than for sons. When we cross the sex of parents and children, we find that persistence is largest when parents and children have the same sex, which suggests the importance of the vertical same-sex intergenerational transmission of gender roles.

Our results are relevant for social welfare reasons and for public policy. Intergenerational mobility is usually related to equality of opportunity. [51] Our findings suggest that the economic status of individuals depends on who their parents are, to a greater extent than in most high-income countries. This is especially true for women.

[51] For a discussion of the literature on inequality of opportunity, see Ramos and Van de gaer (2016); Roemer and Trannoy (2016); Jantti and Jenkins (2015).

References

Acciari, P., Polo, A., and Violante, G. L. (2022). And yet it moves: Intergenerational mobility in italy. *American Economic Journal: Applied Economics*, 14(3):118–63.

Aghion, P., Banerjee, A., and Piketty, T. (1999). Dualism and macroeconomic volatility. *The Quarterly Journal of Economics*, 114(4):1359–1397.

Alvaredo, F. and Gasparini, L. (2015). Recent trends in inequality and poverty in developing countries. *Handbook of income distribution*, 2:697–805.

Amarante, V., Arim, R., and Yapor, M. (2016). Decomposing inequality changes in uruguay: the role of formalization in the labor market. *IZA Journal of Labor & Development*, 5(1):13.

Amarante, V. and Gómez, M. (2016). El proceso de formalización en el mercado laboral uruguayo. *Serie de Estudios y Perspectivas*, 20.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47:1148–1178.

Atkinson, A. B. (2007). Measuring top incomes: methodological issues. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries*, 1:18–42.

Becker, G. S., Kominers, S. D., Murphy, K. M., and Spenkuch, J. L. (2018). A theory of intergenerational mobility. *Journal of Political Economy*, 126(S1):S7–S25.

Becker, G. S. and Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4(3, Part. 2):S1–S39.

Bernuy, V. S. J. and Esteve, A. (2019). Amores imposibles: la brecha entre universitarios y el resto de grupos educativos en los mercados matrimoniales de américa latina, 1970-2010. *Notas de Población*, 46(108):11–36.

Björklund, A., Roine, J., and Waldenstöm, D. (2012). Intergenerational top income mobility in sweden: Capitalist dynasties in the land of equal opportunity?. *Journal of Public Economics*, 96:474–484.

Böhlmark, A. and Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for sweden. *Journal of Labor Economics*, 24(4):879–896.

Bourguignon, F., Ferreira, F. H., and Walton, M. (2007). Equity, efficiency and inequality traps: A research agenda. *Journal of Economic Inequality*.

Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D., and Vaage, K. (2017). A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US. *Sandinvian Journal of Economics*, 119(1):72–101.

Braun, S. and Stuhler, J. (2017). The transmission of inequality across multiple generations: Testing recent theories with evidence from germany. *Economic Journal*, 609(128):576–611.

Breen, R. and García-Peñalosa, C. (2005). Income inequality and macroeconomic volatility: an empirical investigation. *Review of Development Economics*, 9(3):380–398.

Britto, D., Fonseca, A. d. A., Pinotti, P., Sampaio, B., and Warwar, L. (2022). Intergenerational mobility in the land of inequality. *BAFFI CAREFIN Centre Research Paper No. 2022-186, Available at SSRN: https://ssrn.com/abstract=4237631 or http://dx.doi.org/10.2139/ssrn.4237631.*

Burdín, G., De Rosa, M., Vigorito, A., and Vilá, J. (2022). Falling inequality and the growing capital income share: Reconciling divergent trends in survey and tax data. *World Development*, 152:105783.

Bütikofer, A. (2013). Revisiting 'mothers and sons' preference formation and the female labor force in Switzerland. *Labour Economics*, 20:82–91.

Carrasco, P. (2012). El efecto de las condiciones de ingreso al mercado de trabajo en los jóvenes uruguayos: Un análisis basado en la protección de la seguridad social. *Serie Documentos de Trabajo; 13/12*.

Carrasco, P. (2021). Efectos dinámicos de los empleos de baja remuneración en uruguay. *Serie Documentos de Trabajo; 31/21*.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014a). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014b). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, 104(5):141–47.

Cleveland, W. S., Devlin, S. J., and Grosse, E. (1988). Regression by local fitting: Methods, properties, and computational algorithms. *Journal of Econometrics*, 37(1):87–114.

Corak, M. and Piraino, P. (2010). Intergenerational Earnings Mobility and the Inheritance of Employers. IZA Discussion Papers 4876, Institute of Labor Economics (IZA).

Cortés Orihuela, J., Díaz, J. D., Gutiérrez Cubillos, P., Troncoso, P. A., and Villarroel, G. I. (2024). Intergenerational earnings mobility in chile: the tale of the upper tail. *Empirical Economics*, 67(5):2411–2447.

Dahl, M. and DeLeire, T. (2008). The association between children's earnings and fathers' lifetime earnings: Estimates using administrative data. *Discussion Paper No. 1342-08*.

Deutscher, N. and Mazumder, B. (2020). Intergenerational mobility across Australia and the stability of regional estimates. *Labor Economics*, 66:101861.

Durlauf, S. N. and Seshadri, A. (2018). Understanding the great gatsby curve. *NBER Macroeconomics Annual*, 32(1):333–393.

Díaz, J. D., Gutiérrez, P., Troncoso, P., and Villarroel, G. I. (2021). Intergenerational Earnings Mobility in Chile: A tale of two tails. *unpublished manuscript*.

Espino, A., Isabella, F., Leites, M., and Machado, A. (2017). Do women have different labor supply behaviors? evidence based on educational groups in uruguay. *Feminist Economics*, 23(4):143–169.

Farré, L. and Vella, F. (2013). The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labour Force Participation. *Economica*, 80(318):219–247.

Fernandez, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, 1(1):146–77.

Fernández, R., Fogli, A., and Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *Quarterly Journal of Economics*, 119(4):1249–1299.

Flachaire, E., Lustig, N., and Vigorito, A. (2023). Underreporting of top incomes and inequality: A comparison of correction methods using simulations and linked survey and tax data. *Review of Income and Wealth*, 69(4):1033–1059.

Grawe, N. (2004). *Intergenerational mobility for whom? The experience of high- and low-earning sons in international perspective.* Cambridge: Cambridge University Press.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4):1308–1320.

Hertz, T. (2009). *Chapter Five. Rags, Riches, and Race.* Princeton University Press.

Instituto Nacional de Estadística (2026). Series históricas de actividad, empleo y desempleo.

Jantti, M. and Jenkins, S. (2015). Income mobility. In *Handbook of Income Distribution. Vol. 2.*

Krueger, A. B. (2012). *The rise and consequences of inequality in the United States.* Speech, Center for American Progress, Washington, DC.

Leites, M., Salas, G., Rivero, L., Suarez, L., and Vigorito, A. (2018). *Trayectorias educativas y laborales de los jóvenes en Uruguay.* BID-Espacio Público-IDRC.

Lindahl, M., Palme, M., Sandgren-Massih, S., and Sjögren, A. (2015). Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations. *Journal of Human Resources*, 50(1):1–33.

Lundberg, S. (2005). The Division of Labor by New Parents: Does Child Gender Matter? *IZA Discussion Papers*, 1787.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. *The Review of Economics and Statistics*, 87:235–255.

Mazumder, B. (2016). Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data. In *Inequality: Causes and Consequences*, volume 43 of *Research in Labor Economics*, pages 83–129. Emerald Group Publishing Limited.

Mitnik, P., Bryant, V., Weber, M., and Grusky, D. B. (2015). New estimates of intergenerational mobility using administrative data. *Statistics of Income Division working paper, Internal Revenue Service.*

Morrill, M. S. and Morrill, T. (2013). Intergenerational links in female labor force participation. *Labour Economics*, 20:38–47.

Muñoz, E. and van der Weide, R. (2025). Intergenerational income mobility around the world: A new data base. *Policy Research Working Paper 11166.*

Nybom, M. and Stuhler, J. (2017). Biases in Standard Measures of Intergenerational Income Dependence. *Journal of Human Resources*, 52(3):800–825.

Perales, F., Hoffmann, H., King, T., Vidal, S., and Baxter, J. (2021). Mothers, fathers and the intergenerational transmission of gender ideology. *Social Science Research*, 99:102597.

Ramos, X. and Van de gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys*, 30(5):855–883.

Robeyns, I. (2019). What, if anything, is wrong with extreme wealth? *Journal of Human Development and Capabilities*, 20(3):251–266.

Roemer, J. E. and Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4):1288–1332.

Schneider, F. and Boockmann, B. (2025). Die größe der schattenwirtschaft — methodik und berechnungen für das jahr 2025. *IAW Working Paper*.

Stuhler, J. (2018). A review of intergenerational mobility and its drivers. JRC Research Reports JRC112247, Joint Research Centre.

Weiss, J. (2023). Intergenerational transmission of left-right ideology: A question of gender and parenting style? *Frontiers in Political Science*, Volume 5 - 2023.

A Appendix: Additional Tables

Table A.1: Summary statistics: observations and female share by offspring cohort, 2012

	Population coverage			Positive income			Female share			
Cohort	Sample (1)	HH survey (2)	Coverage (3)	Sample (4)	Tax records (5)	Coverage (6)	Sample (7)	Sample inc. (8)	HH survey (9)	Tax inc. (10)
1970	13638	42039	32.4%	7165	25706	27.9%	51.1%	48.8%	52.8%	47.0%
1971	14798	39596	37.4%	8051	26653	30.2%	50.0%	47.9%	51.7%	46.9%
1972	15614	43335	36.0%	8610	26684	32.3%	50.6%	49.4%	55.1%	47.6%
1973	16499	41161	40.1%	9158	27329	33.5%	50.0%	48.9%	50.7%	47.2%
1974	17981	45859	39.2%	10159	28970	35.1%	49.8%	48.7%	51.7%	46.8%
1975	18969	45503	41.7%	10876	30973	35.1%	49.9%	48.2%	53.6%	46.5%
1976	19745	48286	40.9%	11574	32081	36.1%	49.9%	47.8%	54.1%	46.6%
1977	19886	47981	41.4%	11613	31976	36.3%	50.4%	48.8%	52.8%	46.9%
1978	19296	46016	41.9%	11586	32302	35.9%	50.3%	48.7%	50.1%	46.6%
1979	18982	46392	40.9%	11639	32198	36.1%	50.9%	49.3%	52.4%	46.6%
1980	17671	44599	39.6%	10960	31128	35.2%	50.9%	48.9%	53.3%	46.3%
1981	18580	41281	45.0%	11717	31789	36.9%	50.7%	49.4%	53.4%	46.1%
1982	26545	47364	56.0%	16771	32149	52.2%	50.9%	49.1%	49.3%	46.4%
1983	32596	42300	77.1%	20438	32023	63.8%	49.5%	46.9%	50.5%	45.7%
1984	33985	44844	75.8%	21487	32634	65.8%	49.6%	46.7%	52.2%	45.6%
1985	35721	44102	81.0%	22788	34010	67.0%	49.6%	46.8%	51.9%	45.6%
1986	36826	45269	81.3%	23417	34185	68.5%	49.4%	46.5%	51.4%	45.5%
1987	37310	45146	82.6%	23493	33364	70.4%	49.3%	46.2%	50.9%	45.1%
1988	40181	47727	84.2%	24798	34066	72.8%	49.6%	46.3%	50.5%	45.5%
1989	40842	47396	86.2%	24512	32771	74.8%	50.0%	46.2%	50.1%	45.3%
1990	45142	51396	87.8%	26234	32053	81.8%	49.3%	45.1%	49.8%	44.7%
1991	47252	49767	94.9%	25785	29603	87.1%	49.2%	43.6%	50.7%	43.3%
1992	48297	49431	97.7%	24507	27478	89.2%	49.2%	42.2%	50.6%	41.9%
1993	51687	45902	112.6%	23231	25518	91.0%	49.2%	41.2%	50.0%	40.7%
1994	53638	50707	105.8%	15172	16358	92.7%	49.1%	37.8%	48.6%	37.5%
1995	55212	55054	100.3%	2433	2552	95.3%	49.1%	40.0%	47.3%	39.7%
1996	55961	54490	102.7%	1028	1067	96.3%	49.4%	29.9%	47.3%	30.1%

Notes: The table reports the number of observations by offspring birth cohort across datasets. Population coverage compares the family-linkage sample with population projections from household surveys. Positive-income coverage compares the sample with positive earnings to tax records with positive income. Female shares are reported for the full sample, the sample with positive income, household survey projections, and tax records with positive income. *Source:* Based on social security records (BPS), tax records (DGI), and household survey data (INE).

Table A.2: Summary statistics: observations and female share by parent cohort, 2012

	Population coverage			Positive income			Female share			
Cohort	Sample (1)	HH survey (2)	Coverage (3)	Sample (4)	Tax records (5)	Coverage (6)	Sample (7)	Sample inc. (8)	HH survey (9)	Tax inc. (10)
1944	9399	24072	39.0%	1211	3811	31.8%	66.7%	64.8%	56.5%	38.9%
1945	10549	26549	39.7%	1489	4427	33.6%	65.9%	62.2%	57.0%	38.2%
1946	11454	25346	45.2%	1909	5267	36.2%	65.8%	63.1%	55.4%	40.0%
1947	12540	29882	42.0%	2529	6336	39.9%	65.6%	61.6%	55.7%	40.7%
1948	13938	27754	50.2%	3027	7186	42.1%	64.1%	57.4%	54.7%	40.4%
1949	15048	30565	49.2%	3637	8094	44.9%	64.1%	57.7%	52.9%	41.6%
1950	16171	32466	49.8%	4582	9737	47.1%	63.3%	56.5%	54.7%	42.0%
1951	16949	27258	62.2%	5730	11496	49.8%	62.0%	56.0%	51.3%	43.0%
1952	18072	33317	54.2%	7630	14307	53.3%	60.3%	55.7%	55.4%	44.3%
1953	19016	30375	62.6%	8620	15601	55.3%	59.2%	55.0%	55.2%	44.9%
1954	21076	34604	60.9%	10104	17629	57.3%	57.4%	53.4%	52.3%	45.0%
1955	22790	35412	64.4%	11343	19390	58.5%	55.5%	51.0%	53.6%	44.8%
1956	24115	38545	62.6%	11952	20216	59.1%	55.4%	51.7%	54.0%	46.1%
1957	24615	35353	69.6%	12414	20541	60.4%	54.2%	50.8%	53.7%	46.6%
1958	25498	36551	69.8%	12874	21291	60.5%	52.7%	48.2%	54.0%	46.2%
1959	26426	36590	72.2%	13216	21785	60.7%	51.7%	47.7%	52.2%	46.8%
1960	27348	42167	64.9%	13666	22509	60.7%	51.5%	46.2%	51.8%	45.8%
1961	29374	34915	84.1%	14715	23503	62.6%	51.2%	46.5%	54.9%	46.6%
1962	31059	41655	74.6%	15316	24175	63.4%	51.7%	47.3%	55.4%	47.0%
1963	31682	39668	79.9%	15267	23920	63.8%	51.5%	47.5%	50.5%	46.7%
1964	32274	39669	81.4%	15370	24009	64.0%	52.0%	47.8%	51.2%	47.4%
1965	31860	39921	79.8%	14897	23410	63.6%	52.5%	48.7%	54.6%	47.3%
1966	31482	38605	81.5%	14522	22988	63.2%	52.3%	48.3%	52.5%	47.4%
1967	32033	39635	80.8%	14276	23372	61.1%	52.6%	48.0%	53.2%	47.1%
1968	32874	36670	89.6%	14384	23953	60.1%	52.4%	49.0%	53.7%	47.0%
1969	35865	40589	88.4%	15194	25985	58.5%	53.0%	48.5%	52.8%	47.2%

Notes: The table reports the number of observations by parent birth cohort across datasets. Population coverage compares the family-linkage sample with population projections from household surveys. Positive-income coverage compares the sample with positive earnings to tax records with positive income. Female shares are reported for the full sample, the sample with positive income, household survey projections, and tax records with positive income. *Source:* Based on social security records (BPS), tax records (DGI), and household survey data (INE).

Table A.3: Educational Assortative Mating by Years of Formal Education, Couples Aged 25 and Older

		Women					
	Yrs. education	0-6	6-9	9-12	12-16	> 16	Total
Men	0-6	**20%**	9%	4%	1%	1%	35%
	6-9	8%	**13%**	8%	2%	3%	33%
	9-12	2%	5%	**6%**	2%	3%	18%
	12-16	0%	1%	1%	**2%**	2%	6%
	> 16	0%	1%	1%	1%	**5%**	9%
	Total	30%	29%	19%	8%	13%	100%

Notes: Frequencies are based on the National Household Survey. The table reports the distribution of couples by the years of formal education attained by each spouse.

Table A.4: Earnings Assortative Mating by Income Decile, Couples Aged 25 and Older

		Women									
	Decile	1	2	3	4	5	6	7	8	9	10
Men	1	**2,1%**	1,5%	1,2%	1,1%	0,8%	0,7%	0,7%	0,5%	0,4%	0,4%
	2	1,6%	**1,4%**	1,4%	1,0%	0,9%	0,6%	0,4%	0,3%	0,3%	0,2%
	3	1,4%	1,0%	**1,3%**	1,2%	1,5%	0,8%	0,7%	0,6%	0,5%	0,3%
	4	1,3%	0,8%	1,2%	**1,4%**	1,1%	1,5%	0,8%	0,7%	0,4%	0,4%
	5	1,2%	1,0%	1,1%	1,2%	**1,1%**	1,2%	1,2%	1,0%	0,7%	0,4%
	6	0,9%	1,1%	1,1%	1,1%	1,2%	**1,3%**	1,3%	1,1%	0,8%	0,4%
	7	1,2%	1,0%	1,1%	1,0%	1,0%	1,2%	**1,2%**	1,2%	1,1%	0,6%
	8	1,0%	0,6%	0,8%	0,9%	1,0%	1,3%	1,1%	**1,2%**	1,7%	1,0%
	9	0,9%	0,6%	0,7%	0,7%	0,7%	0,9%	1,3%	1,7%	**1,8%**	1,8%
	10	0,6%	0,4%	0,5%	0,4%	0,5%	0,7%	1,0%	1,3%	1,9%	**3,9%**

Notes: Frequencies are based on the National Household Survey. The sample includes couples in which both women and men are aged 25 years or older. The table reports the joint distribution of couples across earnings deciles.

Table A.5: Summary Statistics of Three Samples by Age Group: Offspring Aged 20–29, 2012

	Distribution			Mean test vs. Panel A	
	Mean	Median	SD	Diff.	F (p-value)
Panel A: Parents/children sample					
Labour income	151,084	122,587	145,489	–	–
Fraction no labour income	4.2%	–	–	–	–
Fraction female	45.6%	–	–	–	–
Age	24.3	24.0	2.8	–	–
Observations		247,866		–	–
Panel B: Tax records					
Labour income	153,927	123,884	152,432	-2,843	267.97 (0.000)
Fraction no labour income	4.0%	–	–	0.2	105.34 (0.000)
Fraction female	45.14%	–	–	-0.45	78.81 (0.000)
Age	24.6	25.0	2.8	-0.3	11094.9 (0.000)
Observations		335,594			
Panel C: Household survey					
Labour income	206,342	176,465	151,372	-55,258	1084.73 (0.000)
Fraction no labour income	10.1%	–	–	0.318	309.28 (0.000)
Fraction female	44.49%	–	–	1.1	5.00 (0.025)
Age	24.8	25.0	2.78	-0.49	247.43 (0.000)
Observations		8,959			
Weighted observations		264,633			

Notes: The table reports descriptive statistics for individuals aged 20–29 in 2012 using three alternative data sources: the matched parents/children sample (Panel A), tax records (Panel B), and the household survey (Panel C). Differences and F-tests compare Panels B and C with Panel A. The household survey sample includes only individuals with positive total income in 2012. *Source:* Based on social security records (BPS), tax records (DGI), and household survey data (INE).

Table A.6: Summary Statistics of Three Samples: Parents Aged 45–65, 2012

	Distribution			Mean test vs. Panel A	
	Mean	Median	SD	Diff.	F (p-value)
Panel A: Parents/children sample					
Labour income	294,001	149,397	522,586	–	–
Fraction no labour income	24.9%	–	–	–	–
Fraction female	54.1%	–	–	–	–
Age	53.5	53.0	6.09	–	–
Observations		319,659			
Panel B: Tax records					
Labour income	265,834	128,845	495,319	28,167	2107.00 (0.000)
Fraction no labour income	30.5%	–	–	-5.57	9990.00 (0.000)
Fraction female	49.4%	–	–	4.63	5981.00 (0.000)
Age	54.2	54.0	6.28	-0.65	4508.00 (0.000)
Observations		583,061			
Panel C: Household survey					
Labour income	227,477	142,081	349,865	70,980	633.39 (0.000)
Fraction no labour income	40.1%	–	–	-15.95	1977.00 (0.000)
Fraction female	54.21%	–	–	-0.01	284.23 (0.000)
Age	54.2	54.0	6.30	-0.70	185.66 (0.000)
Observations		21,815			
Weighted observations		558,517			

Notes: The table reports descriptive statistics for individuals aged 45–65 in 2012 using three alternative data sources: the matched parents/children sample (Panel A), tax records (Panel B), and the household survey (Panel C). Differences and F-tests compare Panels B and C with Panel A. The household survey sample includes only individuals with positive total income in 2012. *Source:* Based on social security records (BPS), tax records (DGI), and household survey data (INE).

Table A.7: Analysis of Parent Groups by Whether Spouse's Formal Income Is Observed

	Children aged 20–29				Children aged 30–39			
Composition of parents	Obs.	(%)	IRA	SD	Obs.	(%)	IRA	SD
Fathers' income > mothers' income	61,450	63	0.150	0.004	17,006	62	0.310	0.009
Mothers' income > fathers' income	36,443	37	0.130	0.006	10,470	38	0.315	0.012
Both parents (total)	97,893	100	0.141	0.003	27,476	100	0.306	0.007
Only father's income observed	63,594	50	0.126	0.004	9,088	13	0.259	0.010
Only mother's income observed	64,771	50	0.147	0.004	62,413	87	0.292	0.004
One parent only (total)	128,365	100	0.137	0.003	71,501	100	0.289	0.004
Parental income used in the regressions								
Father's income	125,044	55	0.134	0.003	26,094	26	0.291	0.007
Mother's income	101,214	45	0.135	0.003	72,883	74	0.293	0.004
Total (baseline sample)	226,258	100	0.140	0.002	98,977	100	0.292	0.003

Notes: The table compares subsamples defined by whether formal earnings are observed for one or both parents. IRA denotes the coefficient from an OLS regression of the child's income rank on the parent's income rank. SD reports standard errors. The lower panel shows the parental income measure used in the baseline regressions.

Table A.8: Joint Distribution of Observed Income Years for Parents and Children

Unconditional distribution						
Children aged 20–29						
Parents aged 45–65	Years observed	0 years	1–4 years	5 years	5–8 years	All
	0 years	5%	8%	2%	6%	21%
	1–4 years	3%	9%	3%	6%	20%
	5 years	1%	2%	1%	2%	6%
	5–8 years	5%	22%	7%	18%	53%
	All	14%	42%	13%	32%	100%
Children aged 30–39						
Parents aged 45–65	Years observed	0 years	1–4 years	5 years	5–8 years	All
	0 years	6%	7%	2%	12%	26%
	1–4 years	3%	6%	2%	11%	22%
	5 years	1%	1%	1%	3%	6%
	5–8 years	4%	10%	4%	28%	46%
	All	13%	24%	9%	54%	100%

Notes: The table reports the joint distribution of the number of years with observed income for parents and children. Rows classify parents by observed income years, while columns classify children. Percentages sum to 100 within each panel.

Table A.9: Average IRA for Total and Labor Income: Longer Sample Period (8 Years), Children Aged 30–39 and 20–29

	Total income		Labor income	
	Children 30–39	Children 20–29	Children 30–39	Children 20–29
IRA	0.272***	0.180***	0.257***	0.182***
	(0.003)	(0.002)	(0.003)	(0.002)
Observations	165,826	351,317	165,826	351,317

Notes: The table reports coefficients from OLS regressions of the child's percentile rank on the parent's percentile rank. Controls include child sex and parent age and sex. Standard errors are reported in parentheses. All IRA estimates are statistically significant ($p < 0.001$).

Table A.10: Average IRA for Earnings by Parent and Child Gender (Children Aged 30–39)

	Fathers (maximum income)			Mothers (maximum income)		
	Average	Sons	Daughters	Average	Sons	Daughters
IRA	0.267***	0.277***	0.257***	0.260***	0.227***	0.292***
	(0.006)	(0.009)	(0.009)	(0.004)	(0.006)	(0.005)
Observations	26,094	13,517	12,577	72,883	36,811	36,072
Tests of equality of IRA estimates						
vs. mothers	0.896	20.786	11.725	–	–	–
p-value	0.344	0.000	0.001	–	–	–
vs. daughters	–	2.426	–	–	20.786	–
p-value	–	0.119	–	–	0.000	–

Notes: The table reports coefficients from OLS regressions of the child's income rank on the parent's income rank. Estimates are shown separately by parent and child gender. Standard errors are reported in parentheses. Test statistics for differences across selected estimates are reported in the lower panel.

Table A.11: Average IRA for Total Income by Parent and Child Gender (Children Aged 20–29)

	Fathers (maximum income)			Mothers (maximum income)		
	Average	Sons	Daughters	Average	Sons	Daughters
IRA	0.139***	0.141***	0.137***	0.142***	0.115***	0.170***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Observations	125,044	67,179	57,865	101,214	52,429	48,785
Tests of equality of IRA estimates						
vs. mothers	0.350	18.364	35.244	–	–	–
p-value	0.554	0.000	0.000	–	–	–
vs. daughters	–	0.550	–	–	78.985	–
p-value	–	0.458	–	–	0.000	–

Notes: The table reports coefficients from OLS regressions of the child's income rank on the parent's income rank. Estimates are shown separately by parent and child gender. Standard errors are reported in parentheses. Test statistics for differences across selected estimates are reported in the lower panel.

Table A.12: Average IRA for Earnings by Parent and Child Gender (Children Aged 20–29)

	Fathers (maximum income)			Mothers (maximum income)		
	Average	Sons	Daughters	Average	Sons	Daughters
IRA	0.134***	0.135***	0.132***	0.135***	0.109***	0.163***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Observations	125,044	67,179	57,865	101,214	52,429	48,785
Tests of equality of IRA estimates						
vs. mothers	0.023	20.849	31.473	–	–	–
p-value	0.880	0.000	0.000	–	–	–
vs. daughters	–	0.257	–	–	20.849	–
p-value	–	0.612	–	–	0.000	–

Notes: The table reports coefficients from OLS regressions of the child's income rank on the parent's income rank. Estimates are shown separately by parent and child gender. Standard errors are reported in parentheses. Test statistics for differences across selected estimates are reported in the lower panel.

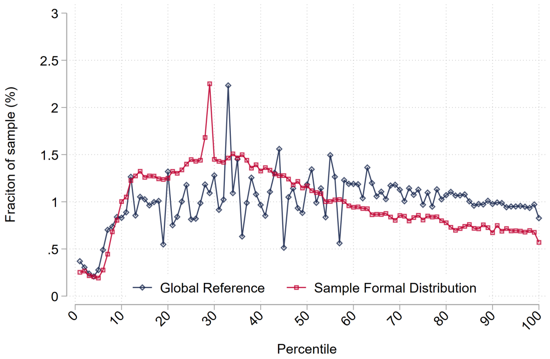
Table A.13: Average IRA for Total Income by Parent and Child Gender (Children Aged 30–39), Longer Sample Period (8 Years)

	Fathers (maximum income)			Mothers (maximum income)		
	Average	Sons	Daughters	Average	Sons	Daughters
IRA	0.270***	0.269***	0.270***	0.272***	0.213***	0.330***
	(0.006)	(0.009)	(0.008)	(0.003)	(0.005)	(0.004)
Observations	38,909	19,298	19,611	126,917	63,020	63,897
Tests of equality of IRA estimates						
vs. mothers	0.071	27.913	31.043	–	–	–
p-value	0.789	0.000	0.000	–	–	–
vs. daughters	–	0.006	–	–	260.147	–
p-value	–	0.941	–	–	0.000	–

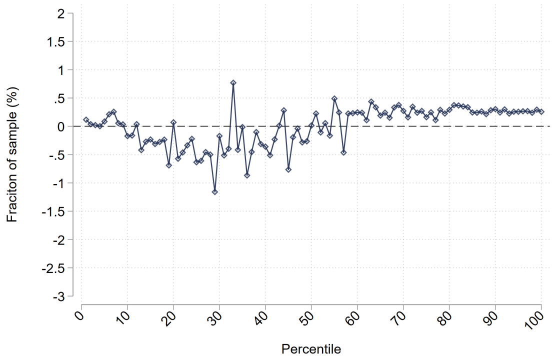
Notes: The table reports coefficients from OLS regressions of the child's income rank on the parent's income rank. Estimates are shown separately by parent and child gender. Standard errors are reported in parentheses. Test statistics for differences across selected estimates are reported in the lower panel. Results use an eight-year income averaging period.

Appendix: Additional Figures

Figure A.1: Fraction of individuals in each percentile of formal income distribution and of combined formal and informal income distribution, children aged 20-29



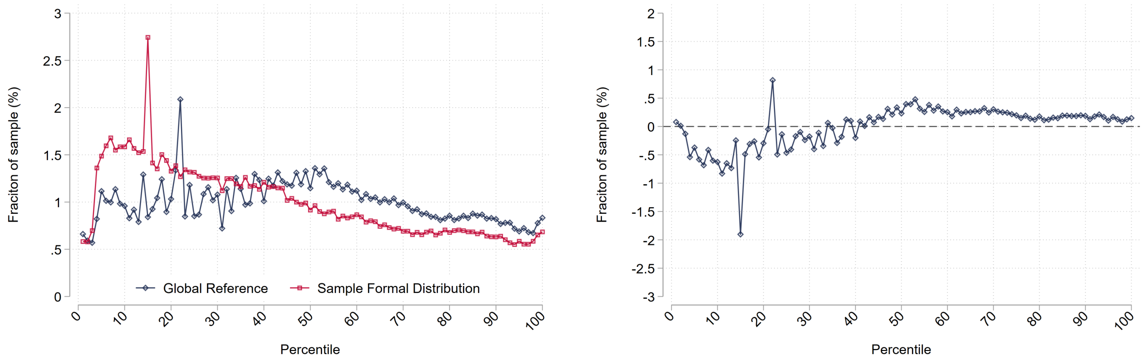
(a) Formal income distribution



(b) Difference

Notes: Panel (a) shows how our observations are distributed across the *global* reference distribution (blue) and the *sample* reference distribution (red). Panel (b) plots the difference between these distributions. The *global* reference distribution combines formal labor income from tax records and informal labor income from survey data, while the *sample* reference distribution includes only formal labor income from tax records. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

Figure A.2: Fraction of individuals in each percentile of formal income distribution and of combined formal and informal income distribution, parents aged 45-65



(a) Formal income distribution

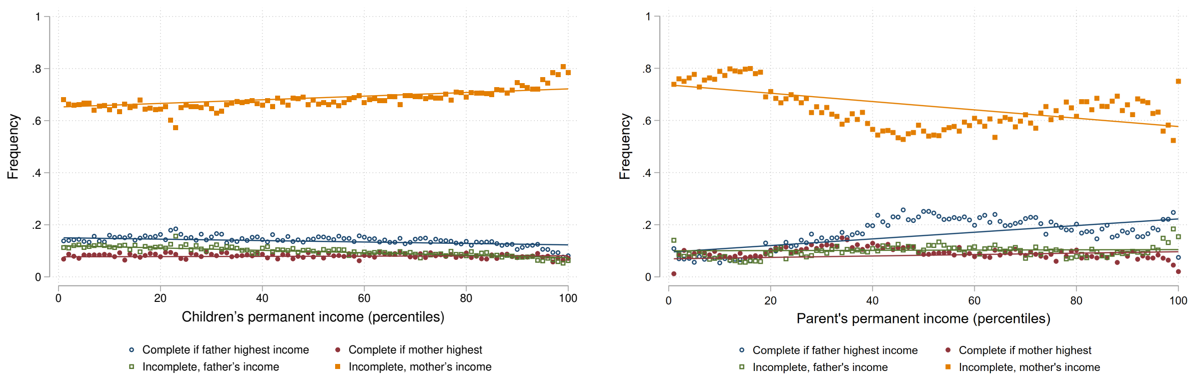
(b) Difference

Notes: Panel (a) shows how our observations are distributed across the *global* reference distribution (blue) and the *sample* reference distribution (red). Panel (b) plots the difference between these distributions. The *global* reference distribution combines formal labor income from tax records and informal labor income from survey data, while the *sample* reference distribution includes only formal labor income from tax records. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

Figure A.3: Distribution of children and parents by type of household. Offspring aged 30- 39

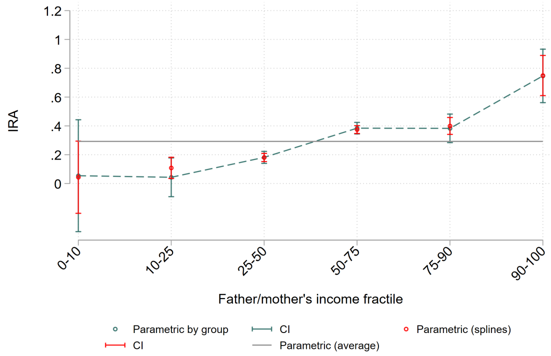
(a) Children

(b) Parents

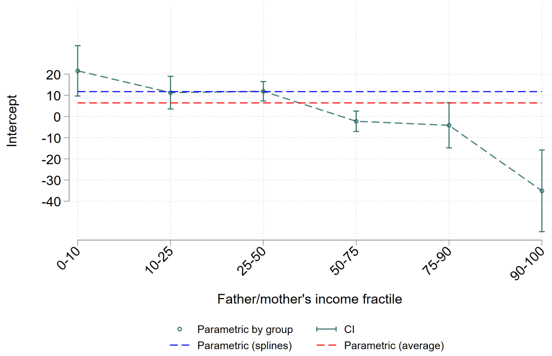


Notes: Panel (a) reports the distribution of offspring across household types by percentile of permanent income. Panel (b) reports the corresponding distribution for parents. Household types are defined according to whether formal income is observed for one or both parents, and whether the highest observed income corresponds to the father or the mother. Categories distinguish complete households (income observed for both parents) from incomplete households (income observed for only one parent). *Source:* Based on social security records (BPS) and tax records (DGI).

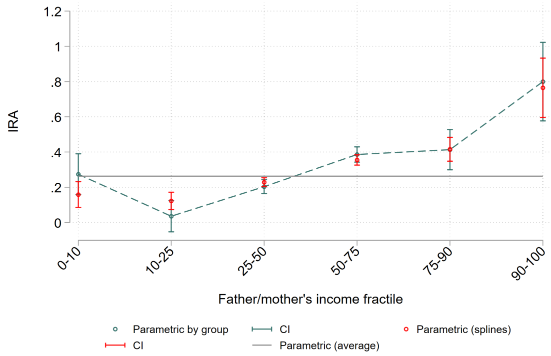
Figure A.4: Slope (IRA) and intercept coefficient estimates from non-linear regressions, children aged 30-39



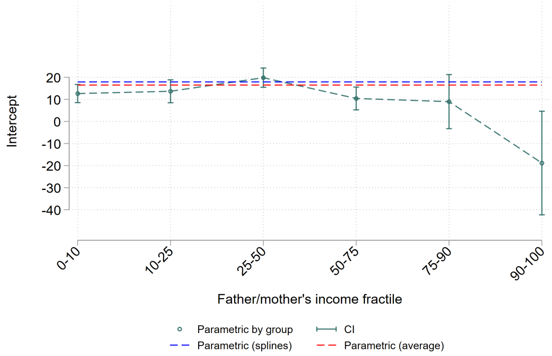
(a) IRA coefficients



(b) Intercepts



(c) IRA coefficients (Universal with zeros)



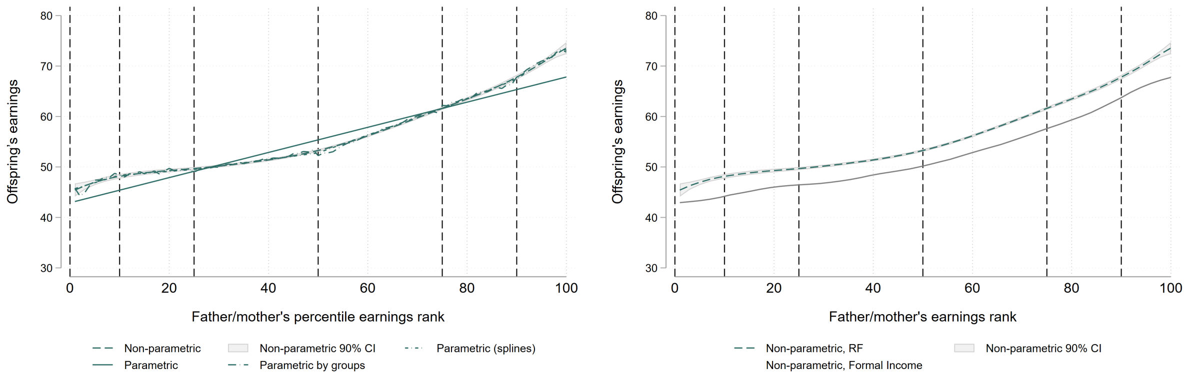
(d) Intercepts (Universal with zeros)

Notes: The figure reports slope (IRA) and intercept estimates from non-linear specifications across parental income fractiles for children aged 30–39. Panels (a) and (b) use the baseline sample, while Panels (c) and (d) use the universal sample including individuals with zero formal income. Panels (a) and (c) show estimated IRA coefficients, and Panels (b) and (d) show the corresponding intercepts. Estimates are obtained from regressions by income group and spline specifications. Vertical bars denote confidence intervals. The spline knots are located at percentiles P10, P25, P50, P75, and P90, which also define the six parental income groups used in the piecewise regressions. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

Figure A.5: Earnings Rank persistence across the earnings distribution. Children aged 30-39

(a) Different estimation methods

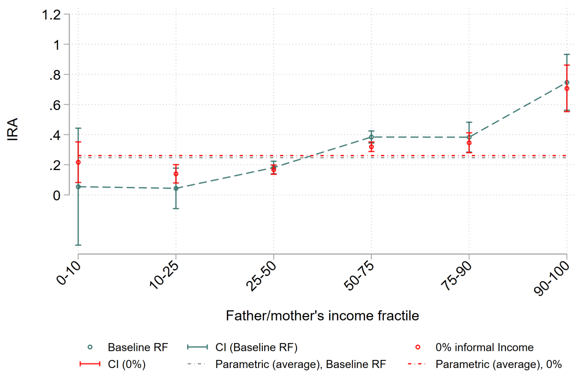
(b) Impact of imputing informal income



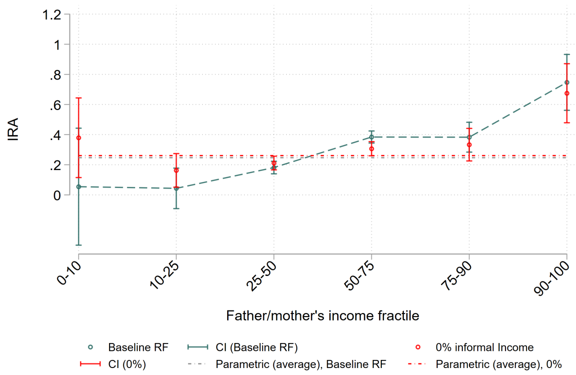
Notes: Panel (a) compares alternative estimation methods for the relationship between parental earnings rank and offspring earnings rank among children aged 30–39. We report non-parametric estimates, linear parametric estimates, estimates by percentile groups, and spline specifications. Panel (b) compares non-parametric estimates using only observed formal labor earnings with estimates that additionally incorporate imputed informal earnings using Random Forest methods. Shaded areas represent 90% confidence intervals. Vertical dashed lines indicate selected parental earnings rank percentiles. *Source:* Based on social security records (BPS), tax records (DGI), and survey data (ECH).

Figure A.6: Income Rank persistence across the income distribution: Impact of imputing informal income. Different estimation methods. Children aged 30-39

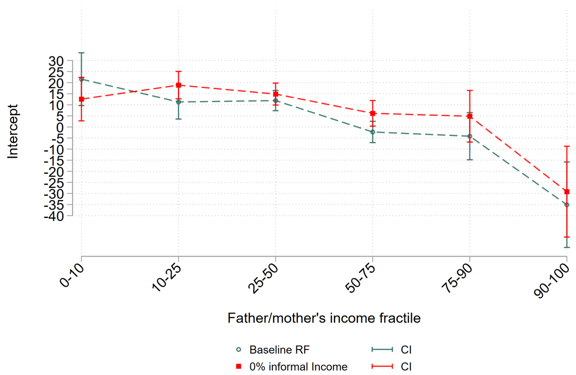
(a) Splines regression model



(b) Parametric by groups (beta)



(c) Parametric by groups (intercepts)



Notes: The figure compares alternative non-linear estimates of income rank persistence for children aged 30–39 under two treatments of informal income: the baseline specification with Random Forest imputation and a specification assigning zero informal income in periods without observed formal earnings. Panel (a) reports spline-based estimates of the IRA coefficient across parental income fractiles. Panel (b) reports group-specific parametric slope coefficients, and Panel (c) reports the corresponding intercepts. Vertical bars denote confidence intervals. The six parental income groups are defined by percentile cutoffs at P10, P25, P50, P75, and P90.
Source: Based on social security records (BPS), tax records (DGI), and survey data (ECH).