INTERGENERATIONAL INCOME MOBILITY IN FRANCE:

A COMPARATIVE AND GEOGRAPHIC ANALYSIS

GUSTAVE KENEDI
LOUIS SIRUGUE

WORKING PAPER N°2023/28

DECEMBER 2023

WORLD INEQUALITY LAB

Intergenerational Income Mobility in France: A Comparative and Geographic Analysis*

Gustave Kenedi — *Sciences Po*

Louis Sirugue — *Paris School of Economics*

This Version: August 2023

Abstract

We provide new estimates of intergenerational income mobility in France for children born in the 1970s using rich administrative data. Since parents' incomes are not observed, we employ a two-sample two-stage least squares estimation. We show, using the Panel Study of Income Dynamics, that this method slightly underestimates rank-based measures of intergenerational persistence. Our results suggest France is characterized by a strong persistence relative to other developed countries. 9.7% of children born to parents in the bottom 20% reach the top 20% in adulthood, four times less than children from the top 20%. We uncover substantial spatial variations in intergenerational mobility across departments, and a positive relationship between geographic mobility and intergenerational upward mobility. The expected income rank of individuals from the bottom of the parent income distribution who moved towards high-income departments is around the same as the expected income rank of individuals from the 75$^{\text{th}}$ percentile who stayed in their childhood department.

Keywords: intergenerational mobility, geographic mobility, spatial variations
JEL Codes: C18, J61, J62, R23

*Kenedi: Sciences Po, 28 rue des Saints-Pères, 75007 Paris France, and Paris School of Economics, 48 boulevard Jourdan, 75014 Paris France, gustavekenedi@gmail.com. Sirugue: Paris School of Economics, 48 boulevard Jourdan, 75014 Paris France, louis.sirugue@psemail.eu, EHESS, CNRS, CI-Migration, INED. This paper was previously circulated under the title "The Anatomy of Intergenerational Income Mobility in France and its Spatial Variations". This work has been funded by public grants overseen by the French National Research Agency as part of the "Investissements d'avenir" program (references ANR-17-EURE-001 and ANR-10-EQPX-17 - Centre d'accès sécurisé aux données - CASD), the ANR JOCE (reference ANR-18-CE41-0003-02), the French Collaborative Institute on Migration coordinated by the CNRS (reference ANR-17-CONV-0001), and PSE's Education policy and social mobility chair. This paper greatly benefited from feedback and suggestions from the editor, Nathan Hendren, three anonymous referees, as well as Brian Bell, Paul Brandily, Pierre Cahuc, Pascale Champalaune, Pierre-Philippe Combes, Clément Dherbécourt, Basile Grassi, Antoine Ferey, Sun Kyoung Lee, Bhash Mazumder, Thomas Piketty, Benoît Schmutz, Michaël Sicsic, Patrick Simon, Gary Solon, Yann Thommen, and Alain Trannoy, as well as participants at the 2021 Augustin Cournot Doctoral Days, the 2021 Congress of the French Economic Association, the 2022 European Meeting of the Urban Economics Association, and the 2022 European Meeting of the Econometric Society; and seminar participants at Sciences Po, Paris School of Economics, and CREST. Replication files available at https://github.com/gustavekenedi/intergenerational-mobility-france.

1 Introduction

To what extent is the income of individuals related to that of their parents? This question has seen renewed interest both in the general public and in academia as rising income inequality raised concerns about equality of opportunity. Examining this link is essential to understand whether children from different socio-economic backgrounds are afforded the same opportunities. It also matters for economic efficiency, as high persistence across generations may reflect an inefficient allocation of talents (so-called "Lost Einsteins"). Intergenerational persistence has now been estimated for a large number of countries, paving the way for insightful cross-country comparisons. Yet, much remains to be known for France, a country with relatively modest post-tax/transfers income inequality in international comparison and largely inexpensive higher education tuition fees.

The few existing studies for France only estimate the traditional intergenerational income elasticity (IGE), which captures the elasticity of child income with respect to parent income, and are based on small-sample surveys with self-reported incomes (Lefranc and Trannoy, 2005; Lefranc, 2018). Using a large sample combining census and tax returns data, we estimate two additional measures of intergenerational mobility: (i) the rank-rank correlation (RRC), increasingly prominent in the literature, which corresponds to the correlation between child and parent income percentile ranks, and (ii) transition matrices, which capture finer mobility patterns along the parent income distribution. While previous studies on France used self-reported labor earnings, we focus on household-level income measures. They provide a better depiction of one's economic resources and allow the inclusion of children raised by single mothers. Integrating these improvements from the "new" intergenerational mobility literature enables us to conduct a detailed international comparison to rank France relative to other advanced economies for which comparable estimates are available.

In addition, we investigate the spatial variations in intergenerational mobility across the 96 metropolitan French departments. Such subnational analyses, pioneered by Chetty et al. (2014), help shed light on the mechanisms that may underlie income persistence across generations. Importantly, they highlight that national level estimates provide an incomplete assessment of a country's intergenerational mobility. We make use of the panel dimension of our data to describe the geographic mobility patterns of individuals and study the relationship between geographic mobility and intergenerational mobility. We investigate the separate roles of moving to a higher-income department from that of climbing the income ladder within departments, conditional on parent income rank.

Our analysis is conducted on almost 65,000 children born between 1972 and 1981, and observed in the Permanent Demographic Sample (EDP). This rich administrative dataset allows us to implement the contributions discussed above and to convincingly address concerns related to lifecycle and attenuation bias (Haider and Solon, 2006; Black and Devereux, 2011; Nybom and Stuhler, 2017). Since parents' incomes are not observed, we use a two-sample two-stage least squares (TSTSLS) estimation which consists in predicting parents' incomes using other parents drawn from the same population but for whom income is observed (Björklund and Jäntti, 1997). This method has been employed previously to estimate the IGE in the French context (Lefranc and Trannoy, 2005; Lefranc, 2018) as well as in many other countries (Jerrim et al., 2016, Table A1).

While studies typically use education and/or occupation to predict parent income, we make use of the richness of our data to also include detailed demographic characteristics of parents (French nationality dummy, country of birth, household structure, and birth cohort), and characteristics of the municipality of residence (unemployment rate, share of single mothers, share of foreigners, population, and population density). Our results are largely insensitive to the set of predictors. Parent income is then defined as the average[1] of father and mother predicted mean pretax wage over ages 35-45, and child income as pretax household income averaged over the same age range between 2010 and 2016. These two income definitions represent the most comprehensive household-level income definitions available for either generation.

TSTSLS Validation Exercise. Using the United States' Panel Study of Income Dynamics (PSID), we find that TSTSLS slightly underestimates rank-based measures of intergenerational persistence relative to what would be obtained if parent income were observed (OLS). The downward bias relative to the OLS estimate for the RRC ranges from 11% when education is the only predictor, to around 3-5% once occupation is also included. Subnational TSTSLS estimates are also fairly close to their OLS counterparts, though they tend to deviate more when the number of observations is small. Our results highlight that in settings like ours, where parent income cannot be directly observed, rank-based measures of intergenerational mobility obtained with TSTSLS likely provide lower bounds that are reasonably close to the true estimates. These findings confirm those obtained in different settings and samples by Cortes-Orihuela et al. (2022) and Jacome et al. (2023). We find that this reasoning also applies to the transition matrix.

National Results. Our main finding is that France exhibits relatively strong intergenerational income persistence compared to other developed countries. Our baseline estimate of the intergenerational elasticity in household income is 0.527, suggesting that on average, a 10% increase in parent income is associated with a 5.27% increase in

[1]See Section 3.3 for an explanation for why we take the *average* rather than the *sum*.

child income. Put differently, if one's parents earn 10% more than the average of parents' incomes, then one is expected to preserve about 50% of that relative advantage. This estimate should be interpreted with caution considering our validation exercise suggests the TSTSLS IGE is significantly greater than the true estimate. Applying the correction factor we find, the IGE decreases to 0.396.

Moving to the rank-rank relationship, we find that the conditional expectation of child income percentile rank with respect to parent income percentile rank is linear throughout most of the parent income distribution, with steeper relationships at the tails. Our baseline estimate of the rank-rank correlation is 0.303, implying that a 10 percentile increase in parent income rank is associated, on average, with a 3.03 percentile increase in child income rank. This estimate is of similar magnitude to that found for Italy (0.3; Acciari et al. (2022)), somewhat smaller than for the United States (0.341; Chetty et al. (2014)), and markedly greater than existing estimates for other advanced economies such as Sweden (0.197; Heidrich (2017)), Australia (0.215; Deutscher and Mazumder (2020)) or Canada (0.242; Corak (2020)). Applying the correction factor we find in the validation exercise gives an RRC of 0.314 which does not affect France's relative position.

Intergenerational persistence, as captured by the transition matrix, is strongest at the tails of the parent income distribution: 9.7% of children from the bottom 20% of the parent income distribution reach the top 20% as adults. This probability is almost 4 times greater for children born to parents in the top 20% (38.4%). In comparison, the probability for a child born to a family in the bottom 20% to reach the top 20% in adulthood is 7.5% in the United States (Chetty et al., 2014) and 12.3% in Australia (Deutscher and Mazumder, 2020). Moreover, persistence at the top becomes stronger and stronger as we zoom in on the right tail of the parent income distribution. As with the RRC, the validation exercise suggests these estimates represent upper (lower) bounds on mobility (persistence).

We show that our baseline results are robust to potential biases. Foremost, we evaluate how sensitive they are to the parent income prediction specification. In particular, we check whether varying the set of predictors or using non-parametric estimation methods influences our estimates. IGE estimates are overinflated when using only education as a predictor, while the RRC and transition matrices remain surprisingly stable regardless of the set of predictors used. Slightly improved prediction from using flexible models does not quantitatively alter our estimates. Moreover, we assess our estimates' sensitivity to the lifecycle and attenuation biases by varying the ages at which child and parent incomes are measured as well as the number of parent income observations used. Our baseline results do not appear to under- nor over-estimate intergenerational mobility due to measuring child and/or parent incomes too early or too late in the lifecycle nor because of averaging incomes over too few years.

Subnational Results. We uncover substantial spatial variations in intergenerational mobility across departments, comparable to those observed across countries. We define individuals' location as their department of residence in 1990, when they are between 9 and 18 years old. Higher levels of mobility are typically found in the West of France, and lower levels in the North and South. While the IGEs range from 0.30 to 0.45 in departments in Brittany (West), they range from 0.42 to 0.70 in departments in Hauts-de-France (North). The distribution of department-level RRCs is tighter than that of IGEs, but displays very similar spatial patterns.

We also characterize departments' absolute upward mobility (AUM), defined as the expected income rank of children born to parents at the 25$^{\text{th}}$ percentile, which is obtained from the fitted values of the department-level rank-rank regression (Chetty et al., 2014). Absolute upward mobility ranges from the 36.8 in Pas-de-Calais (North) to 54.4 in Haute-Savoie (East). The Paris department stands out in terms of AUM (49.8) but exhibits around average intergenerational persistence levels in terms of IGE (0.51) and RRC (0.28). The cross-department correlation between the IGE and RRC is only 0.65, and -0.55 with AUM. This highlights the importance of using a variety of intergenerational mobility measures to characterize a country's income persistence across generations (Deutscher and Mazumder, forthcoming).

As a first step to understand the sources underlying these cross-department variations in intergenerational mobility, we undertake a simple correlational analysis. We find that absolute upward mobility exhibits much stronger relationships with department characteristics in general, than either the IGE or the RRC. This suggests that factors that affect absolute mobility might differ from those that affect relative mobility. The only characteristic consistently negatively correlated with intergenerational mobility is the unemployment rate. Intriguingly, we find no evidence of a within France "Great Gatsby Curve"[2] with respect to the IGE nor the RRC. This contrasts with findings from other countries (Acciari et al., 2022; Chetty et al., 2014; Corak, 2020).

Lastly, we conduct a descriptive analysis of the relationship between intergenerational income mobility and geographic mobility. We document important gains in expected income rank for movers, which are slightly decreasing in parent income rank. For children from families in the bottom decile, movers have an expected rank approximately 5.6 percentiles greater than stayers, while this difference is of roughly 4.4 percentiles for children from families in the top decile. These gains are partly attributable to movers locating in higher-income departments in adulthood relative to stayers, but also to movers reaching local ranks in their adulthood department that

[2]The "Great Gatsby Curve" refers to the positive correlation between intergenerational income persistence (defined by the IGE) and income inequality (defined by the Gini index) found across countries (Corak, 2013).

are further away from the rank of their parents in the childhood department. Destination departments are on average characterized by higher income levels than origin departments only at the tails of the parent income distribution. However, regardless of parent income rank, the absolute upward mobility gains associated with moving to a higher-income department appear to be large and increasing with average income in the destination department. All these findings combine self-selection and causal effects, and we leave the disentangling of these two channels for future research.

The rest of the article is organized as follows. Section 2 describes the intergenerational income mobility measures we estimate and the main sources of bias they are subject to. The data, the parent income prediction procedure and validation exercise, and the sample and variable definitions are presented in Section 3. Section 4 reports our baseline estimates at the national level, while Section 5 assesses their robustness to various sources of bias. In Section 6, we investigate the spatial variations in intergenerational income mobility, their correlation with local characteristics, and describe the relationship between geographic and intergenerational mobility. Section 7 concludes.

2 Measuring Intergenerational Mobility

Intergenerational income mobility can be characterized using a variety of statistics.[3] In this section we (i) describe the statistics we employ, and (ii) discuss the two major biases inherent to most intergenerational persistence estimators, namely lifecycle bias and attenuation bias.

2.1 Main Measures

Intergenerational persistence measures primarily aim to characterize the joint distribution of children and their parents' lifetime incomes with a parsimonious set of practical statistics. We summarize intergenerational persistence using the following statistics.

Intergenerational Income Elasticity (IGE). The traditional intergenerational income elasticity is obtained by regressing children's log lifetime income on their parents' log lifetime income. An IGE of 0.4 implies that a 10% increase in parent income is associated, on average, with a 4% increase in child income. Importantly, this estimator is sensitive to differences in inequality across generations. This can be seen in the following equation, where y_p and y_c are parent and child log lifetime incomes:

[3] See for example Corak (2020), where nine statistics of intergenerational mobility are put into perspective. More elaborate discussions on the properties of the different intergenerational mobility estimators can also be found in Black and Devereux (2011), Chetty et al. (2014), Nybom and Stuhler (2017), and Deutscher and Mazumder (forthcoming).

$$\text{IGE} = \frac{\text{Cov}(y_c, y_p)}{\text{Var}(y_p)} = \text{Corr}(y_c, y_p) \times \frac{\text{SD}(y_c)}{\text{SD}(y_p)}. \tag{1}$$

The empirical literature has highlighted that IGEs are particularly sensitive to life-cycle and attenuation biases, sample selection criteria, non-linearities along the parent income distribution, income definitions, and to the treatment of negative/zero incomes (Couch and Lillard, 1998; Chetty et al., 2014; Landersø and Heckman, 2017; Helsø, 2021).

Rank-Rank Correlation (RRC). The increasingly popular rank-rank correlation is obtained by regressing children's percentile rank in lifetime income on their parents' percentile rank in lifetime income. A RRC of 0.4 means that a 10 percentile increase in parent rank is associated, on average, with a 4 percentile increase in child rank. Unlike the IGE, the RRC is unaffected by inequality levels in either generation. This can be seen in the following equation, where p_p and p_c are parent and child percentile ranks in their respective lifetime income distributions:

$$\text{RRC} = \frac{\text{Cov}(p_c, p_p)}{\text{Var}(p_p)} = \text{Corr}(p_c, p_p) \times \frac{\text{SD}(p_c)}{\text{SD}(p_p)} = \text{Corr}(p_c, p_p). \tag{2}$$

Consequently, the greater the degree of inequality in the child generation relative to the parent generation, the greater the IGE relative to the RRC. In addition, the same RRC in two countries with large differences in inequality would hide that in one country the distance between ranks in monetary terms is actually much larger than in the other. The RRC owes its recent popularity to its robustness to specification variations, common biases, and treatment of negative/zero incomes (Dahl and DeLeire, 2008; Chetty et al., 2014; Nybom and Stuhler, 2017).

Transition Matrices. To get a finer picture, one can use transition matrices, which report the probability of ending up in a given quantile as an adult conditional on coming from a family in a given quantile. Typically, they are reported by quintile and are of particular interest to seize non-linearities in children mobility along the parent income distribution.

2.2 Main Sources of Bias

The vast majority of currently available data sources do not cover the whole lifetime of children's and/or parents' incomes, leading researchers to approximate lifetime income based on shorter time spans. This data limitation generates the following two fundamental biases, which we extensively investigate in Section 5.

Attenuation Bias. A direct implication of relying on a limited number of income observations to approximate parent lifetime income is the attenuation bias arising from classical measurement error (Solon, 1992; Zimmerman, 1992). This leads to downward-biased estimates of intergenerational persistence. Mazumder (2005, 2016) and Nybom and Stuhler (2017) find that the attenuation bias can be very large for the IGE but affects the RRC only mildly, while O'Neill et al. (2007) show that it affects most the corner elements of the transition matrix. The common solution to lessen this bias is to average parent income over as many years as possible.

Lifecycle Bias. The second common bias relates to the age at which child and parent incomes are observed (Grawe, 2006; Haider and Solon, 2006). In particular, lifecycle bias arises in the presence of heterogeneous age-income profiles, which is observed empirically as high lifetime income individuals tend to experience steeper earnings profiles than low lifetime income individuals. As such, observing child or parent incomes either too early or too late in the lifetime is likely to bias intergenerational persistence estimates. The IGE is particularly sensitive to lifecycle bias, especially if incomes are measured before age 35, while it affects the RRC only moderately so long as incomes are measured at least in the late 20s/early 30s. Just as for the attenuation bias, the corner elements of the transition matrix are most sensitive to lifecycle bias (Chetty et al., 2014; Nybom and Stuhler, 2016, 2017).

3 Data

We use data from the Permanent Demographic Sample (EDP), which combines several administrative data sources on individuals born on the first four days of October. We refer to individuals born on one of these days as *EDP individuals*. We describe below the most relevant details for each data source we use and provide additional technicalities in Appendix A.

Civil Registers. They contain information from birth certificates of EDP individuals and their children, including gender, date and place of birth, and parents' date and place of birth, nationality and occupation.

1990 Census. It contains socio-demographic information about EDP individuals and members of their household. Importantly, it reports parents' education level, occupation, and other demographic characteristics if EDP individuals live with their parents in 1990.

All Employee Panel. It gathers worker-year level information on all private (since

1967) and public (since 1988) sector employees in metropolitan France, except those in the agricultural sector. Prior to 2001, only individuals born on an even year are covered. Our results are robust to the late coverage of civil servants (see Appendix C.1.1).

Tax Returns. They provide tax information on incomes earned between 2010 and 2016 for individuals in dwellings where an EDP individual is known either from their income tax form or their main housing tax. Income variables are available both at the household level and at the individual level. An advantage of the information being gathered at the dwelling level is that household income is observed for all couples, regardless of whether they file their taxes jointly.

3.1 Parent Income Prediction

The measures of intergenerational mobility laid out in Section 2.1 cannot be estimated directly with our data since we do not observe parents' incomes. We therefore rely on the two-sample two-stage least squares (TSTSLS) strategy introduced by Björklund and Jäntti (1997), and previously used in the French context by Lefranc and Trannoy (2005) and Lefranc (2018), and in many other countries (Jerrim et al., 2016, Table A1). It consists in predicting individuals' parents' incomes from a sample of other parents whose incomes are observed using a set of common observed characteristics. We refer to these other parents as *synthetic* parents.

Let Z denote a set of characteristics observed both for parents and synthetic parents. Their log lifetime incomes y can be expressed as:

$$y_i = \beta Z_i + \varepsilon_i. \tag{3}$$

We estimate this first-stage equation by OLS on our sample of synthetic parents, and predict parents' log lifetime incomes using the resulting $\hat{\beta}$ as $\hat{y}_i = \hat{\beta} Z_i$. Z includes parents' (i) education (8 categories), (ii) 2-digit occupation (42 cat.; includes inactivity status), (iii) demographic characteristics (birth cohort, French nationality dummy, country of birth (6 cat.), and household structure (6 cat.)), and (iv) characteristics of the municipality of residence (unemployment rate, share of single mothers, share of foreigners, population, and population density). For the geographic analysis, we drop the municipality characteristics to ensure they do not spuriously drive any spatial patterns, though this has virtually no impact on the estimates. All characteristics are observed in the 1990 census. To reduce the potential for lifecycle and attenuation bias, synthetic parents' income is defined as average pretax wage between 35 and 45 with at least 2 income observations over this age range in the All Employee Panel. The model is estimated separately on synthetic mothers (adj. $R^2 = 0.37$) and fathers (adj. $R^2 = 0.36$). We extensively test the robustness of our baseline results to using more flexible models and to varying the set of first-stage regressors in Section 5.1.

Method Validity. To assess how reliable TSTSLS estimates are relative to their OLS counterparts (i.e., using *observed* parent income), we need a dataset that includes parents' actual incomes as well as predictors of parents' incomes. Since such a dataset does not exist for France, we follow Jerrim et al. (2016), Bloise et al. (2021) and Jacome et al. (2023), and conduct a validation exercise using the United States' Panel Study of Income Dynamics.[4] We describe this analysis in detail in Appendix B. We provide comparisons both at the national level and, due to sample size constraints, by Census Bureau regions (Northeast, Midwest, South, and West). Our sample and definition choices aim to be as close as possible to our main analysis setting while at the same time maximizing sample size.

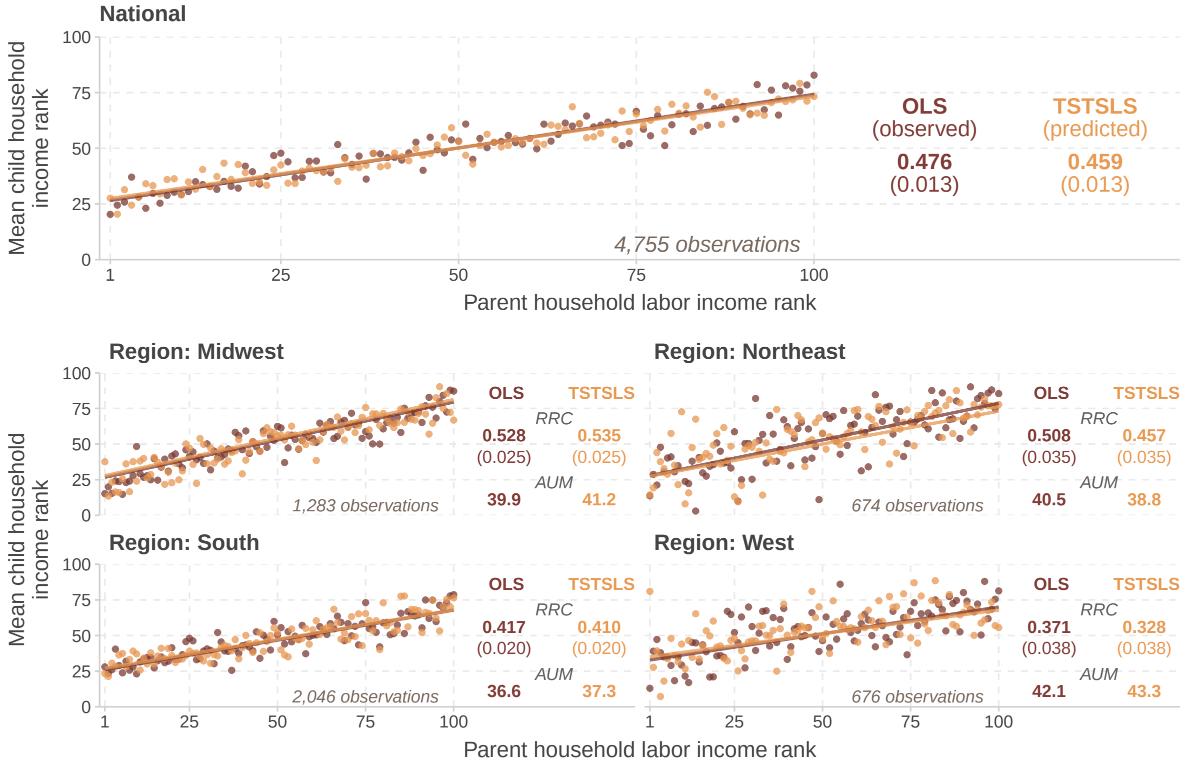
Specifically, our sample of children consists in individuals born between 1963 and 1988. We define parent income as the sum of father and mother mean labor income over ages 30-50, and child income as mean family total income over ages 30-50. The results are quantitatively similar when computing parent and child incomes over ages 35-45 as in the main analysis, despite the smaller sample size. We use education, 3-digit occupation (including inactivity status), birth year, race, and state of residence as first-stage predictors. These predictors are the closest we could find to those used in the main analysis.

Figure 1 presents the main results from our validation exercise. At the national level, the TSTSLS RRC estimate (0.459) is 4% smaller than the OLS estimate (0.476), a very moderate difference. Moreover, and importantly, the TSTSLS estimate of the RRC appears to understate persistence, i.e., they provide an upper bound for intergenerational mobility, as also found by Cortes-Orihuela et al. (2022) and Jacome et al. (2023). The same applies for estimates of the transition matrix presented in Appendix Figure B.3. At the Census Region level, the RRC obtained by TSTSLS are again reasonably similar to those obtained by OLS, with a slightly larger underestimation for the Northeast and West regions where the number of observations is smaller. The same applies to absolute upward mobility, defined as the expected rank of children from families at the 25th percentile.

The TSTSLS RRC estimate is smaller than the true OLS estimate likely because parents from the very top (bottom) of the income distribution can only be mispositioned downwards (upwards) when using predicted incomes. Assuming a monotonic relationship between parents and child income ranks, this mechanically flattens the rank-rank relationship and biases the rank-rank correlation downwards. This can be seen in Figure 2, which shows the conditional expectation of out-of-sample predicted labor

[4] Acciari et al. (2022) and Cortes-Orihuela et al. (2022) also conduct validation exercises of TSTSLS using administrative data from Italy and Chile respectively.

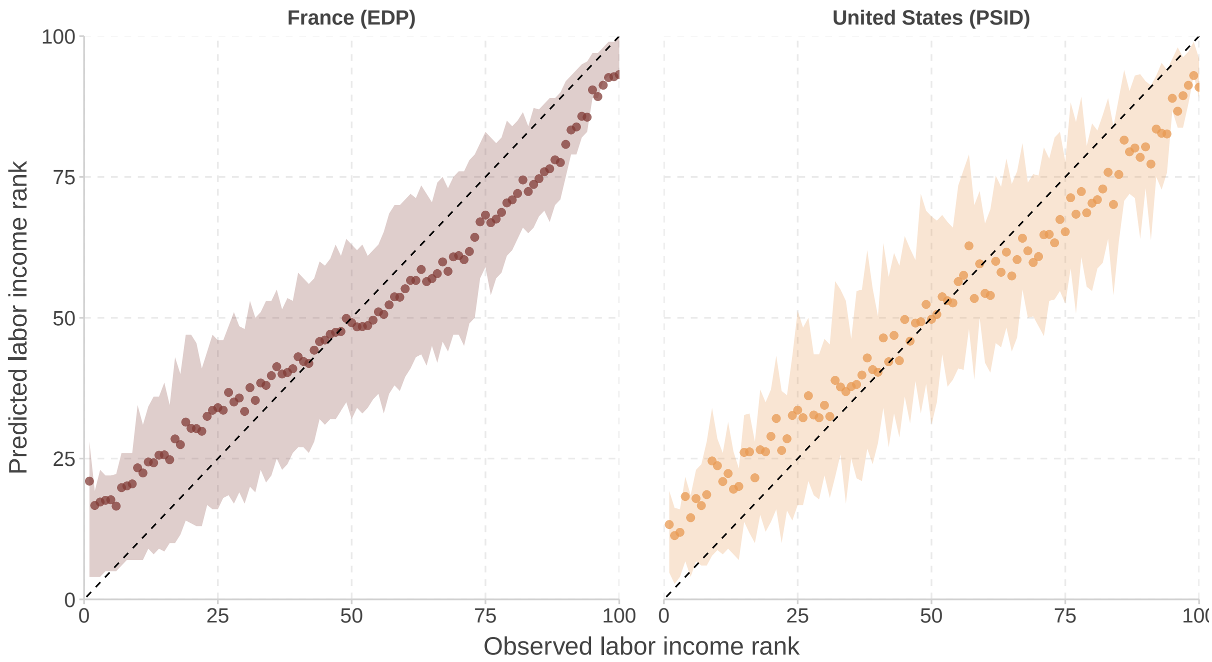
Figure 1: OLS vs. TSTSLS RRC - National and Census Regions in the United States



Notes: This figure presents rank-rank correlations obtained when parent income is observed (OLS) and when it is predicted using two-sample two-stage least squares (TSTSLS), at the national level and by Census Bureau Regions in the United States. They are computed on the Panel Study of Income Dynamics (PSID). The sample used is restricted to children born between 1963 and 1988 who are observed at least once as children in a family unit and at least once as a reference person or partner in a family unit over ages 30-50. Child income is the mean of family total income over ages 30-50. Parent income is the sum of father and mother (predicted) mean labor income over ages 30-50. For TSTSLS estimates, parent income is predicted separately for males and females using an OLS model including education (7 cat.; highest years of school completed), 3-digit occupation (334 cat.; most common occupation (incl. inactivity status) between 30 and 50 years old), parents' demographic characteristics in 1990 (birth cohort and race (5 cat.; most recent observation) and state fixed effects (most common state of residence between 30 and 50 years old). The fitted lines correspond to the regression line obtained on the microdata. We report coefficients and naive standard errors (in parenthesis) obtained from OLS regressions of child income rank on parent income rank with child cohort fixed effects, on the microdata for the full sample.

income rank with respected to observed labor income rank, as well as the interquartile range of the prediction. Indeed, percentile ranks tend to be overestimated at the bottom of the parents income distribution and underestimated at the top. We obtain very similar out-of-sample predictions in the EDP as in the PSID, suggesting we can reasonably apply the estimated TSTSLS biases of our validation exercise to the main analysis. Note that the IGE is sensitive to another bias because, all else equal, it is decreasing in the variance of parents' incomes (as highlighted in equation (1)). As such, since the distribution of predicted parent incomes is narrower than the true distribution, this puts an upwards pressure on the IGE.

Figure 2: Out-of-Sample Predicted Labor Income Rank - France (EDP) and United States (PSID)



Notes: This figure presents the conditional expectation of out-of-sample predicted labor income rank with respect to observed labor income rank, for both the PSID validation exercise (United States - PSID) and our own parent income prediction (France - EDP). See Figure 1's notes for details on data, sample and income definitions for the PSID analysis, and Figure 3's note for details on our analysis (EDP).

Inference. Since we are in a two-stage setting, standard inference is inappropriate. Inoue and Solon (2010) derive an analytical formula for TSTSLS standard errors. However, their method cannot be applied in our setting as we use a non-standard transformation of the first-stage outcome variables. Indeed, because labor income is observed for synthetic parents *individually* but is not observed for their spouse, we can only estimate equation (3) on *individual* income. We then aggregate mother and father predicted incomes to obtain a measure of *household* income, which we use as the regressor in the second stage rather than using the fitted values from the first stage as is. We thus report bootstrapped standard errors for all individual-level regressions, which, for the same reason, cannot be clustered at the family level. Specifically, we draw one bootstrap sample for synthetic fathers and one for synthetic mothers separately. We then run the first-stage regression, and predict parent income on a bootstrap sample of children. We iterate this process 1,000 times. These bootstrapped standard errors are of the same order of magnitude though slightly larger than naive ones.

3.2 Sample Definitions

Hereinafter we rely on the Permanent Demographic Sample (EDP) to estimate intergenerational persistence in France. Our samples of interest are defined as follows.

Sample of Children. It consists of EDP individuals who are (i) born between 1972 and 1981 in metropolitan France,[5] (ii) observed with their parents in the 1990 census, (iii) whose parents are neither farmers nor in a liberal profession[6], and (iv) observed in the tax returns data at least once between 35 and 45 years old.[7] Restriction (i) is made to observe individuals with their parents in the 1990 census[8] and to have a reasonably large sample size for the subnational analysis. Restriction (ii) enables us to retrieve their parents' characteristics, and (iii) is due to the fact that farmers and liberal professions are not covered by the All Employee Panel from which we obtain synthetic parent income. Restriction (iv) aims to minimize lifecycle bias. The final sample contains 64,571 children.[9] Overall, they have very similar socio-economic characteristics as the representative sample of EDP individuals satisfying only restriction (i), except for under-representing children of farmers by definition, as shown in Appendix Section C.1.2.

Sample of Synthetic Parents. It is constructed such that synthetic parents come from the same overarching population as actual parents. It therefore consists of EDP individuals who (i) had at least one child born between 1972 and 1981 in metropolitan France, (ii) are observed in the 1990 census, (iii) are neither farmers nor in a liberal profession in 1990, and (iv) have at least two pretax wage observations between 35 and 45 years old in the All Employee Panel.[10] As such our sample excludes individuals born in an odd year since they were not covered by the All Employee Panel prior to 2001.

[5] Metropolitan France refers to the part of France that is geographically in Europe.

[6] Liberal professions encompass activities that are not salaried, agricultural, commercial or artisanal, and carried out by self-employed service providers (e.g., lawyers, notaries, private doctors, etc.). 5.08% of EDP individuals satisfying (i) and (ii) have at least one parent who is a farmer and 2.41% have at least one parent who is in a liberal profession. As raised by Lefranc (2018), the fact that farmers tend to face relatively low incomes and a strong occupational inheritance (Lefranc et al., 2009) makes the exclusion of farmers likely to bias intergenerational persistence downwards.

[7] 6.73% of EDP individuals satisfying (i) and (ii) are not observed in the tax returns data between 35 and 45 years old.

[8] See Appendix Figure E.1 for the position in the family in the 1990 census by child birth cohort.

[9] See Appendix Table F.1 for the sample size at each additional restriction. Parent income cannot be predicted for 23 children because one of their parents has an occupation not represented in the sample of synthetic parents of the corresponding gender, hence the very slight discrepancy with this table.

[10] In Appendix Table F.2 we compare average characteristics of parents and synthetic parents. To ensure appropriate comparability of the two samples, no restriction on wage observations for synthetic parents or children is applied. Average characteristics are remarkably similar for most variables, even for 2-digit occupation (Appendix Table F.3), which confirms the assumption that actual and synthetic parents are random subsets of the same population.

The final sample contains 31,423 synthetic parents.[11]

Descriptive Statistics. Appendix Table F.6 provides statistics on our sample of synthetic parents and children. On average, fathers are around 42 in 1990 and mothers 39. This assures that we predict income based on observable characteristics measured sufficiently late in their lifecycle.

3.3 Variable Definitions

The variables we use are constructed as follows. All incomes are expressed in 2015 euros, and are measured before taxes but after the deduction of employer- and employee-level payroll taxes.

Parent Income. We define the income of one parent as predicted average pretax wage over ages 35 to 45. This income is predicted according to the methodology described in Section 3.1. We then compute income at the household level (regardless of marital status) by taking the average of father and mother predicted incomes if the child is observed with both parents in the 1990 census, and income of the only parent otherwise. We take the *average* of father and mother predicted incomes rather than the *sum* (the standard in the literature), to correct for the fact that otherwise single-headed households would be over-represented in the bottom of the income distribution (when using the sum, there are virtually no single-headed households above rank 50). Indeed, while in other studies parent income is typically observed repeatedly over several years, in our setting a parent observed as single in 1990 can by definition only be predicted their *individual* income for their entire lifetime even if their marital status actually changes later on. We refer to this income definition as parent household wage and use it as our main parent income measure. We also report results using father predicted income, which we refer to as father wage.

Child Income. Our main measure of child income, computed from the tax returns, corresponds to the sum of labor earnings (wages and self-employment income), taxable and imputed non-taxable capital income[12], unemployment insurance, retirement, and alimony, at the household level.[13] Just as for parents, a household is defined as

[11] See Appendix Table F.4 for the sample size at each additional restriction.

[12] Financial incomes not subject to any tax reporting are predicted by the French National Institute of Statistics and Economic Studies (INSEE) from a model estimated on the *Enquête Patrimoine*. In particular, they predict capital income for seven financial products (various tax-exempt savings accounts and life insurance) using household-level observed characteristics (income, age, family situation, ...). Excluding this income source from our child income definition does not affect the results.

[13] Social benefits such as family allowances, social minima (e.g., RSA, disability benefits) and housing benefits are not included in our main measure of child income.

individuals living in the same dwelling. To mitigate the potential for lifecycle bias, we average over 2010-2016 only for incomes declared when the individual is between 35 and 45 years old. We refer to this income definition as household income and use it as our main child income measure. We also report results using the following alternative child income definitions: (i) household wage, which is equivalent to the parent household wage definition, (ii) individual income, which we define as the sum of all individual-level incomes: labor earnings (wages and self-employment income), unemployment benefits, retirement, and alimony, and (iii) individual wage.

Income Definition Discussion. Our preferred parent and child income definitions represent the most comprehensive household-level income definitions possible for either generation. Defining incomes at the household level is important in order to (i) better capture the economic conditions of individuals and their parents, (ii) allow the inclusion of children raised by single mothers, and (iii) enable the analysis of daughters, whose labor incomes alone may not be an appropriate measure of their economic outcomes. These income definitions are not identical but the results are qualitatively similar when using the same income definition, household wage, for both children and parents.

Percentile Ranks. We rank children within their birth cohort, and parents relative to other parents with children in the same birth cohort. To avoid individuals (in a given cohort) earning the same income (e.g., 0, or the minimum wage) being assigned different income ranks, we define the income rank of such individuals as the ceiling of the median income rank of individuals with that income level.[14]

4 Results at the National Level

We start by analyzing intergenerational mobility at the national level. For our baseline results, we use data on children born on the first four days of October between 1972 and 1981 and measure parent income as household-level predicted average annual pretax wage over ages 35-45, and child income as pretax household income averaged over the same age range between 2010 and 2016. We include child birth cohort fixed effects in the log-log and rank-rank regressions.[15]

[14] For example, if there are 3.65% of children with zero income, their median rank is 2, and thus they are assigned a rank of 2. In our samples, 0.06% of children have negative or zero household income (see Appendix Table F.5), while no parent has negative or zero predicted wage.

[15] In practice, these fixed effects have virtually no influence on the coefficients of interest.

4.1 Intergenerational Income Elasticity (IGE)

Figure 3 panel A displays the conditional expectation of log child income with respect to log parent income. Children with negative or zero incomes are excluded. This is of minor importance when defining child income as household income as such cases are exceedingly rare. Nonetheless, we assess the influence of zero incomes in Appendix Figure C.9. The log-log CEF is pretty linear throughout the middle 80% of the parent income distribution, with some mild non-linearities at the tails.[16] This S-shaped relationship is also observed in the United States (e.g., Chetty et al. (2014)), Denmark (e.g., Helsø (2021)) or Sweden (e.g., Björklund et al. (2012)). It implies that the elasticity is not constant over the whole parent income distribution, with smaller magnitudes at the tails, and is sensitive to the inclusion or exclusion of parents at the tails of their income distribution.[17]

Our baseline IGE estimate is 0.527, meaning that a 10% increase in parent income is associated, on average, with a roughly 5% increase in child income. This estimate should be interpreted with caution as our validation exercise presented in Section 3.1 suggests TSTSLS estimates of the IGE can be quite inflated relative to the true value. Thus this baseline IGE is not well-suited for international comparisons. Appendix Figure E.2 shows our estimates of the intergenerational income elasticity for every child and parent income definition, and for sons and daughters separately. Our father-son wage IGE estimate is relatively similar to existing ones for France despite important differences in methodology and data (see Appendix Table F.7). Intergenerational persistence estimates are larger for household income than for individual income or wage, which could be the result of assortative mating. IGEs are very similar when defining parent income as father wage, despite the fact that by construction, estimates based on father wage exclude children only observed with their mother in the 1990 census (about 10% of observations). The IGE is significantly lower for sons (0.478) than for daughters (0.577). This phenomenon is not systematic across countries, but is also observed in Germany (Bratberg et al., 2017) and the Netherlands (Carmichael et al., 2020), for instance.

4.2 Rank-Rank Correlation (RRC)

Figure 3 panel B plots the conditional expectation of child income rank with respect to parent income rank. It is relatively linear, with slight non-linearities at the tails as observed in many countries (Chetty et al., 2014; Bratberg et al., 2017; Helsø, 2021).

[16] Appendix Figure C.2 shows that these non-linearities are not driven by the set of first-stage predictors.

[17] Appendix Figures C.10a and C.10c show how trimming the top and bottom of the parent/child income distribution influences our estimates.

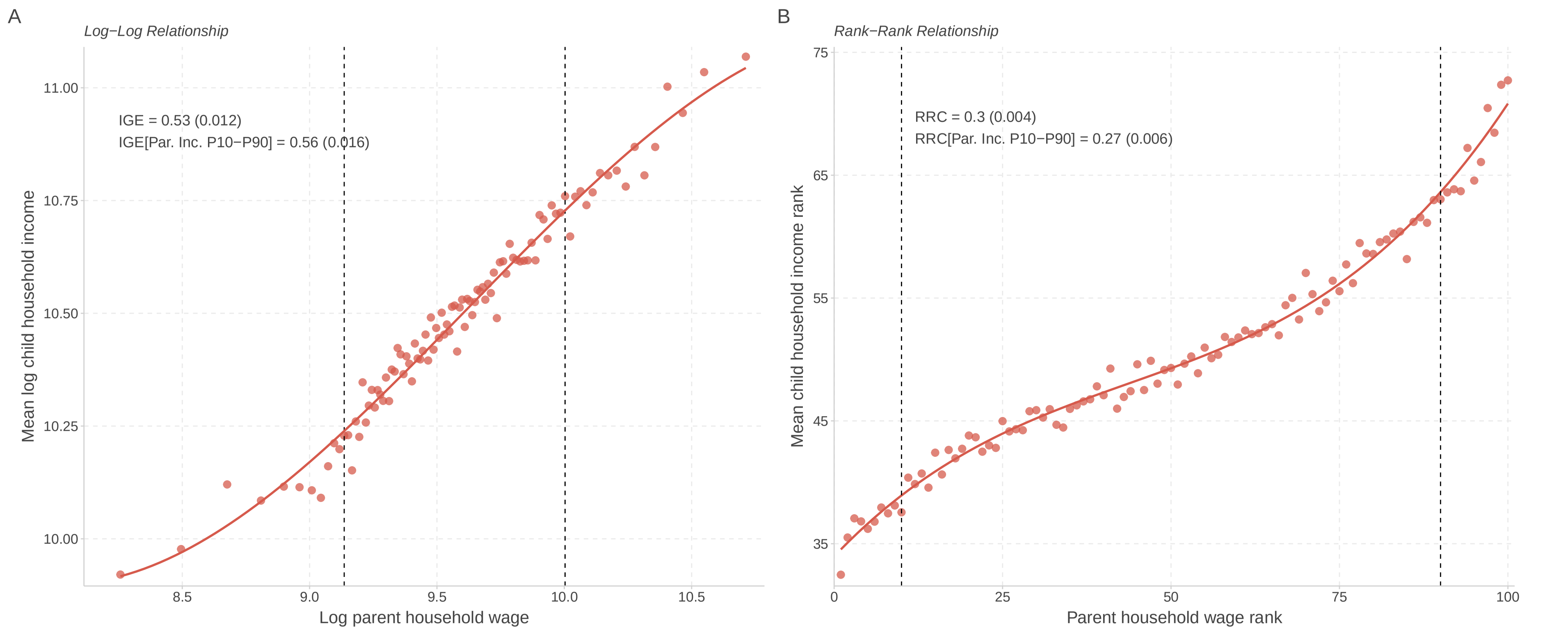


Figure 3: Conditional Expectation Functions for Log-Log and Rank-Rank Relationships in France

Notes: This figure presents non-parametric binned scatter plots of the relationship between log child income and log parent income (panel A), and child income rank and parent income rank (panel B) in France. It is computed on the Permanent Demographic Sample, a dataset of individuals born on the first four days of October. The sample used is restricted to children born between 1972 and 1981. Child income is the mean of 2010–2016 household income (with age restricted to 35-45). Parent income is the sum of each parent predicted wage divided by the number of parents. Parent income is predicted separately for males and females using an OLS model including parents' education (8 cat.), 2-digit occupation (42 cat.), demographic characteristics in 1990 (birth cohort, French nationality dummy, country of birth (6 categories), and household structure (6 cat.)) and characteristics of the municipality they lived in in 1990 (unemployment rate, share of single mothers, share of foreigners, population, and population density). These municipality characteristics are excluded for the geographic analysis. It is estimated on a sample of synthetic parents whose average wage at ages 35-45 (at least 2 income observations) is used as the dependent variable. Incomes are in 2015 euros. To construct panel A, children with negative or zero incomes are excluded (0.06% of the sample) and we bin parent incomes into 100 equal-sized bins and plot mean log child income versus mean log parent income within each bin. To construct panel B, children are ranked relative to other children in the same birth cohort while parents are ranked relative to other parents with children in the same birth cohort. We then plot mean child income rank versus parent income rank. The dashed lines represent the 10^{th} and 90^{th} percentiles of parents' income. We report coefficients and bootstrapped standard errors (in parenthesis) obtained from OLS regressions of log child income on log parent income (panel A) and child income rank on parent income rank (panel B), both with child cohort fixed effects, on the microdata for the full sample and for parents between the 10^{th} and 90^{th} percentiles. The fitted line is a 3^{rd} order polynomial fit through the conditional expectations.

Our baseline estimate of the rank-rank correlation is 0.303, meaning that a 10 percentile increase in parent income rank is associated, on average, with a 3.03 percentile increase in child income rank. Note that this estimate corresponds to a lower bound, as the validation exercise suggests the TSTSLS methodology slightly underestimates the RRC. Applying the estimated correction factor of 3.7% leads to a corrected baseline RRC coefficient of 0.314. Appendix Figure E.3 shows our baseline estimates of the rank-rank correlation for every child and parent income definition, and for sons and daughters separately. The estimates are slightly higher for daughters (0.310) than for sons (0.296), and are also slightly higher when defining parent income as household wage rather than as father wage. The estimates are smaller when defining child income as household wage or individual income and smallest when using individual wage, a pattern observed in other countries (Chetty et al., 2014; Deutscher and Mazumder, 2020; Landersø and Heckman, 2017), again possibly due to assortative mating.

To the best of our knowledge, this is the first time the RRC is estimated for France.[18] In Table 1 we compare RRC estimates for countries for which estimates exist (see Appendix Figure E.5 for a visual representation). To enable comparability we only keep studies which pool sons and daughters together, define parent income at the household level and use comprehensive income definitions. Note that for child income some studies only observe *individual* rather than *household* income which might result in lower RRC estimates (as we find for France, and Chetty et al. (2014) for the United States). Even though they are not directly comparable due to important differences in data and sample selection rules, we believe that it is a relevant exercise given the relative stability of the RRC to specification variations and common data limitations.

This international comparison suggests that (i) France exhibits strong persistence across generations in international comparison, given that it is the country with the second highest available RRC estimate behind the United States, and (ii) there is less variation across countries in the rank-rank slope than in the intergenerational elasticity, which is coherent with the fact that the RRC is not influenced by changes in inequality across generations, and is less sensitive to sample restrictions.

4.3 Transition Matrices

The last measure of intergenerational income persistence we estimate is a quintile-by-quintile transition matrix, which documents the conditional probabilities of be-

[18] A recent report (in French) by Abbas and Sicsic (2022) now also provides rank-based intergenerational mobility estimates for France. They use the same data as us and their sample consists in individuals born in 1990 (i) who are still claimed as dependent in their parents' tax return at age 20, (ii) whose parents' income can be observed around age 50, and (iii) whose individual income is observed at age 28 in their own tax return. They compare their results to ours and despite different sample definitions, when using the same income definition and measuring child income at the same age (i.e., 28), they find very similar results.

Country	RRC ↓	# obs.	Data	Child Income Definition[1]	Child Cohort	Child Age or Year at Income Measurement	Parent Age or Year at Income Measurement	Source
Switzerland	0.14	667,047	Social Security Earnings Records	Average total pretax *individual* income	1967-1982	32-34	when child between 15-20	Kalambaden and Martınez (2021, Table 3)
Switzerland	0.14	923,262	Social Security Earnings Records	Average total pretax *individual* income	1967-1984	30-33	when child between 15-20	Chuard-Keller and Grassi (2021, Figure 1)
Spain	0.195	1,492,107	Atlas de Oportunidades	Total pretax *individual* income	1980-1986	2016	1998	Soria Espín (2022, Figure 1)
Sweden	0.197	778,484	SIMSAM database[2]	Average total pretax *individual* income	1968-1976	32-34	34-50	Heidrich (2017, Table 2)
Denmark	0.203	157,543	Danish register data	Average total pretax *family* income	1980-1982	2011-2012	1996-2000	Helsø (2021, Table 1)
Australia	0.215	1,025,800	Federal income tax returns	Average total pretax *family* income	1978-1982	2011-2015	1991-2001	Deutscher and Mazumder (2020, Table 2)
Sweden	0.215	252,745	35% random sample from admin. data	Average total pretax *household* income	1957-1964	1996-2007[3]	1978-1980	Bratberg et al. (2017, Table 3)
Norway	0.223	324,870	Full population admin. data	Average pretax *family* earnings	1957-1964	1996-2006	1978-1980	Bratberg et al. (2017, Table 3)
Canada	0.242	2,115,150	Intergenerational Income Data	Average total pretax *family* income	1963-1970	2004-2008	when child between 15-19	Corak (2020, Table 5)
Germany	0.245	1,128	German Socio-Economic Panel	Average total pretax *household* income	1957-1976	2001-2012	1984-1986	Bratberg et al. (2017, Table 3)
Denmark	0.253	$\approx$ 410,000	Danish register data	Average total pretax *individual* income	1973-1979	2010-2012	when child between 7-15	Landersø and Heckman (2017, Table A6)
Denmark	0.257	205,625	Full population admin. data	Average total pretax *individual* income	1973-1975	2010-2012	when child between 7-15	Eriksen (2018, Table 3.2)
Italy	0.30[4]	1,719,483	Electronic database of Personal Income returns	Average total pretax *individual* income	1979-1983	2016-2018	1998-2000	Acciari et al. (2022, p.145)
France	0.303[5]	64,571	Permanent Demographic Sample	Parents: (predicted) *household* wage; Children: average total pretax *household* income	1972-1981	2010-2016 (between 35-45)	35-45	
United States	0.341	9,867,736	Federal income tax records, 1996-2012	Average total pretax *family* income	1980-1982	2011-2012	1996-2000	Chetty et al. (2014, Table 1)
United States	0.395	6,414	NLSY79	Average total pretax *family* income (self-reported)	1957-1964	1996-2008	1978-1980	Bratberg et al. (2017, Table 3)

Notes:
[1] The parent income definition is always at the family level.
[2] Swedish Initiative for Research on Microdata in the Social and Medical Sciences.
[3] Only even years.
[4] This estimate corresponds to the one when adjusting for lifecycle bias, incomplete coverage of taxpayers and tax evasion as reported on p.28. The baseline RRC estimate reported in Table 3 is 0.22.
[5] Assuming that the bias induced by the TSTSLS methodology is the same in France as in the United States, our validation exercise performed on the PSID suggests the OLS counterpart to our baseline estimate would equal to $0.303 \times \frac{0.476}{0.459} = 0.314$ (see Figure 1).

Table 1: Rank-Rank Correlation in International Comparison

ing in each income quintile as an adult given any parent income quintile. Figure 4 presents our baseline estimates of the transition matrix for France, along with available estimates for the United States (Chetty et al., 2014) and Australia (Deutscher and Mazumder, 2020). To the best of our knowledge, this is the first time transition matrices are estimated for France.[19]

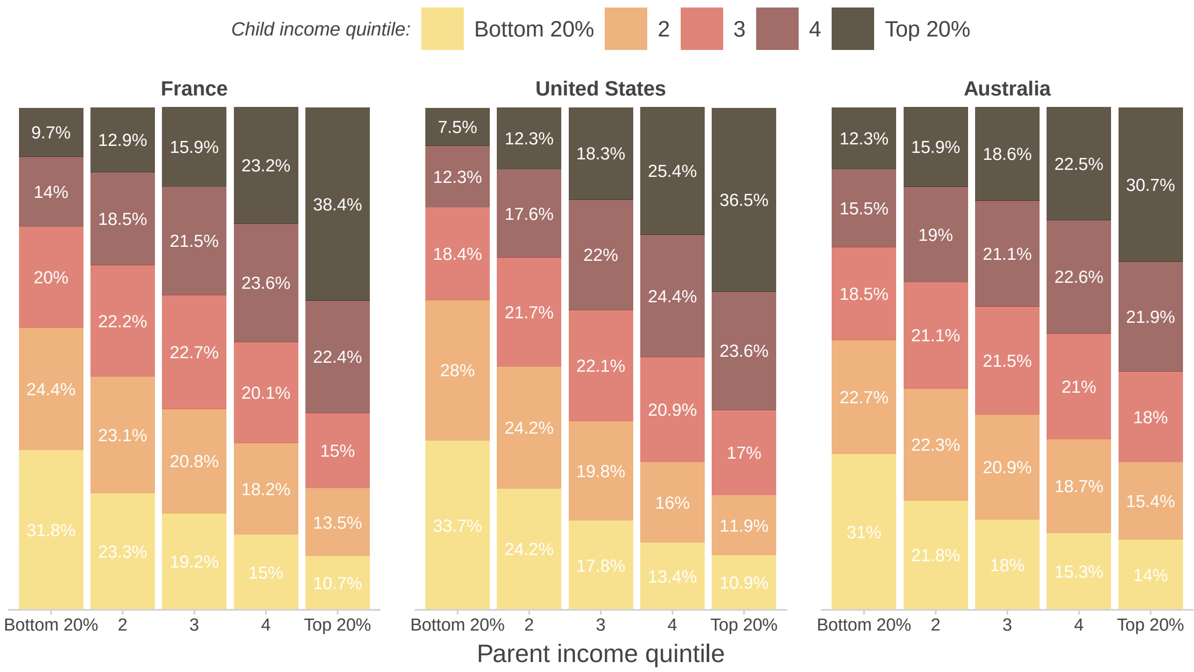


Figure 4: Baseline Quintile Transition Matrix for Different Countries

Notes: The first panel of this figure presents our baseline intergenerational transition matrix estimates. Bootstrapped standard errors are presented in Appendix Figure E.4. See Figure 3's notes for details on data, sample and income definitions. Each cell documents the share of children belonging to the quintile indicated by the color legend among children born to parents whose income falls in the quintile indicated on the x-axis. We present these estimates along with those put forward by Chetty et al. (2014) for the United States (second panel) and Deutscher and Mazumder (2020) for Australia (third panel). While we rely on at most 11 income observations (7 on average) for parents and at most 7 income observations (5 on average) for children, Deutscher and Mazumder (2020) use 11 income observations for parents and 5 for children, and Chetty et al. (2014) use 5 income observations for parents and 2 for children.

We find that 9.7% of children born to parents in the bottom 20% reach the top 20% in their forties. This share is 7.5% in the United States and 12.3% in Australia. In comparison, 31.8% remain in the bottom 20% of the income distribution. Regarding children born to the top 20%, 38.4% remain at the top, while only 10.7% move down to the bottom of the income distribution, much less than in Australia (14%). As a reference point, in a society where an individual's income is completely independent of parent income, the probability of being in any quintile given a parent quintile would by

[19] Alesina et al. (2018) estimated father-son wage transition probabilities from the bottom quintile only, using the TSTSLS methodology and data from the *Formation et Qualification Professionnelle* survey for earlier cohorts (1963-1973).

Country	P(Child Top 20% | Parent Bot. 20%) ↓	P(Child Bot. 20% | Parent Bot. 20%)	P(Child Top 20% | Parent Top 20%)	Source
United States	7.5%	33.7%	36.5%	Chetty et al. (2014, Table 2)
Italy[1]	8.6%[2]	36.7%	27.8%	Acciari et al. (2022)
France	9.7%	31.8%	38.4%	
Denmark	10.7%	30.7%	34.8%	Eriksen (2018, Figure 3.3*)
Netherlands	11.3%	29.8%	33.1%	Carmichael et al. (2020, Table 1*)
Canada	11.4%	30.1%	32.3%	Corak (2020, Table 6)
Switzerland	11.9%	23.7%	30.3%	Chuard-Keller and Grassi (2021, Table 2)
Spain	12.3%	25.3%	33.3%	Soria Espín (2022, Table A.5)
Australia	12.3%	31%	30.7%	Deutscher and Mazumder (2020, Table 3)
Switzerland	12.8%	24.5%	28.8%	Kalambaden and Martınez (2021, Table 5)
Sweden[3]	15.7%	26.3%	34.5%	Heidrich (2017, Figure 10, Appendix B)

Notes: See Table 1 for details about samples and income definitions used in each study.
[1] As the authors point out, this paper's baseline estimates are likely to overestimate upward mobility and underestimate persistence at the bottom and at the top because of lifecycle bias, the omission of taxpayers and tax evasion. The reported P(Top 20% | Bottom 20%) here corresponds to the estimate accounting as best as possible for these three sources of bias. For the other two measures, we report the estimates correcting for missing tax returns and tax evasion obtained from the authors.
[2] Obtained by multiplying the "Q1Q5" estimate found in the last column of Table 14 by the ratio of the two rows in Table 11, i.e., $0.100 \times 0.099/0.115$.
[3] Child incomes are measured relatively early in the lifecycle (32-34 years old), thus these estimates may suffer from lifecycle bias (i.e., overestimating upward mobility and underestimating persistence). By comparison, the father-son P(Child Top 20% | Parent Bot. 20%) estimate in Nybom and Stuhler (2017, Figure 1, Panel D) is essentially 10%, a much lower estimate of upward mobility.
* The authors very kindly shared more detailed estimates than reported in their papers.

Table 2: Transition Matrix in International Comparison

definition be 20%. We analyze persistence at the top of the parent income distribution in more detail in Appendix Section C.5.

Note that among the corner elements of the transition matrix, the estimates of mobility (i.e., P(Child Top 20% | Parent Bot. 20%) and P(Child Bot. 20% | Parent Top 20%)) are likely to be upper bounds, while estimates of persistence (i.e., P(Child Bot. 20% | Parent Bot. 20%) and P(Child Top 20% | Parent Top 20%)) are likely to be lower bounds. This is because the potential measurement error in parent rank prediction induced by TSTSLS can only go in one direction for the bottom and top quintiles. Parents in the bottom 20% necessarily have a true rank in the bottom 20% or above, but not below, as ranks take positive values by definition. Reasonably assuming that the probability of reaching the top 20% is increasing in parent income rank, our estimate of P(Child Top 20% | Parent Bot. 20%) is therefore likely to be an upper bound. In line with this intuition, the PSID validation exercise suggests that TSTSLS transition matrices overstate mobility relative to observed transition matrices (see Appendix Table B.4). The same reasoning can be applied to the other corner elements of the transition matrix.

In Table 2 we compare conditional probabilities of interest with those found for other developed countries. In France income persistence across generations is particularly strong, both at the top and at the bottom. While France does better than the United States when it comes to upward mobility from the bottom quintile (9.7% vs. 7.5%), a point we discuss in Section 4.4, it fares significantly worse than countries such as Canada (11.4%), Switzerland (11.9%) or Australia (12.3%). It also displays one of the strongest persistence at the bottom and at the top of the income distribution.

4.4 Discussion of Baseline Results

International Comparison. Our findings confirm the conventional wisdom that France exhibits strong income persistence across generations relative to many OECD countries (OECD, 2018). This is true not only with respect to the IGE, which has been the main focus for cross-country comparisons in the literature (e.g., see Corak (2016)), but also for the RRC, and in terms of transition matrices. This raises the question of the underlying mechanisms. Indeed, one apparent puzzle is that various studies have found positive effects of government spending on intergenerational mobility (Mayer and Lopoo, 2008; Huang et al., 2021). Yet, despite significant government spending, France displays relatively little intergenerational mobility.

However, though the IGE and RRC estimates are fairly similar for France and the United States, the two countries differ in terms of the probability of reaching the top 20% conditional on having parents in the bottom 20%. Given the large dissimilarities in their higher education systems, part of the explanation could stem from differences in access to, and graduation from, higher education along the parent income distribution.

Access to and Graduation from Higher Education. Using the yearly census surveys available since 2004 in the EDP, we can observe children's last obtained diploma when they are between 23 and 45.[20] Figure 5 compares higher education *graduation rates* in France with *enrollment rates* in the United States (defined by Chetty et al. (2020) as attending college at least at some point between ages 18-21) by parent income rank. To avoid capturing the direct effect of parent education (independent from parent income) on child higher education graduation, we use parent income ranks obtained when excluding parent education from the set of first-stage predictors. This has virtually no effect on the result. Graduation rates in France are lower than enrollment rates in the United States, which is expected considering that a sizable share of students who enroll in higher education eventually drops out. While the relationship between parent income rank and enrollment is linear in the United States, obtaining a higher education degree appears to be a convex function of parent income rank in France. In particular, it is flatter at the bottom of the distribution.[21] This convex relationship is all the more striking since children from low-income families are probably more likely to drop out from higher education, and therefore not earn a higher education degree.

This comparison does not allow us to assess directly whether higher education may explain the gap in upward mobility between France and the United States, since the relationship between college completion and parent income rank for the latter is

[20] We observe this information for 86% of the sample. The share of missing values is pretty well uniformly distributed along the parent income rank distribution.

[21] Appendix Figure E.6 documents the graduation rate for each cell of the quintile-by-quintile transition matrix. It shows that the convexity in the relationship between family background and graduation rate holds within child income quintile.

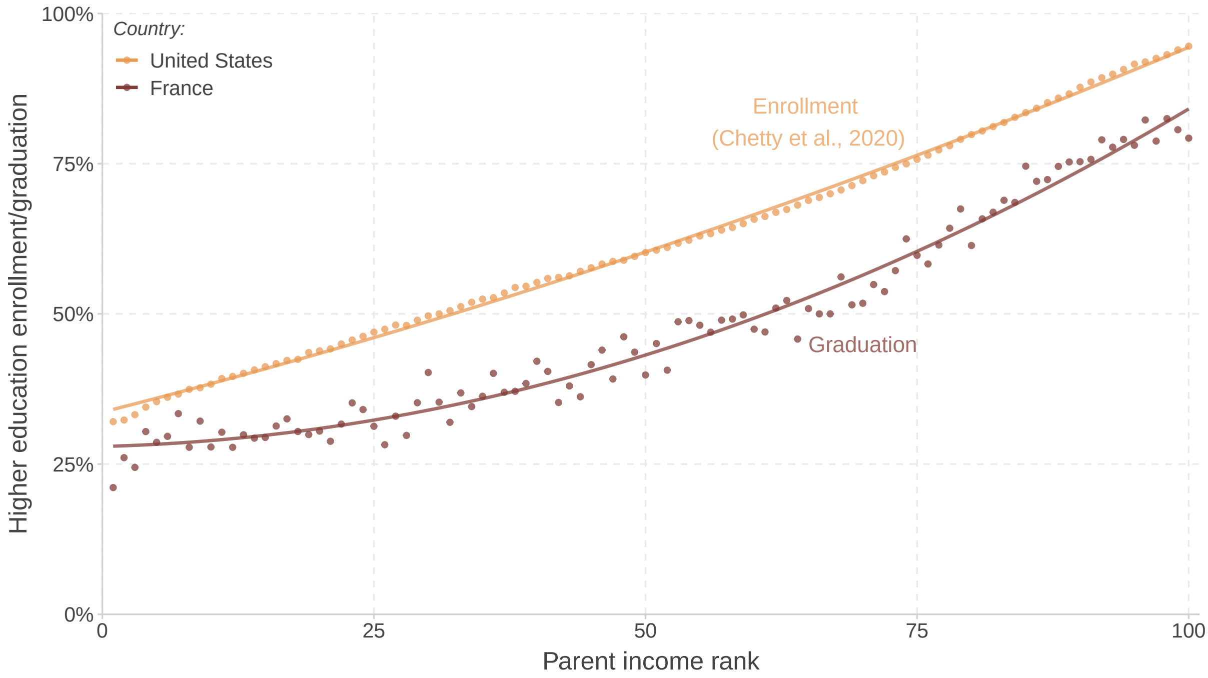


Figure 5: Graduation From/Enrollment In Higher Education by Parent Income

Notes: This figure presents higher education graduation in France vs. enrollment rates in the United States (Chetty et al., 2020) by parent income rank. See Figure 3's notes for details on data, sample and income definitions. In this figure parent income ranks are computed without parent education in the set of first-stage predictors to avoid capturing the effect of parent education independent from that of parent income.

not available. Using a French survey of roughly 6,000 18-24 year olds, Bonneau and Grobon (2022) find that enrollment rates in higher education by parent income rank are very similar in France compared to the United States. Therefore, if higher education were to explain part of the upward mobility gap observed between the two countries, it must necessarily be trough differences in dropouts rates and/or heterogeneous returns to higher education along the parent income distribution.

5 Robustness of Baseline Results

In addition to the method validity exercise presented in Section 3.1, we assess the sensitivity of our baseline results to the TSTSLS method by (i) varying the set of instruments, and (ii) relaxing parametric assumptions. Moreover, as discussed in Section 2.2, two statistical biases may affect our baseline estimates: lifecycle and attenuation bias. The former relates to heterogeneous lifecycle earnings profiles among parents and children, while the latter refers to classical measurement error in parent income. We therefore assess how our estimates vary with the age at which child and parent incomes are measured, and with the number of synthetic parent income observations used. We discuss additional potential biases (i.e., data coverage, treatment of zero incomes, and top and bottom income trimming) in Appendix C.

5.1 Two-Sample Two-Stage Least Squares

First-Stage Predictors. We first estimate the IGE, RRC, and transition matrices using only education as the first-stage predictor. We then add successively to the set of first-stage predictors: parents' (i) 2-digit occupation, (ii) demographic characteristics, and (iii) municipality-level characteristics. Our baseline specification corresponds to the one including the full set of predictors. The results are shown in Appendix Figure C.2.

Overall, our estimates are largely insensitive to the set of first-stage regressors, except for the IGE which is significantly larger when using only education in the first-stage. For example, the RRC (IGE) when using only education is 0.284 (0.679) compared to 0.303 (0.527) in our baseline. The transition matrices are also mostly unchanged: when using only education the P(Top 20% | Bot. 20%) is 10.8% compared to 9.7% in our baseline. These results are consistent with our validation exercise using the PSID where we find that the TSTSLS RRC estimate increases slightly once (3-digit) occupation is included as a predictor and the transition matrices are largely unaffected by the set of first-stage regressors (see Appendix Table B.4).

Functional Form. We estimate the first-stage using the three following flexible methods: (i) generalized additive model (GAM), (ii) gradient boosted tree, and (iii) the ensemble method. The results are shown in Appendix Figure C.3. These more flexible models yield essentially identical estimates and they do not lead to gains in terms of (out-of-sample) mean squared error.

5.2 Lifecycle and Attenuation Bias

Child Lifecycle Bias. Figure 6 presents our estimates of intergenerational income mobility when varying the age at which child income is measured. In addition to household income from the tax returns data, we exploit the longer time series wage data provided by the All Employee Panel. Each point represents the estimate of the measure of intergenerational income mobility when measuring child income at a given age. For the transition matrix, we only present the analysis for the conditional probability of being in the top or bottom 20% for children born to parents in the top or bottom 20%. The broad pattern that emerges in Figure 6 is that the estimated persistence (mobility) increases (decreases) sharply when child incomes are measured early in the lifecycle and stabilizes roughly when child income is measured in their mid-thirties.[22]

[22] By construction, each age estimate is obtained from a different sample since we only measure child incomes in the tax returns data between 2010 and 2016, and in the All Employee Panel from 1967 to 2015 (though only for individuals born in even years before 2001). The observed slight decline in the

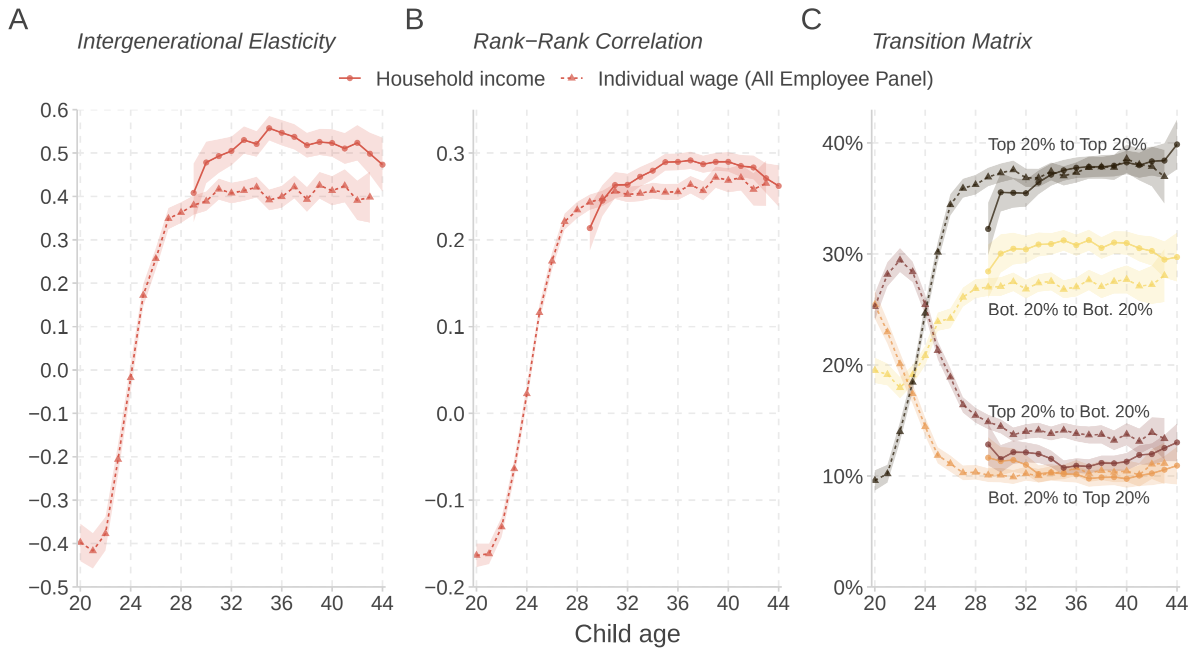


Figure 6: Child Lifecycle Bias

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to changes in the age at which child income is measured. Shaded areas represent the 95% bootstrapped confidence intervals. See Figure 3's notes for details on data, sample and income definitions.

Parent Lifecycle Bias. We assess the sensitivity of our baseline estimates to varying the age at which parent income is measured. Since we predict parent income rather than observe it, we vary the age at which synthetic parent income is measured in the first-stage regression. Specifically, we run the first-stage regression (equation (3)) defining synthetic parent income at a given age between 25 and 60 years old. Figure 7 shows that the relationship between age at which parent income is measured and persistence is concave, strongly increasing between 25 and the late thirties and then stabilizing until the mid to late fifties. Relative to our baseline estimate, it does not appear that our choice of measuring synthetic parent income as the average between 35 and 45 years old is either too early or too late in the lifecycle.[23]

Attenuation Bias. We evaluate the extent to which our baseline estimates are sensitive to the number of observations used to compute parent lifetime income. The main source of attenuation bias comes from measurement error in parent income.[24] Ap-

IGE and RRC estimates when children are in their forties for household income appears to mostly reflect changes in the underlying cohort sample rather than a real decrease in the estimate (see Appendix Figure C.4 where we reproduce the All Employee Panel estimates keeping the sample of children constant).

[23] In Appendix Section C.3.2 we study how our measures of intergenerational persistence vary with the age at which child and synthetic parent income is measured jointly.

[24] We also check in Appendix Section C.3.4 the sensitivity of intergenerational mobility to the number

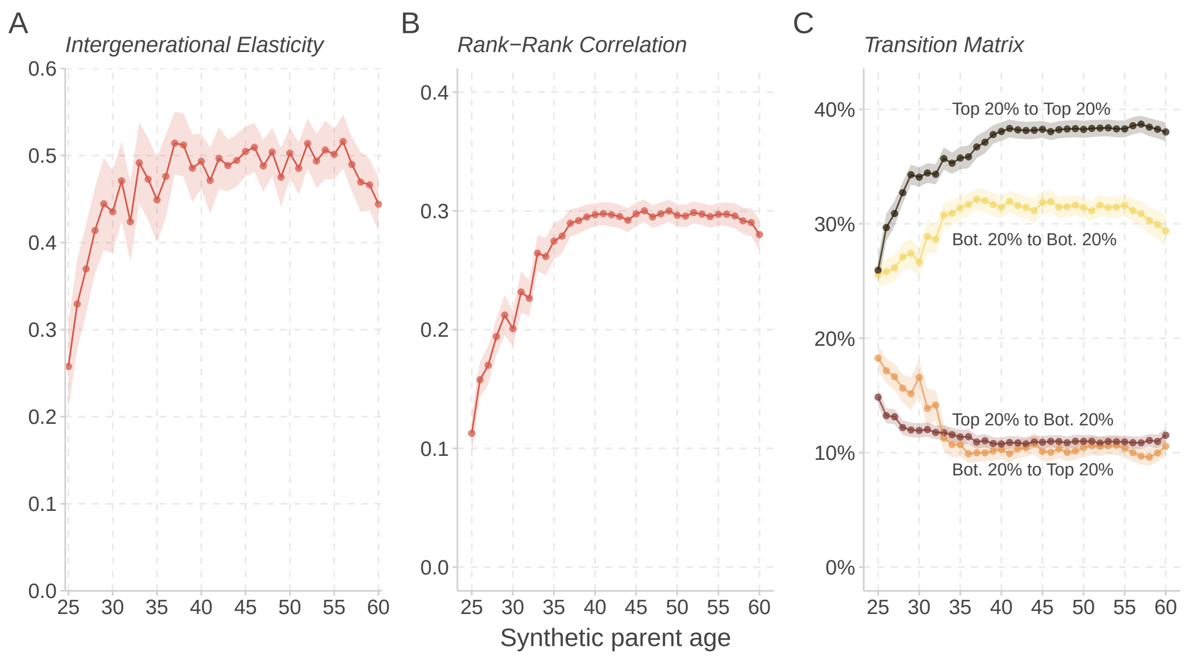


Figure 7: Parent Lifecycle Bias

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to changes in the age at which synthetic parent income is measured. Shaded areas represent the 95% bootstrapped confidence intervals. See Figure 3's notes for details on data, sample and income definitions.

pendix Figure C.6 plots estimates of our persistence measures varying the number of synthetic parent income observations used in the first-stage regression from 1 to 11 (see details in Appendix Section C.3.3). The rank-based measures, whether the RRC or the transition matrix cells, are remarkably unaltered by increasing the number of income observations over which synthetic parent income is averaged. However, the IGE increases gradually with the number of income observations, which largely rests on how mothers' incomes are predicted. In the context of TSTSLS estimation, this appears to be a strength of rank-based measures since it suggests that in cases where parent income is not observed, predicting it using only one synthetic parent income observation is likely to provide sufficiently accurate estimates. This is indeed what we find in our validation exercise, where the TSTSLS RRC bias is largely unchanged when increasing the number of parent income observations.

of child income observations and confirm that it only plays a very minor role.

6 Geographic Analysis

6.1 Heterogeneity Across Departments

A first step in understanding the sources of intergenerational mobility in France is to investigate where persistence is highest and lowest. We study the geographic variations of intergenerational mobility at the department level. Departments divide metropolitan France into 96 territories.[25] Departments have the advantage of covering the whole of metropolitan France, and their borders have not changed over the study period. In addition, considering finer geographic units such as commuting zones would imply dropping a sizable amount of areas due to insufficient sample size.

Children are assigned to their department of residence in 1990, when they were between 9 and 18 years old. This is the best proxy we have for the department they grew up in. To ensure our estimates are sufficiently reliable, we focus on the 85 departments with at least 200 observations.[26] Individuals are still ranked within the national income distribution.

Hereinafter we use parent income predicted without municipality characteristics in the first stage. This is to make sure that they do not spuriously drive any spatial patterns.[27] Moreover, we find that spatial variations in intergenerational mobility are not driven by differences in prediction accuracy of the first-stage across departments. Indeed, as shown in Appendix Table F.8, the department-level mean-squared errors of the first-stage predictions are not significantly related with department-level intergenerational mobility measures.

The statistics we use at the subnational level are (i) the IGE, (ii) the RRC, and (iii) the expected income rank for individuals whose parents locate at the 25$^{\text{th}}$ percentile, which we refer to as *absolute upward mobility* (AUM) following Chetty et al. (2014). We favor absolute upward mobility over specific cells of the transition matrix because of the size of our department samples. Indeed, while absolute upward mobility is estimated using all the observations in a given department, any cell of the quintile transition matrix is by construction estimated using only a fifth of these observations. Denoting $p_{c,d}$ the percentile income rank of children observed in department d during childhood, and $p_{p,d}$ the percentile income rank of their parents, local RRCs are obtained

[25] For practical reasons, we treat Corsica as a single department. Appendix Figure E.7 shows a map of French departments.

[26] The number of observations per department is reported in Appendix Table F.9.

[27] The removal of municipality characteristics from the first stage does not alter our national estimates (see Appendix Figure C.2) nor the first-stage R^2. Moreover, the cross-department correlation with and without municipality characteristics is above 0.97 for all three intergenerational mobility measures (IGE, RRC, AUM).

from the following OLS regression:

$$p_{c,d} = \alpha_d + RRC_d \times p_{p,d} + \varepsilon_d \tag{4}$$

The expected income rank for individuals whose parents locate at the 25$^{\text{th}}$ percentile then writes:

$$\text{AUM} := \mathbb{E}[p_{c,d} \mid p_{p,d} = 25] = \hat{\alpha_d} + R\hat{R}C_d \times 25 \tag{5}$$

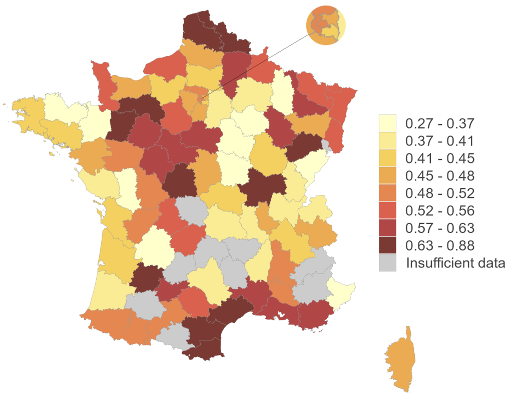
Appendix Figure E.8 graphically illustrates how this intergenerational mobility measure is computed for the Nord department, the most populated one in 1990. The conditional expectation functions for the most populated departments are available in Appendix Figures E.9 and E.10. Even at the department level, it appears that the rank-rank relationship is well approximated by a linear function.

Geographic Variations. Figure 8 depicts department-level intergenerational mobility as captured by the three estimators mentioned above. It reveals substantial variations, though not necessarily statistically significant likely due to a lack of statistical power.[28] The distribution of department-level RRCs ranges from 0.17 to 0.40 and is tighter than that of IGEs, which ranges from 0.27 to 0.88. Both vary across departments just as much as they vary across countries. The range of our estimates of absolute upward mobility, from rank 37 to rank 54, is almost identical to that observed in Italy using a comparable geographic unit (from 35 to 57 (Acciari et al., 2022)).

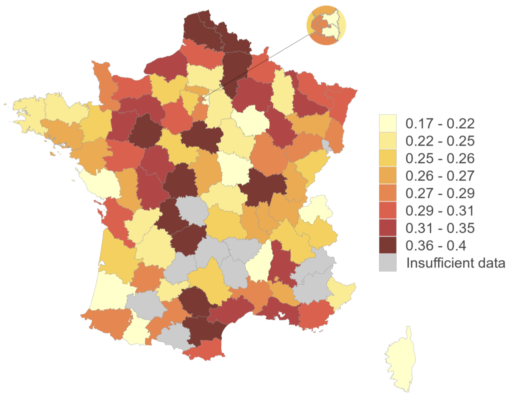
Intergenerational persistence is particularly high in the North and in the South of France, and relatively low in the West. For instance, the IGEs range from 0.30 to 0.45 in departments in Brittany (West), from 0.42 to 0.70 in departments in Hauts-de-France (North), and from 0.63 to 0.77 in the former region of Languedoc-Roussillon (South). This pattern is observed not only in terms of relative mobility (IGE and RRC), but also in terms of absolute upward mobility. Indeed, while children with modest socio-economic backgrounds have relatively high expected income ranks in Brittany (AUM $\in$ (43.3; 44.7)), they tend to remain lower in the income distribution in Hauts-de-France (AUM $\in$ (36.8; 44.1)) and Languedoc-Roussillon (AUM $\in$ (36.9; 39.3)).

However, a high relative mobility is not systematically associated with a high absolute upward mobility. For instance, such a discrepancy is observed for the municipality-department of Paris, the third highest department in terms of AUM, but where intergenerational mobility levels in terms of IGE and RRC are close to the department-level average. The conditional expectation functions in Appendix Figure E.10 provide an explanation to this idiosyncrasy. They reveal that the Parisian CEF is both shifted up-

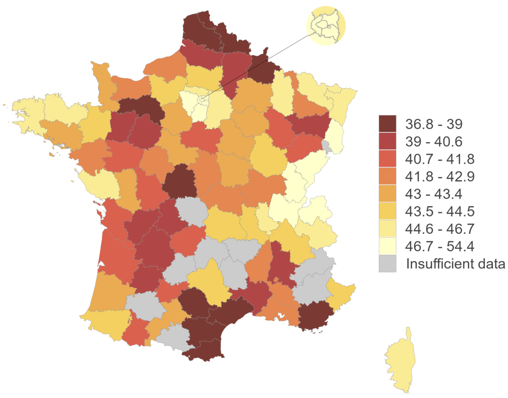
[28] Department-level estimates are reported in Appendix Table F.9. Department-level IGE, RRC and AUM are represented graphically with their confidence intervals in Appendix Figures E.11 to E.13.



(a) Intergenerational Elasticity



(b) Rank-Rank Correlation



(c) Absolute Upward Mobility

Figure 8: Spatial Variations in Intergenerational Mobility

Notes: This figure presents department-level estimates of our intergenerational mobility measures. To compute local estimates, individuals are assigned to their department of residence in 1990, when they were between 9 and 18 years old. Departments with less than 200 observations are considered as having insufficient data. See Figure 3's notes for details on data, sample and income definitions.

wards relative to other large departments, and flatter at the lower end of the parent income distribution. The combination of these two features results in relatively good prospects for children whose parents locate at the 25$^{\text{th}}$ percentile without implying particularly high relative mobility. The cross-department correlation between the IGE and RRC is 0.65, and is -0.55 with AUM (see Appendix Table F.10), which highlights the importance of using a variety of intergenerational mobility measures to characterize a country's income persistence across generations (Deutscher and Mazumder, forthcoming).

Correlation with Local Characteristics. To pin down potential sources of the spatial variations in intergenerational mobility, we explore the department characteristics that it might correlate with. We consider 14 variables, measured as close to 1990 as possible, classified into 5 groups: demographic, economic, inequality, education, and social capital variables. There are three main takeaways from this correlational analysis (additional details can be found in Appendix D).

First, the IGE appears to be only significantly related to the unemployment rate. This correlation is indeed striking visually when comparing the department-level unemployment rate in 1990, displayed in Appendix Figure E.14, with Figure 8a. Second, absolute upward mobility tends to exhibit much stronger relationships with department characteristics in general, than either the IGE or the RRC. This suggests that factors that affect absolute mobility might differ from those that affect relative mobility. A lasso analysis, detailed in Appendix D.3, yields similar insights.

Third, we find no evidence of a within France "Great Gatsby Curve", which refers to the positive correlation between intergenerational income persistence (defined by the IGE) and income inequality (defined by the Gini index) found across countries (Corak, 2013). The Gini index is significantly positively related to absolute upward mobility, the opposite sign one might expect if inequality is detrimental to intergenerational mobility. This contrasts with findings from Italy (Acciari et al., 2022) and North America (Chetty et al. (2014) for the United States and Corak (2020) for Canada).

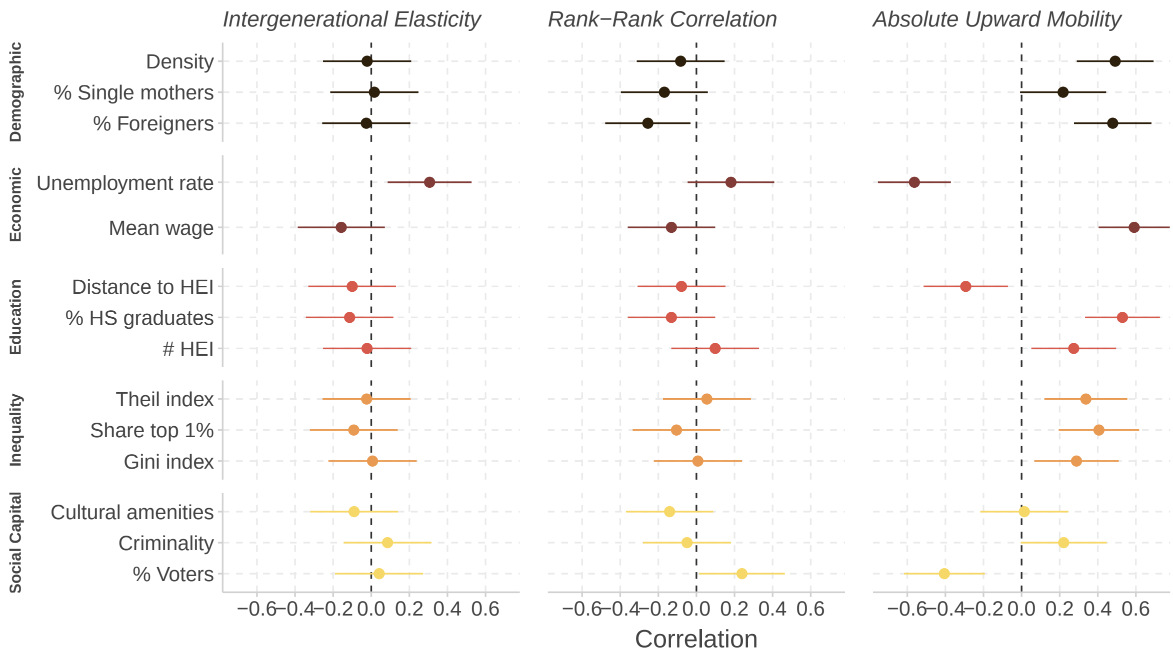


Figure 9: Intergenerational Mobility and Department Characteristics - Separate Estimation

Notes: This figure presents the regression coefficient between department-level intergenerational mobility and department characteristics. Each coefficient is obtained from a separate regression. Both the department intergenerational mobility estimates and the characteristics are standardized, implying that the coefficients can be interpreted as correlations. Horizontal lines represent the 95% confidence intervals. See Figures 3 and 8's notes for details on data, sample and income definitions, and Appendix Table D.1 for definitions and sources of the department characteristics.

6.2 Geographic Mobility

Few studies have explored the relationship between geographic mobility and intergenerational mobility.[29] We consider individuals as geographically mobile if their adulthood department of residence is different from their childhood department of residence. The childhood department of residence is observed in the 1990 census, when individuals were aged from 9 to 18 years old. The adulthood department of residence is the one indicated on individuals' tax return. If the individual has lived in several departments over 2010-2016, we consider the most common department of residence. In case of ties, we consider the most recent of the most common departments. According to this definition, 40.8% of individuals are geographically mobile. This share is relatively homogeneous across males (40.2%) and females (41.3%). The percentage of movers by parent household wage rank is presented in Appendix Figure E.15.

Intergenerational Mobility Gains from Geographic Mobility. Figure 10 shows the conditional expectation of child household income rank with respect to parent household wage rank for movers and stayers. The CEF is slightly flatter for movers than for stayers, and importantly, movers have systematically higher expected income ranks than stayers throughout the parent household wage rank distribution. The difference between the two CEFs is slightly decreasing in parent income and is particularly pronounced at the bottom of the distribution. This difference is the result of the combination of individuals self-selecting into migration and the causal effect of moving.

To characterize the relationship between intergenerational and geographic mobility, we estimate the following regression model:

$$p_{c,i} = \alpha + \beta p_{p,i} + \gamma \text{Mover}_i + \delta p_{p,i} \times \text{Mover}_i + X_i' \lambda + \varepsilon_i, \tag{6}$$

where $p_{c,i}$ is the household income rank of individual i, $p_{p,i}$ is individual i's parents' household wage rank, Mover_i is a binary variable taking the value 1 if individual i lives in a different department from the one they grew up in and 0 otherwise, and X_i is a set of control variables. Table 3 reports the corresponding regression results.

Column (1) shows the estimates from equation (6). Living in a different department from one's childhood department is associated, on average, with a $\mathbb{E}[\hat{\gamma} + \hat{\delta} p_{p,i}] = 5.89$ percentile rank increase in the national household income distribution. The point estimate of the rank-rank slope is slightly lower for movers when controlling for parents characteristics (coefficient $\hat{\delta}$ col. (4)-(5)), but not statistically significantly so. In the last specification, the difference in expected income rank between movers and stayers is

[29] Soria Espín (2022) analyzes this relationship in Spain, but other existing studies rather exploit geographic mobility to estimate the causal impact of location on upward mobility (Chetty and Hendren, 2018; Laliberté, 2021).

Figure 10: Intergenerational Mobility and Geographic Mobility

Notes: This figure represents the conditional expectation of child household income rank with respect to parent household wage rank separately for individuals whose adulthood department of residence is different or not from their childhood department of residence. Percentile ranks are computed according to the national income distribution, which implies that the share of movers and stayers is not constant throughout the parent income distribution. The childhood department of residence is observed in the 1990 census, when individuals were aged from 9 to 18 years old. The adulthood department of residence is the one indicated on individuals' tax return. If the individual has lived in several departments over 2010-2016, we consider the most represented department of residence. In case of ties, we consider the most recent of the most represented departments. See Figure 3's notes for details on data, sample and income definitions.

decreasing in parent income (5.36 at the 25$^{\text{th}}$ percentile and 4.71 at the 75$^{\text{th}}$ percentile).

The Role of Mobility Toward Richer Departments at the Aggregate Level. There are several potential reasons for the better intergenerational mobility outcomes movers tend to experience. One explanation may be that movers simply migrate to departments where wages are higher. To investigate this channel, we compute two statistics: (i) the mean parent household wage rank in the origin department, and (ii) the mean child household income rank in the destination department. Figure 11 displays the average of these two statistics for movers for each ventile of the parent household wage rank distribution.

There are three takeaways from this figure. First, the difference in average income rank in the destination and origin departments is highest at the top and bottom of the parent income distribution. Second, these differences are relatively small, reaching at most 2 percentile ranks for the top ventile. Third, the origin and destination departments of movers from the middle of the parent income distribution have very similar

	Dependent variable: Child household income rank				
	(1)	(2)	(3)	(4)	(5)
Parent income rank ($\hat{\beta}$)	0.278***	0.278***	0.268***	0.163***	0.138***
	(0.005)	(0.005)	(0.005)	(0.008)	(0.017)
Mover ($\hat{\gamma}$)	5.836***	5.858***	5.539***	5.716***	5.681***
	(0.472)	(0.472)	(0.475)	(0.472)	(0.475)
Parent income rank × Mover ($\hat{\delta}$)	0.001	0.0003	0.001	−0.012	−0.013
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	34.087***	33.780***	38.123***	29.195***	30.509***
	(0.258)	(0.274)	(1.228)	(1.659)	(1.782)
Birth cohort	✓	✓	✓	✓	✓
Gender		✓	✓	✓	✓
Department FE			✓	✓	✓
Parents' education				✓	✓
Parents' 2-digit occupation					✓
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_p] = \hat{\gamma} + \hat{\delta} \times 50.5$	5.89	5.87	5.59	5.11	5.02
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_p \mid p_p = 25]$	5.86	5.86	5.56	5.42	5.36
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_p \mid p_p = 75]$	5.91	5.91	5.61	4.82	4.71
Observations	64,571	64,571	64,571	64,571	64,571
Adjusted R^2	0.098	0.098	0.106	0.119	0.125

Notes: This table provides the estimates from regression child household income rank on their parents' income rank, a dummy variable indicating whether the individual is a mover, and the interaction between these two variables. Columns (2) to (5) progressively include control variables. See Figure 10 for details on variable and sample definitions. Bootstrapped standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 3: Intergenerational & Geographic Mobility

average income ranks. Put in parallel with the slight monotonic decrease in the gains from geographic mobility along the parent income rank distribution, it seems that these gains are not only due to individuals moving to higher-income departments.

Another way to test this hypothesis consists in comparing the conditional expectation functions of movers and stayers ranked either at the *national* and *department* level. Indeed, ranking individuals at the national level allows individuals born to parents who earn the median income of their department to be upward mobile by earning the median income of a higher-income department in adulthood. This channel can be removed by ranking individuals and their parents within departments. When doing so, movers can only be more intergenerationally mobile than stayers if they reach income ranks in their adulthood department that are further away from the rank of their parents in their childhood department. Finding no expected gains associated with geographic mobility when ranking individuals according to their department income distribution would suggest that the expected increase in income rank associated with mobility is fully driven by movers ending up in higher-income departments, but reaching on expectation a local income rank in their destination department that is not

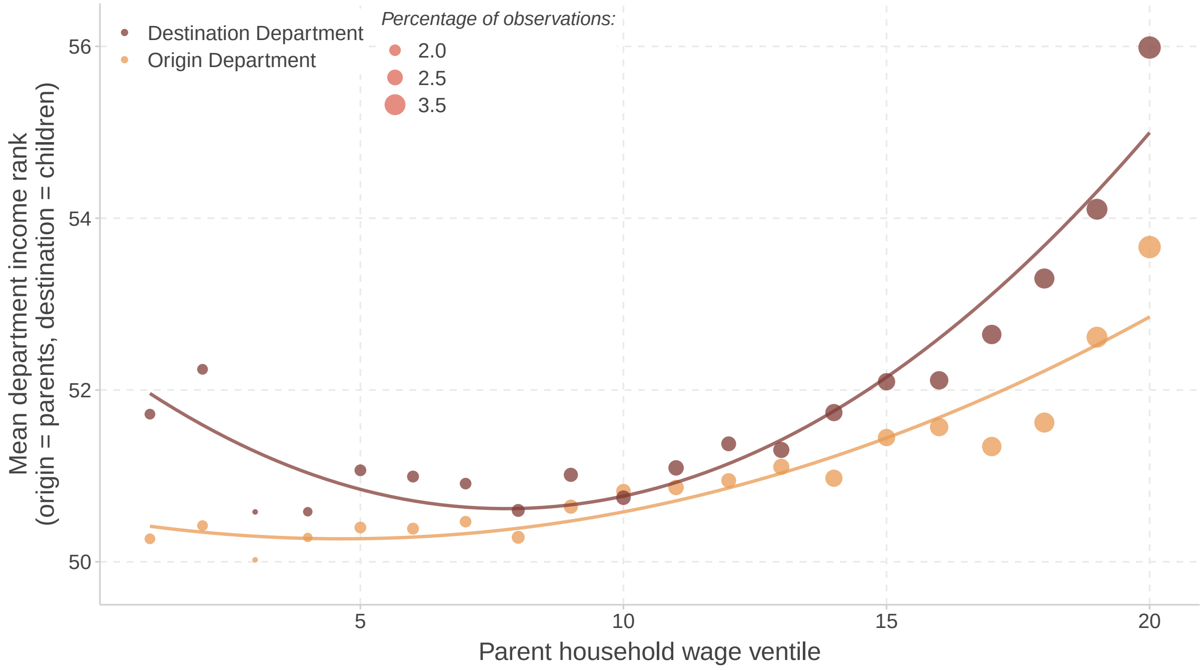


Figure 11: Mean Income Rank of Origin and Destination Departments of Movers

Notes: This figure represents the conditional expectation of income rank with respect to parent household wage rank for movers, separately by origin and destination departments. Origin department mean income rank is computed as the average income rank of residents in the parent sample, while destination mean income rank is computed as the average income rank of residents in the child sample. See Figures 3 and 10's notes for details on data, sample and income definitions.

further away from that of their parents, relative to stayers.

The regression results of equation (6) using percentile ranks computed at the department level rather than at the national level are reported in Appendix Table F.11 (Appendix Figure E.16 shows the corresponding conditional expectation functions). When considering ranks in the department distribution, the gap between the conditional expectation functions of movers and stayers shrinks but does not vanish completely. While the expected national-rank increase associated with mobility amounts to 5.89, it drops to 3.87 when considering local ranks. This suggests that the intergenerational mobility gains associated with geographic mobility are partly attributable to movers locating in higher-income departments in adulthood relative to stayers, but also to movers reaching local ranks in their adulthood department that are further away from the rank of their parents in the childhood department.

The Role of Mobility Toward Richer Departments at the Individual Level. While geographic mobility patterns between low- and high-income departments only partially explain the gap between movers and stayers at the aggregate level, characteristics of the destination department may be decisive at the individual level. To investigate this hypothesis we classify destination departments into three groups according to the average income rank of their residents from the child sample: (i) *low-income*, destination departments with an average income rank below 50 (70 departments - 49% of movers), (ii) *medium-income*, those with an average income rank between 50 and 60 (20 departments and overseas departments - 33% of movers), and (iii) *high-income*, those with an average income rank above 60 (5 departments and foreign countries - 18% of movers). This high-income group of departments greatly overlaps with the Parisian region as it comprises Essonne, Hauts-de-Seine, Paris, and Yvelines.

Figure 12 shows the conditional expectation of child income rank with respect to parent income ventile for the three destination department categories and for stayers. Results of the corresponding regression are reported in Appendix Table F.12. Except for the top ventiles, the CEFs of movers by destination department category are virtually parallel. Movers thus experience similar levels of relative mobility regardless of the income category of their destination department. However, movers' absolute upward mobility increases with the average income of the destination department, such that the expected income rank of a mover from the bottom of the parent income distribution to a high-income department is around the same as the expected income rank of a stayer from the 75th percentile of the parental income distribution. Still, such transitions are the exception: most movers to high-income departments come from high-income families, while low-income movers go predominantly to low- or medium-income departments. Another noteworthy finding is that expected income ranks are essentially the same for movers to low-income departments as for stayers, highlighting the potential role of the destination department's characteristics in generating upward intergenerational mobility for movers. All these findings combine self-selection and causal effects, and we leave the disentangling of these two channels for future research.

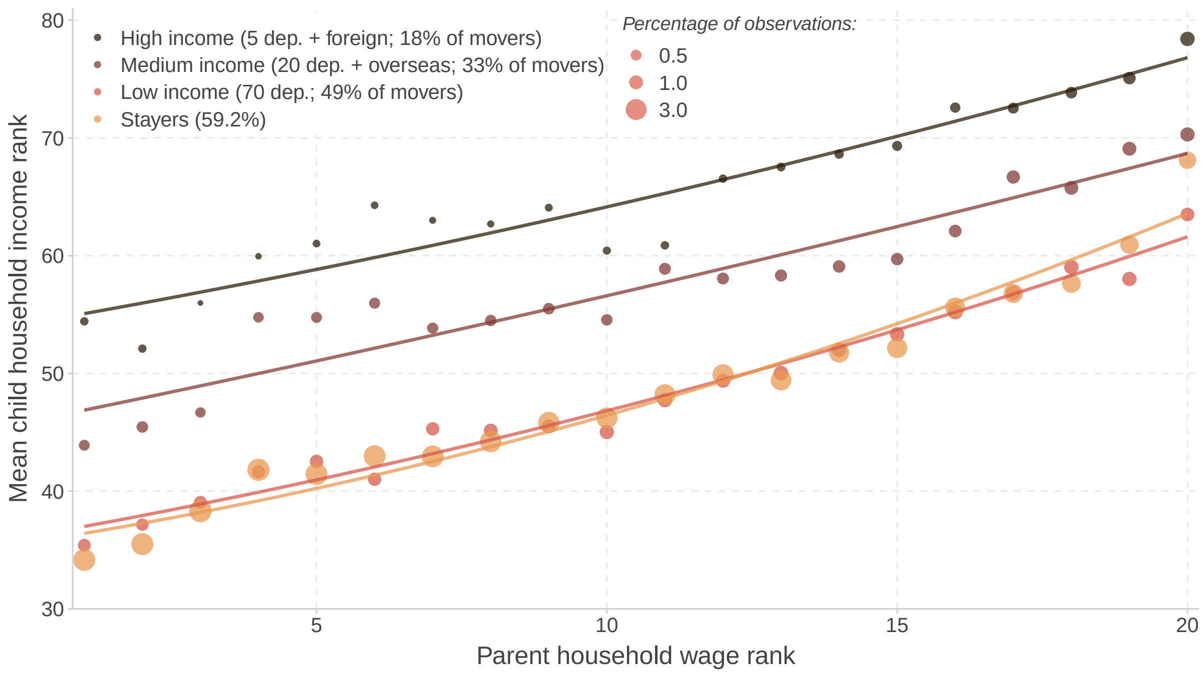


Figure 12: Mean Child Income Rank by Destination Department Mean Income

Notes: This figure represents the conditional expectation of child household income rank with respect to parent household wage rank for stayers and for movers to departments of different mean income categories. Solid lines represent second-order polynomial fits. Low income destination departments are destination departments with an average income rank below 50, medium income are those with an average income rank between 50 and 60, and high income are those with an average income rank above 60. See Figures 3 and 10's notes for details on data, sample and income definitions.

7 Conclusion

France is an interesting case study for intergenerational income mobility considering its relatively modest income inequality and the specificity of its higher education system. Yet, it has been the focus of few studies due to important data limitations. We use administrative data to provide an overview of intergenerational income mobility in France for individuals born in 1972-1981. Relative to existing studies, the richness of these data enables us to apply two-sample two-stage least squares (TSTSLS) using a much larger set of individual characteristics, and to extensively assess the robustness of the resulting estimates. Using the Panel Study of Income Dynamics (PSID) we find that the TSTSLS methodology slightly underestimates rank-based measures of intergenerational persistence relative to what would be obtained if parent income was observed.

Moreover, we provide the first estimates of the rank-rank correlation and transition matrix for France, and conduct a comparative analysis with other countries for which such statistics are available. Our results reveal that France exhibits a relatively strong intergenerational income persistence at the national level. It ranks among the highest in OECD countries, with Italy and the United States, and far from Australia, Canada, and Scandinavian countries.

This high intergenerational income persistence at the national level hides substantial geographic heterogeneity across departments. We observe about as much variation across French departments as across countries. Intergenerational persistence appears to be particularly high in the North and South, and relatively low in the Western part of the country. Yet, only *absolute* mobility, as opposed to *relative* mobility, significantly correlates with local characteristics.

We also provide novel descriptive evidence on a new mechanism that could explain some features of intergenerational mobility: geographic mobility. We find that the difference in expected income ranks between geographically mobile individuals and stayers is large and slightly decreasing in parent income. This difference appears not to be solely due to individuals moving to higher income departments but to be also the result of individuals moving up the local income rank ladder. Destination departments are on average characterized by higher income levels than origin departments only at the tails of the parent income distribution. However, regardless of parent income rank, conditional on moving the absolute upward mobility gains associated with moving to a higher-income department appear to be large and increasing with average income in the destination department. Even though not causal, we believe that these descriptive findings constitute promising avenues for future research to better understand intergenerational income mobility.

References

Abbas, Hicham and Michaël Sicsic, "Qui Gravit l'échelle Des Revenus Par Rapport à Ses Parents ?," Documents de Travail N° 2022 - F04, INSEE, Paris 2022.

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante, ""And Yet, It Moves": Intergenerational Mobility in Italy," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 118–163.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso, "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, February 2018, *108* (2), 521–554.

Barbieri, Teresa, Francesco Bloise, and Michele Raitano, "Intergenerational Earnings Inequality: New Evidence From Italy," *Review of Income and Wealth*, 2020, *66* (2), 418–443.

Bellégo, Christophe, David Benatia, and Louis-Daniel Pape, "Dealing with Logs and Zeros in Regression Models," Technical Report, SSRN Electronic Journal 2021.

Björklund, Anders and Markus Jäntti, "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review*, 1997, *87* (5), 1009–1018.

__, Jesper Roine, and Daniel Waldenström, "Intergenerational Top Income Mobility in Sweden: Capitalist Dynasties in the Land of Equal Opportunity?," *Journal of Public Economics*, June 2012, *96* (5), 474–484.

Black, Sandra E. and Paul J. Devereux, "Recent Developments in Intergenerational Mobility," in "Handbook of Labor Economics," Vol. 4, Elsevier, 2011, pp. 1487–1541.

Bloise, Francesco, Paolo Brunori, and Patrizio Piraino, "Estimating Intergenerational Income Mobility on Sub-Optimal Data: A Machine Learning Approach," *The Journal of Economic Inequality*, December 2021, *19* (4), 643–665.

Bonneau, Cécile and Sébastien Grobon, "Unequal Access to Higher Education Based on Parental Income: Evidence from France," Technical Report 2022/01, World Inequality Lab, Paris 2022.

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage, "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US," *The Scandinavian Journal of Economics*, 2017, *119* (1), 72–101.

Carmichael, Fiona, Christian K. Darko, Marco G. Ercolani, Ceren Ozgen, and W. Stanley Siebert, "Evidence on Intergenerational Income Transmission Using Complete Dutch Population Data," *Economics Letters*, 2020, *189*, 108996.

Chetty, Raj and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates," *The Quarterly Journal of Economics*, August 2018, *133* (3), 1163–1228.

— **, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Income Segregation and Intergenerational Mobility Across Colleges in the United States," *The Quarterly Journal of Economics*, August 2020, *135* (3), 1567–1633.

— **, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Chuard-Keller, Patrick and Veronica Grassi, "Switzer-Land of Opportunity: Intergenerational Income Mobility in the Land of Vocational Education," Technical Report, SSRN Electronic Journal 2021.

Corak, Miles, "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, 2013, *27* (3), 79–102.

— , "Inequality from Generation to Generation: The United States in Comparison," Discussion Paper 9929, IZA, Bonn, Germany 2016.

— , "The Canadian Geography of Intergenerational Income Mobility," *The Economic Journal*, 2020, *130* (631), 2134–2174.

Cortes-Orihuela, Javier, Juan Díaz, Pablo Gutierrez Cubillos, Pablo Troncoso, and Gabriel Villarroel, "How Much Should We Trust TSTSLS Intergenerational Mobility Estimates?: Evidence From A Developing Country," *SSRN Electronic Journal*, 2022.

Couch, Kenneth A. and Dean R. Lillard, "Sample Selection Rules and the Intergenerational Correlation of Earnings," *Labour Economics*, 1998, *5* (3), 313–329.

Dahl, Molly and Thomas DeLeire, "The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data," Technical Report No. 1342-08, Institute for Research on Poverty 2008.

Deutscher, Nathan and Bhashkar Mazumder, "Intergenerational Mobility across Australia and the Stability of Regional Estimates," *Labour Economics*, 2020, *66*, 101861.

— **and** — , "Measuring Intergenerational Income Mobility: A Synthesis of Approaches," *Journal of Economic Literature*, forthcoming.

Eriksen, Jesper, "Finding the Land of Opportunity: Intergenerational Mobility in Denmark," Technical Report, Aalborg University 2018.

Espín, Javier Soria, "Intergenerational Mobility, Gender Differences and the Role of Out-Migration: New Evidence from Spain," 2022, p. 98.

Grawe, Nathan D., "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence," *Labour Economics*, 2006, *13* (5), 551–570.

Haider, Steven and Gary Solon, "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 2006, *96* (4), 1308–1320.

Heidrich, Stefanie, "Intergenerational Mobility in Sweden: A Regional Perspective," *Journal of Population Economics*, 2017, *30* (4), 1241–1280.

Helsø, Anne-Line, "Intergenerational Income Mobility in Denmark and the United States," *The Scandinavian Journal of Economics*, 2021, *123* (2), 508–531.

Huang, Xiao, Shoujun Huang, and Ailun Shui, "Government Spending and Intergenerational Income Mobility: Evidence from China," *Journal of Economic Behavior & Organization*, November 2021, *191*, 387–414.

Inoue, Atsushi and Gary Solon, "Two-Sample Instrumental Variables Estimators," *The Review of Economics and Statistics*, 2010, *92* (3), 557–561.

Jacome, Elisa, Ilyana Kuziemko, and Suresh Naidu, "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century," 2023.

Jerrim, John, Álvaro Choi, and Rosa Simancas, "Two-Sample Two-Stage Least Squares (TSTSLS) Estimates of Earnings Mobility: How Consistent Are They?," *Survey Research Methods*, 2016, *10* (2), 85–102.

Kalambaden, Preetha and Isabel Z Martınez, "Intergenerational Mobility Along Multiple Dimensions Evidence from Switzerland," 2021.

Laliberté, Jean-William, "Long-Term Contextual Effects in Education: Schools and Neighborhoods," *American Economic Journal: Economic Policy*, May 2021, *13* (2), 336–377.

Landersø, Rasmus and James J. Heckman, "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US," *The Scandinavian Journal of Economics*, 2017, *119* (1), 178–230.

Lee, Chul-In and Gary Solon, "Trends in Intergenerational Income Mobility," *Review of Economics and Statistics*, 2009, *91* (4), 766–772.

Lefranc, Arnaud, "Intergenerational Earnings Persistence and Economic Inequality in the Long Run: Evidence from French Cohorts, 1931-75," *Economica*, 2018, *85* (340), 808–845.

— **and Alain Trannoy**, "Intergenerational Earnings Mobility in France: Is France More Mobile than the US?," *Annales d'Économie et de Statistique*, 2005, (78), 57–77.

— **, Nicolas Pistolesi, and Alain Trannoy**, "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France," *Journal of Public Economics*, December 2009, *93* (11-12), 1189–1207.

Mayer, Susan E. and Leonard M. Lopoo, "Government Spending and Intergenerational Mobility," *Journal of Public Economics*, February 2008, *92* (1-2), 139–158.

Mazumder, Bhashkar, "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data," *Review of Economics and Statistics*, 2005, *87* (2), 235–255.

— , "Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data," in Lorenzo Cappellari, Solomon W. Polachek, and Konstantinos Tatsiramos, eds., *Research in Labor Economics*, Vol. 43, Emerald Group Publishing Limited, 2016, pp. 83–129.

Nybom, Martin and Jan Stuhler, "Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation," *Journal of Human Resources*, January 2016, *51* (1), 239–268.

— **and** — , "Biases in Standard Measures of Intergenerational Income Dependence," *Journal of Human Resources*, 2017, *52* (3), 800–825.

OECD, *A Broken Social Elevator? How to Promote Social Mobility*, OECD, June 2018.

O'Neill, Donal, Olive Sweetman, and Dirk Van de gaer, "The Effects of Measurement Error and Omitted Variables When Using Transition Matrices to Measure Intergenerational Mobility," *Journal of Economic Inequality*, 2007, *5* (2), 159–178.

Robert-Bobée, Isabelle and Natacha Gualbert, "L'échantillon Démographique Permanent : En 50 Ans, l'EDP a Bien Grandi !," in "Courrier Des Statistiques N6" number 6, Paris: INSEE, 2021.

Solon, Gary, "Intergenerational Income Mobility in the United States," *American Economic Review*, 1992, *82* (3), 393–408.

Zimmerman, David J, "Regression Toward Mediocrity in Economic Stature," *American Economic Review*, 1992, *82* (3), 409–429.

Online Appendix

A Data - Details

The Permanent Demographic Sample (EDP) is a panel of individuals which the French statistical office, INSEE, started in 1968.[1] It combines several administrative data sources on individuals born on the first four days of October.[2] Individuals born on one of these days are called EDP individuals. The EDP gathers data from 5 administrative sources: (i) civil registers since 1968; (ii) population censuses since 1968 (exhaustive in 1968, 1975, 1982, 1990 and 1999, and yearly rotating 20% random samples since 2004); (iii) the electoral register since 1990; (iv) the *All Employee Panel* since 1967; and (v) tax returns since fiscal year 2011.

Each time an individual born on the first four days of October appears in one of these administrative datasets, the information contained in it is added to their individual identifier in the EDP. Therefore all these datasets can be matched together using a common individual identifier. For our analysis we use data from civil registers, the 1990 census, the All Employee Panel and tax returns. We describe each data source in detail below.

Civil Registers. They contain information from birth certificates of EDP individuals and their children, as well as death and marriage certificates of EDP individuals, since 1968. We use birth certificates of EDP individuals and their children which include the child's gender, date and place of birth, and information on each parent including date and place of birth, nationality and occupation. There are no data breaks or missing certificates for the years under study (1972-1981).

1990 Census. It contains socio-demographic information about EDP individuals, as well as, though to a lesser extent, about members of their household. These include the individual's date and place of birth, nationality, education, occupation, marital status, household structure, dwelling characteristics, building when relevant, and municipality.

All Employee Panel. It combines two sources of data: the annual declarations of social data (*déclarations annuelles des données sociales* - DADS) and data on central government employees (*fichiers de paie des agents de l'état* - FPE). All businesses are obliged to annually communicate the declarations of social data about their employees to a network of private organizations (*Unions de recouvrement des cotisations de sécurité sociale et d'allocations familiales* - URSSAF) coordinated by a government agency (*Agence centrale des organismes de sécurité sociale* - ACOSS). The All Employee Panel data are reported at the worker-year level, aggregated by INSEE from data at the worker-firm-year level. As such, annual pretax wage and annual hours worked correspond to the sum over all the individual's salaried activities. The job characteristics correspond to the year's "main" job, that is the job for which the pay period was the longest and, in case of a tie, the job with the highest wage.

Between 1967 and 2001, data is only available for individuals born on an even year. The scope of workers covered by the All Employee Panel has varied over time. Since 1967 in metropolitan France, all private sector employees, except those in the agricultural sectors, and including employees of public enterprises, are covered. The hospital public service is integrated in 1984, the state civil service and local authorities in 1988. France Télécom and La Poste employees appear only in 1988 as well. See Appendix C.1 for a robustness check to this public sector coverage evolution. The agricultural sector and overseas territories are included in 2002, and employees of private employers in 2009. Unemployment insurance is included from 2008 onwards. Lastly, because of increased workload due to the population censuses of 1982 and

[1] The EDP user guide (in French) can be found here.

[2] The EDP selection criterion has progressively widened to include individuals born on the first days of January, April, and July. See Robert-Bobée and Gualbert (2021) for a detailed description of the dataset.

1990, the All Employee Panel data were not compiled by INSEE in 1981, 1983 and 1990.

Tax Returns. They are compiled using housing and income tax forms filed for incomes earned from 2010 to 2016. In particular, household-level tax returns information is constructed based on dwellings where an EDP individual is known either from the income tax return or from the principal housing tax (*taxe d'habitation principale*). The location of the individual is that declared on January 1^{st} of the fiscal declaration year. Income variables are available at the household-level as well as at the individual level. Since the information is gathered based on living in the same dwelling, household income is computed not only for couples who file their taxes jointly, but also for couples who live together, an increasingly common arrangement. This departs from existing studies based on tax returns data which can only assign households based on marital status (Chetty et al., 2014). The scope of fiscal households excludes individuals living in collective structures (retirements homes, religious communities, student accommodations, prisons, etc.) as well as those most in distress, who live in precarious housing (worker hostels, etc.) or are homeless.

B PSID Validation Exercise

We use the Panel Study of Income Dynamics (PSID) to assess the extent to which OLS and two-sample two-stage least squares (TSTSLS) estimates of rank-based intergenerational mobility measures differ from one another. Our sample and definition choices aim to be as close as possible to our main analysis setting while at the same time maximizing sample size. Note also that for this reason we use all of the PSID, rather than only the nationally-representative Survey Research Center (SRC) component. The main conclusions of our baseline results are robust to using only the SRC sample or to using various weighing schemes as shown in Section B.7.7.

B.1 Sample Definitions

Sample of Children. It consists of individuals who are (i) born between 1963 and 1988, (ii) observed as children in a family unit at least once, and (iii) observed at least once as reference person or partner in a family unit between 30 and 50 years old. Restriction (i) enables us to identify parents, while restriction (ii) enables us to observe children's incomes. The final sample contains 5,655 children.[3]

Sample of Parents. Following Chetty et al. (2014), for each child, we define parent(s) as the reference person and partner of the family unit in which the child is first observed.[4] We then follow these individuals' incomes over time.[5] As Chetty et al. (2014), for simplicity, we fix each child's parent assignment regardless of any potential subsequent changes to the child's family unit reference person and partner. The final sample contains 5,785 (unique) parents.

B.2 Variable Definitions

All income variables are measured in 2019 dollars, adjusting for inflation using the consumer price index (CPI-U). Following Lee and Solon (2009) and Mazumder (2016), we exclude income observations obtained by "major assignment". We opt for larger age ranges than in our main analysis (30-50 vs 35-45) to increase our sample size. However, our baseline results are robust to averaging over 35-45 as in the main analysis (see Appendix Table B.5).

Parent Income. We rely on two parent income definitions. First, as a benchmark, we measure parent income as total pretax income at the household level, which we label parent family income. Specifically, we define parent family income as the sum of taxable income of the family unit's reference person and partner, and total transfer income of the reference person and partner.[6] Taxable income is equal to the sum of reference person's labor income, the partner's labor

[3] See Appendix Table B.2 for the sample size at each additional restriction.

[4] 90% of individuals born in 1963-1988 are first observed as children in a family unit prior to age 18.

[5] Note that this differs from the following studies using the PSID: Lee and Solon (2009) (use the family taxable income in which the children find themselves between ages 15 and 17), Mazumder (2016) (uses the PSID's Family Identification Mapping System (FIMS) to identify fathers), Jerrim et al. (2016) (do not explain exactly how fathers are identified; to be precise, the authors write "[...] we only include sons whose father can be identified," (Jerrim et al., 2016, p.89)), and Bloise et al. (2021) (do not explain exactly how fathers are identified; to be precise, the authors write "we include only sons whose real fathers have at least five years of positive earnings [...]," (Bloise et al., 2021, p.650)).

[6] The accuracy of the family's taxable income is missing in 1993-1996 and in 2001-2019. Total transfers are missing in 1968 and 1969. Total transfers include aid to families with dependent children, supplemental security income, other welfare payments, social security payments, other retirement, pensions and annuities, unemployment pay, workmen's compensation, child support, help from relatives, and other transfer income.

income, income from assets, and net profit from farm or business. This measure enables us to obtain benchmark estimates that the TSTSLS estimation strategy is supposed to yield.

Second, since in TSTSLS strategies parent family income is rarely observed, we also define parent labor income as the sum of family unit's reference person and partner's individual labor incomes (money income from labor, including self-employment income).[7] This follows very closely the setting adopted in the main analysis.

For both parent family income and parent labor income, we average income values over 30 and 50 years old. Specifically, we take the sum of the average for the father and the average for the mother if both parents are observed, and take the average of the only observed parent otherwise.

Child Income. We define child income in the same way as parent family income, again averaging over income observations between 30 to 50 years old.

Adjustment for Household Size. When defining income variables we follow Chetty et al. (2014), and do not account for household size (i.e., whether there is also a partner in the family unit). This way of defining parent income mechanically hinders single-headed households, both parents and children.[8] We therefore show in Table B.5 results when dividing family income measures by the number of observed reference person and partner in that year.

Descriptive Statistics Appendix Table B.1 displays some descriptive statistics for our sample of parents and children. Parents' incomes are observed at a slightly older age (39) than that of our children (34). In both cases, incomes are measured sufficiently late in the lifecycle to limit lifecycle bias.

Table B.1: Descriptive Statistics

	N	Missing (%)	Mean	Std. Dev.	25th pctile	Median	75th pctile
Parents							
Family income (average 30-50 yrs old)	5,785	5.88	82,047	66,121	42,976	72,081	105,523
Number of family income observations	5,785	5.88	13	5	10	15	18
Mean age at family income obs.	5,785	5.88	39	3	38	39	40
Labor income (average 30-50 yrs old)	5,785	5.62	39,679	39,946	13,575	30,800	55,074
Number of labor income observations	5,785	5.62	14	5	10	15	19
Mean age at labor income obs.	5,785	5.62	39	3	38	39	40
Fraction single parents	20.19%						
Fraction female among single parents	92.21%						
Mother's age at child birth	3,135	0.00	25	5	21	25	29
Father age at child birth	2,650	0.00	28	6	24	28	32
Children							
Family income (average 30-50 yrs old)	5,655	3.02	80,539	75,072	34,936	64,092	104,517
Number of family income observations	5,655	3.02	5	3	2	4	8
Mean age at family income obs.	5,655	3.02	34	3	32	34	38
Fraction female	53.60%						

Notes: See Sections B.1 and B.2 for details on sample construction and income definitions. Missing income observations can also correspond to values obtained by 'major assignment'.

[7]The accuracy of the reported value for the reference person is missing in 1994-1996. Moreover, for partners, there was a small change in income definition in 1994: total labor income became total labor income excluding farm and business income.

[8]Interestingly, this is an issue Raj Chetty alludes to in his conversation with Tyler Cowen in his 2017 *Conversations with Tyler* podcast episode. Indeed, Chetty noticed that daughters from affluent families in the Bay area have low *household* incomes but have very high *individual* incomes because they are significantly less likely to be married than if they had grown up somewhere else.

B.3 Benchmark Estimates

We first estimate the rank-rank correlation (RRC) and transition matrix using the family income definitions for both parents and children (results for the intergenerational income elasticity (IGE) are presented in Appendix Figure B.5). Recall that in the TSTSLS setting the parent income definition is parent labor income while we are actually interested in parent family income, a more comprehensive parent income measure. In theory, the extent to which the additional incomes included in parent family income relative to parent labor income generate large rank reversals is ambiguous. Moreover, TSTSLS estimates necessitate restricting the analysis to the sample of children for whom parents' characteristics are observed (e.g., education and/or occupation, etc.). Such restrictions could potentially induce some biases relative to the statistic one is actually interested in measuring.

National Results. Appendix Figure B.1 displays the benchmark RRC and transition matrix for the baseline parent and child income definitions. The baseline RRC is 0.504, compared to 0.34 found in Chetty et al. (2014). Such a high RRC likely reflects the fact that the PSID contains oversamples of disadvantaged families (see Appendix Figure B.7 for estimates obtained only on the Survey Research Center (SRC) component of the PSID). The benchmark transition matrix confirms this intuition. The share of children from the bottom 20% who reach the top 20% in adulthood is 4%, close to half the share found by Chetty et al. (2014) (7.5% for children born in 1980-1982). Persistence at the bottom and top are also very strong at roughly 45%.

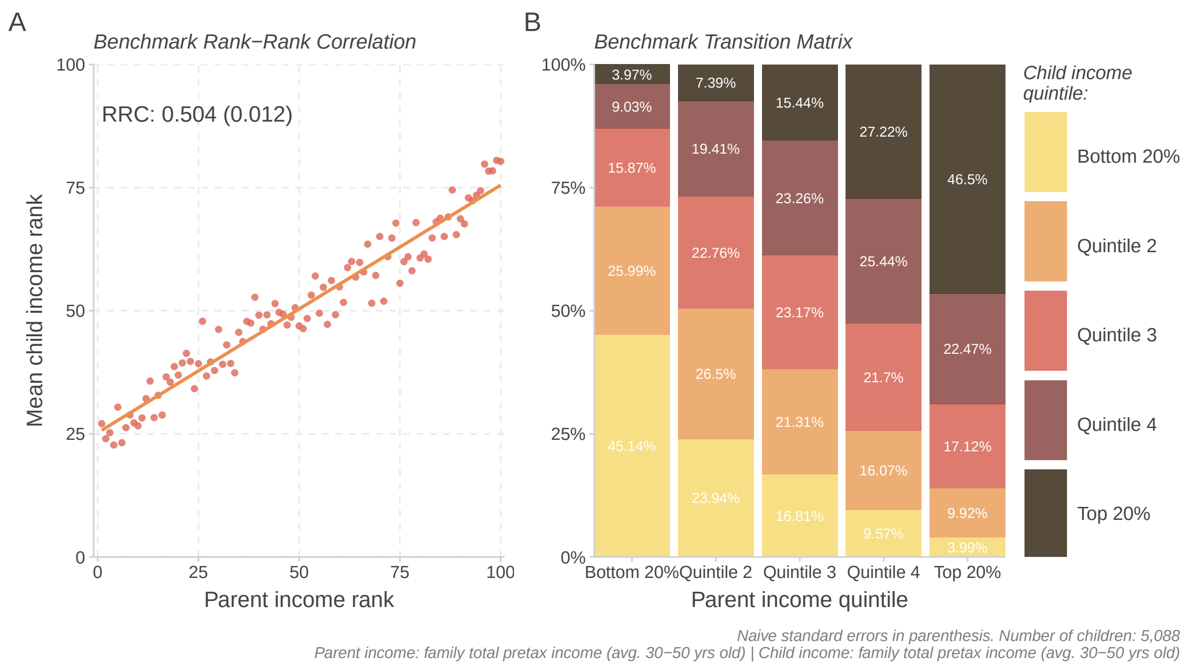


Figure B.1: Benchmark Rank-Rank Correlation and Transition Matrix

Notes: This figure presents the rank-rank correlation (panel A) and the transition matrix (panel B). It is computed on the Panel Study of Income Dynamics (PSID). The sample used is restricted to children born between 1963 and 1988 who are observed at least once as children in a family unit and at least once as a reference person or partner in a family unit over ages 30-50. Child income is the mean of family total income over ages 30-50. Parent income is the sum of father and mother mean family total income over ages 30-50. In panel A, the fitted line is a linear fit through the conditional expectation. We report coefficients and naive standard errors (in parenthesis) obtained from OLS regressions of child income rank on parent income rank with child cohort fixed effects, on the microdata for the full sample.

Subnational Results. Due to sample size constraints we explore geographic heterogeneity in intergenerational by Census Region (Northeast, Midwest, South, and West). Specifically, we define a child's Census Region as the most common region of residence until age 18 (included). Appendix Figure B.2 displays the benchmark RRC and absolute upward mobility (AUM) estimates by Census Region. AUM is defined as in Chetty et al. (2014) as the expected income rank for children at the 25th percentile of the parent income distribution.

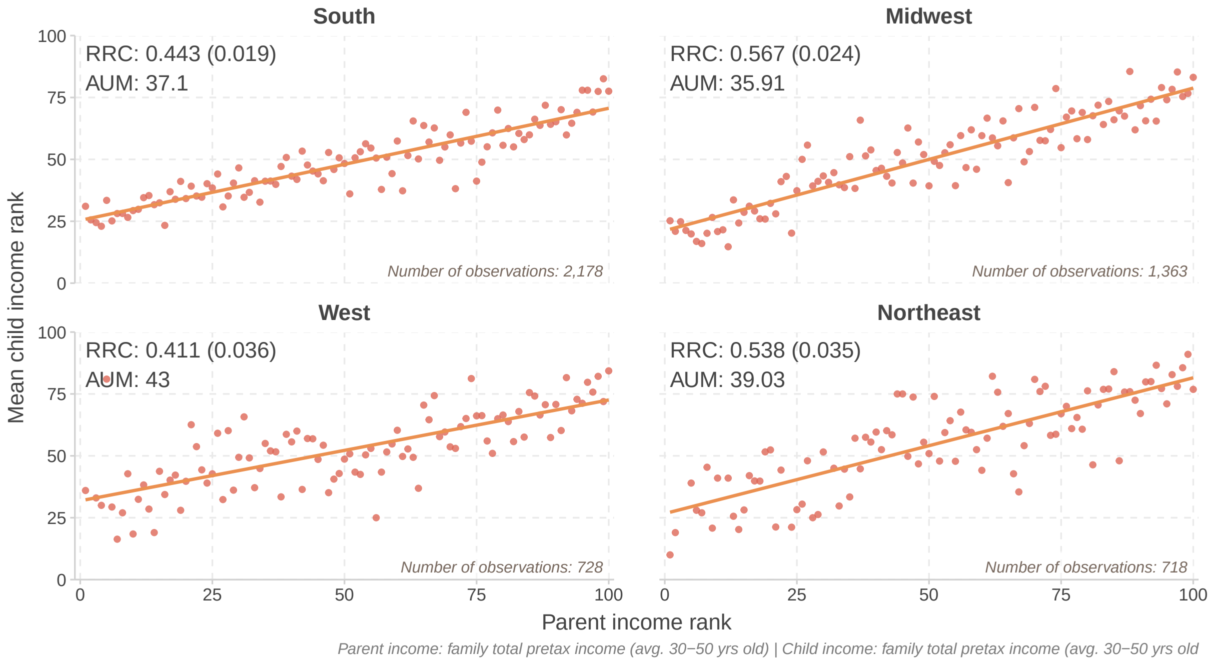


Figure B.2: Benchmark Rank-Rank Correlation and Absolute Upward Mobility by Census Region

Notes: This figure presents Census Region-level estimates of the rank-rank correlation (RRC) and absolute upward mobility (AUM). To compute local estimates, individuals are assigned to their most common Census Region of residence until age 18 (included). See Appendix Figure B.1's notes for details on data, sample and income definitions.

B.4 OLS vs. TSTSLS Comparison

We now turn to the comparison between estimates obtained with OLS and those obtained with TSTSLS. The PSID enables us to compare estimates of intergenerational mobility we obtain when observing parents' incomes and when predicting them using observable characteristics such as education and occupation. Since in the main analysis and in virtually all TSTSLS studies only parents' labor incomes or wages are observed, we define parents' income as individual labor income, while keeping in mind the benchmark estimates presented in the previous section. We follow the main analysis' definitions as closely as possible. We proceed in the following way.

B.4.1 Parent Income Prediction

Let Z denote a set of characteristics observed for parents. We can express their labor incomes y as $y_i = \beta Z_i + \epsilon_i$. We estimate this first-stage equation by OLS on our sample of parents, and predict out of sample using a 5-fold cross-validation approach. Specifically, we split the sample of parents in five random subsamples of equal size, and for each subsample we predict income using the first-stage estimated on the remaining four subsamples. As such all predicted incomes are conceptually made from a random sample of parents taken from the same population. We see these out-of-sample predictions as imitating very closely settings in which researchers do not observe the actual parents' incomes but observe the incomes of other parents taken from the same population (i.e., with children born in the same years).

We define parent income y as log mean (individual) labor income over ages 30 to 50. Once we have predicted labor incomes for children's father and/or mother, we compute a measure of labor income at the household level as the sum of father and mother predicted labor incomes if we have identified two parents, and predicted labor income of the only parent otherwise.[9] We display parents' (out-of-sample) predicted labor incomes against observed labor incomes in Appendix Section B.7.3.

For our baseline results, we define Z in the most similar way as possible as to our paper. Specifically, Z includes (i) education (7 categories; highest years of school completed), (ii) 3-digit occupation (334 cat.; most common occupation, including inactivity status, between 30 and 50 years old), (iii) demographic characteristics (birth cohort, race (5 cat.; most recent observation)), and (iv) state fixed effects (most common state of residence between 30 and 50 years old). The precise details of the construction of each of these variables are described in Appendix Section B.6.2. This set of predictors departs from the ones used in the main analysis because (i) we were unable to find a cross-walk between the 3-digit classification and a 2-digit classification, (ii) nationality is not available in the PSID, and (iii) country of birth is not available in the PSID. We replaced these variables with race. In Appendix Table B.4 we present results when incrementally including these predictors and find that the TSTSLS bias stabilizes once occupation is included.

B.4.2 National Results

Appendix Figure B.3 presents the main results from our validation exercise. Our TSTSLS estimate of the RRC is 0.459. On the exact same sample the OLS estimate is 0.476. Our benchmark RRC from the previous section was 0.504. The TSTSLS estimate is therefore roughly 4% smaller than the OLS estimate on the same sample, and 9% smaller than the benchmark OLS estimate (defining parent income as parent family income). These differences are quite small relative to the large differences in RRC estimates observed across countries (as well as within country across studies). Moreover, and importantly, the TSTSLS estimate appears to *understate* persistence, suggesting it provides a lower bound for intergenerational persistence.

The TSTSLS estimates for the transition matrix also appear to represent upper bounds on intergenerational (upward) mobility. The P(Top 20% | Bot. 20%) is roughly 6% in the TSTSLS case and 4% in the OLS case (4% as well in the benchmark), P(Bot. 20% | Bot. 20%) is 40% vs. 44% (45%) and the P(Top 20% | Top 20%) is 44% vs. 46% (47%). In Appendix B.7.6 we show that the TSTSLS bias of the RRC is largely unaffected by the number of parent income observations used. Moreover, Table B.5 shows our results are qualitatively robust to (i) using nationally-representative Survey Research Center (SRC) sample of the PSID, (ii) computing parent and child incomes over ages 35-45 as in the main analysis, (iii) dropping income observations equal to zero when computing parent and child incomes, and (iv) accounting for household size in the income definitions (additional details in Section B.7.5). Moreover, Table B.6 shows that using the longitudinal or cross-sectional weights moderately increases the TSTSLS RRC downward bias.

[9] Results when dividing by the number of parents are presented in Appendix Table B.5 (col. (5)).

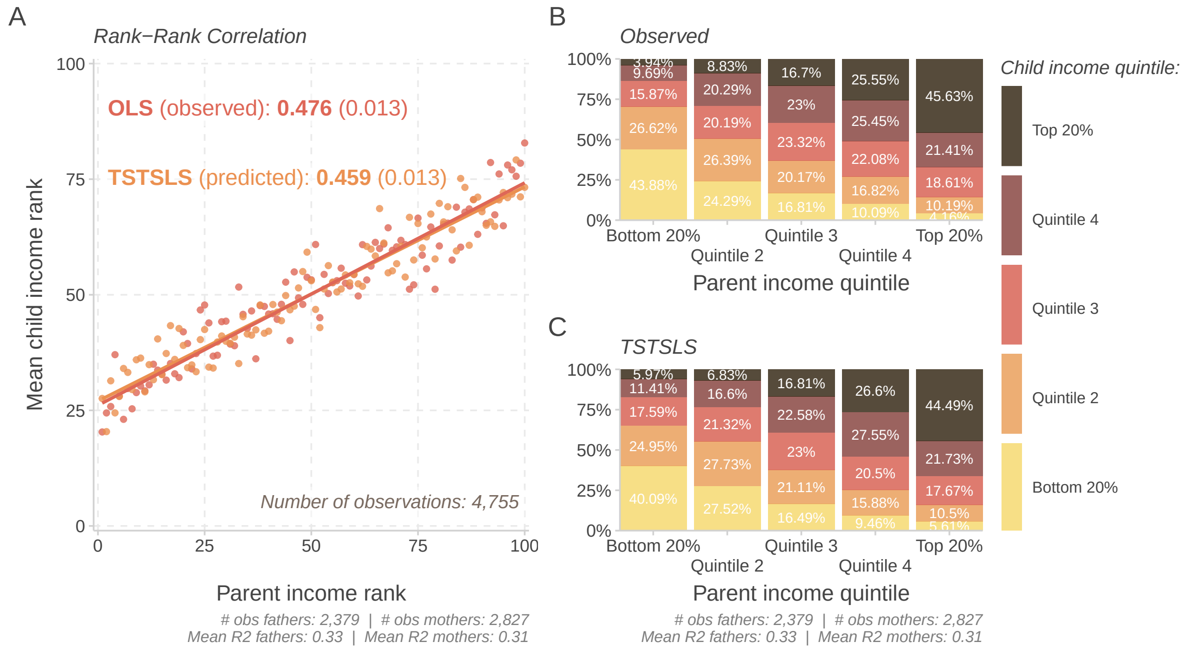


Figure B.3: OLS vs. TSTSLS RRC and Transition Matrix

Notes: This figure presents the rank-rank correlation (panel A) and the transition matrix (panels B and C) obtained when parent income is observed (OLS/observed) and when it is predicted using two-sample two-stage least squares (TSTSLS). It is computed on the Panel Study of Income Dynamics (PSID). The sample used is restricted to children born between 1963 and 1988 who are observed at least once as children in a family unit and at least once as a reference person or partner in a family unit over ages 30-50. Child income is the mean of family total income over ages 30-50. Parent income is the sum of father and mother (predicted) mean labor income over ages 30-50. For TSTSLS estimates, parent income is predicted separately for males and females using an OLS model including education (7 cat.; highest years of school completed), 3-digit occupation (334 cat.; most common occupation (incl. inactivity status) between 30 and 50 years old), demographic characteristics in 1990 (birth cohort and race (5 cat.; most recent observation) and state fixed effects (most common state of residence between 30 and 50 years old). In panel A, the fitted line is a linear fit through the microdata. We report coefficients and naive standard errors (in parenthesis) obtained from OLS regressions of child income rank on parent income rank with child cohort fixed effects, on the microdata for the full sample.

B.4.3 Regional Results

Appendix Figure B.4 shows the results obtained by Census Region. The RRC obtained by TSTSLS is remarkably similar to that obtained by OLS, with a slight underestimation for the Northeast and West regions. The same applies to the AUM which again is very similar in the TSTSLS setting relative to the OLS case (and the benchmarks). Compared to the benchmark estimates presented in the previous section, differences in RRCs are a bit larger but the rank-ordering of regions is preserved.

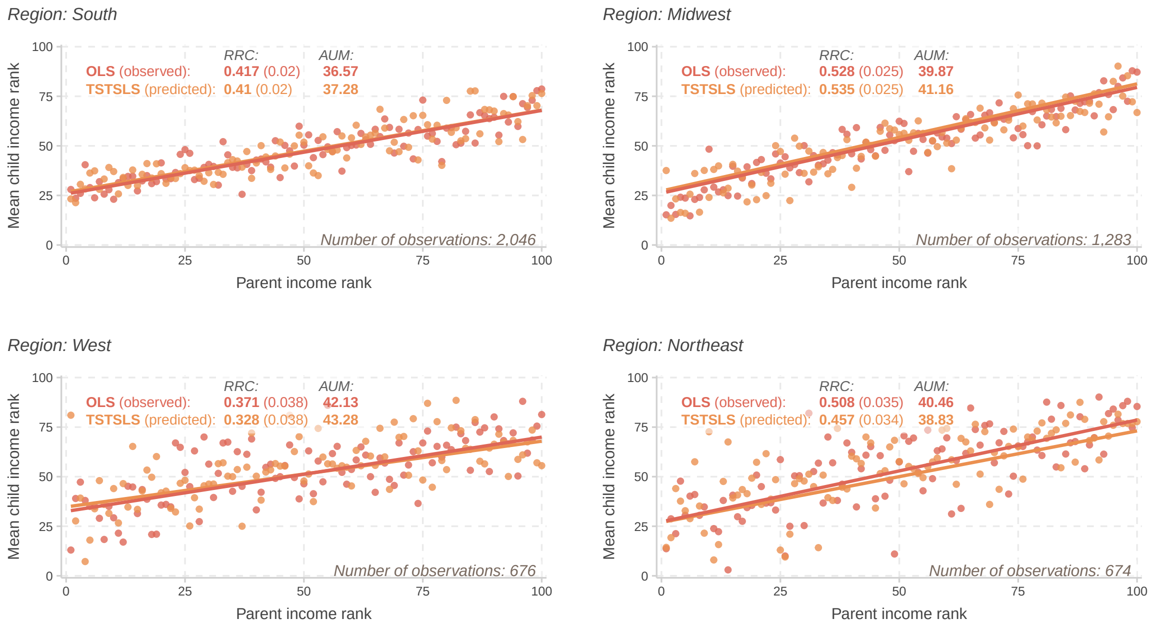


Figure B.4: OLS vs. TSTSLS RRC and AUM

Notes: This figure presents Census Region-level estimates of the rank-rank correlation (RRC) and absolute upward mobility (AUM). To compute local estimates, individuals are assigned to their most common Census Region of residence until age 18 (included). See Appendix Figure B.3's notes for details on data, sample and income definitions.

B.5 Discussion

Overall, the results presented in this analysis suggest that using TSTSLS for rank-based measures of intergenerational mobility leads to reasonably close estimates relative to OLS estimates, both at the national and subnational levels. Specifically TSTSLS estimates appear to slightly underestimate intergenerational persistence, from 4% to 10% depending on the set of predictors (see Appendix Section B.7.4 for all results when varying the set of first-stage predictors). Moreover, they seem to represent lower bounds for intergenerational persistence (i.e., upper bounds for mobility). In Appendix Table B.5, we show these findings are also robust to dropping income observations equal to 0, as well as to accounting for the number of reference person and partner when defining incomes for children and parents.

B.6 Details on Samples and Variable Definitions

B.6.1 Sample Construction Details

Table B.2: Sample Size at Each Restriction

	# obs.	%
Raw sample	82,573	100
+ born 1963-1988	30,186	36.56
+ observed at least once as child in a family unit	18,612	61.66
+ observed at least once as head/spouse 30-50	5,655	30.38
+ at least one family total income observation 30-50	5,484	96.98
+ at least one observation for parent total income observation 30-50	5,088	92.78

Notes: child and parent income observations exclude those obtained by "major assignment".

B.6.2 Details on Variable Constructions

Age: since prior to 1983, only age (rather than birth year) was reported, we use the following rule to obtain individuals' birth year: (i) if at least 1 birth year value: most common value; (ii) otherwise: most common value obtained from year - age (by definition this will equal birth year or birth year + 1).

Parent education: maximum grade completed over all observations, and classified following Jerrim et al. (2016) / PSID classification of grades into education levels.[10]
Categories: Grades 1-5, Grades 6-8, Grades 9-11, Grade 12 (HS completion), Some college / associate degree (grades 13-15), College degree (grade 16), Advanced college degree (grade 17).

Parent occupation: most common 3-digit occupation (1970 classification) or detailed inactivity status between 30 and 50 years old.[11] Occupation variables with a consistent classification are available for all individuals between 1981 and 2001, and are only available for a selected sample of PSID heads and wives/"wives"[12] between 1968 and 1980. In order to prevent bias from focusing only on employed parents[13], we use information from employment status variables from 1981 onwards.[14]
Categories: 441 3-digit occupations + 5 detailed inactivity status (Unemployed, Housewife, Student, Retired/Permanently disabled, Other).

[10] Note the grade completed variable is missing for 1969.

[11] In cases where an individual has several most common occupations, we assign the one for which the individual is the oldest on average, and choose one at random if average age is the same.

[12] Criteria: (i) original sample Heads and Wives/"Wives still living by 1992 who reported main jobs in at least three waves during the period 1968-1992, with at least one of those reports prior to 1980; and (ii) additionally, original sample Heads and Wives/"Wives" who had reported at least one main job between 1968 and 1980 but were known to have died by 1992. Those who were still living but had reported only one or two jobs during the period of interest were excluded, as were all nonsample Heads and Wives/"Wives".

[13] By definition, occupations are only available for employed individuals.

[14] Employment status is only available for heads between 1968 and 1978; from 1979 onwards, it is available for heads and wives/"wives". To prevent any bias, employment status is used only after 1980, i.e., when occupation is not restricted to a selected sample.

Parent race: most recent race observation.
Categories: White, African American, Asian/Pacific Islander, Native American, Other.

Parent region: most common state between 30 and 50 years old.

Child region: most common state between 0 and 18 years old.

Table B.3: Detail of Variables Used in Each Year

								Occupation			Race			Labor income						
Year	Int. # (ind. file)	Int. # (fam. files)	Sequence #	Rel. to ref. person	Age	Birth year	Education	Ref. person	Partner	Employment status	Ref. person	Partner	State of res.	Ref. person	Accu.	Partner	Accu.	Tax. inc.	Accu. tax. inc.	Total transfers
1968	ER30001	V3	-	ER30003	ER30004	-	ER30010	V197_A	V243_A	-	V181	-	V93	V74	V85	V75	V86	V76	V87	-
1969	ER30020	V442	ER30021	ER30022	ER30023	-	-	V640_A	V609_A	-	V801	-	V537	V514	V515	V516	V517	V518	V519	-
1970	ER30043	V1102	ER30044	ER30045	ER30046	-	ER30052	V1279_A	V1367_A	-	V1490	-	V1103	V1196	V1192	V1198	V1199	V1205	V1206	V1220
1971	ER30067	V1802	ER30068	ER30069	ER30070	-	ER30076	V1984_A	V2074_A	-	V2202	-	V1803	V1897	V1893	V1899	V1900	V1906	V1907	V1922
1972	ER30091	V2402	ER30092	ER30093	ER30094	-	ER30100	V2582_A	V2672_A	-	V2828	-	V2403	V2498	V2494	V2500	V2501	V2507	V2508	V2523
1973	ER30117	V3002	ER30118	ER30119	ER30120	-	ER30126	V3115_A	V3183_A	-	V3300	-	V3003	V3051	V3047	V3053	V3054	V3060	V3061	V3076
1974	ER30138	V3402	ER30139	ER30140	ER30141	-	ER30147	V3968_A	V3601_A	-	V3720	-	V3403	V3463	V3459	V3465	V3466	V3472	V3473	V3488
1975	ER30160	V3802	ER30161	ER30162	ER30163	-	ER30169	V4459_A	V4055_A	-	V4204	-	V3803	V3863	V3859	V3865	V3866	V3872	V3873	V3889
1976	ER30188	V4302	ER30189	ER30190	ER30191	-	ER30197	V5374_A	V4605_A	-	V5096	-	V4303	V5031	V4374	V4379	V4380	V4386	V4387	V4404
1977	ER30217	V5202	ER30218	ER30219	ER30220	-	ER30226	V5873_A	V5507_A	-	V5662	-	V5203	V5627	V5284	V5289	V5290	V5297	V5298	V5316
1978	ER30246	V5702	ER30247	ER30248	ER30249	-	ER30255	V6497_A	V6039_A	-	V6209	-	V5703	V6174	V5783	V5788	V5789	V5796	V5797	V5815
1979	ER30283	V6302	ER30284	ER30285	ER30286	-	ER30296	V7100_A	V6596_A	ER30293	V6802	-	V6303	V6767	V6392	V6398	V6399	V6408	V6407	V6426
1980	ER30313	V6902	ER30314	ER30315	ER30316	-	ER30326	V3529	V7198_A	ER30323	V7447	-	V6903	V7413	V6982	V6988	V6989	V6998	V6997	V7016
1981	ER30343	V7502	ER30344	ER30345	ER30346	-	ER30356	V7712	V7885	ER30353	V8099	-	V7503	V8066	V7574	V7580	V7581	V7590	V7589	V7608
1982	ER30373	V8202	ER30374	ER30375	ER30376	-	ER30384	V8380	V8544	ER30382	V8723	-	V8203	V8690	V8266	V8273	V8274	V8283	V8282	V8301
1983	ER30399	V8802	ER30400	ER30401	-	ER30404	ER30413	V9011	V9194	ER30411	V9408	-	V8803	V9376	V8874	V8881	V8882	V8891	V8890	V8909
1984	ER30429	V10002	ER30430	ER30431	-	ER30434	ER30443	V10460	V10678	ER30441	V11055	-	V10003	V11023	V10257	V10263	V10264	V10277	V10276	V10305
1985	ER30463	V11102	ER30464	ER30465	-	ER30468	ER30478	V11651	V12014	ER30474	V11938	V12293	V11103	V12372	V11398	V11404	V11405	V11419	V11418	V11461
1986	ER30498	V12502	ER30499	ER30500	-	ER30503	ER30513	V13054	V13233	ER30509	V13565	V13500	V12503	V13624	V12797	V12803	V12804	V12818	V12817	V12868
1987	ER30535	V13702	ER30536	ER30537	-	ER30540	ER30549	V14154	V14329	ER30545	V14612	V14547	V13703	V14671	V13899	V13905	V13906	V13920	V13919	V13970
1988	ER30570	V14802	ER30571	ER30572	-	ER30575	ER30584	V15162	V15464	ER30580	V16086	V16021	V14803	V16145	V14914	V14920	V14921	V14935	V14934	V14985
1989	ER30606	V16302	ER30607	ER30608	-	ER30611	ER30620	V16663	V16982	ER30616	V17483	V17418	V16303	V17534	V16414	V16420	V16421	V16435	V16434	V16485
1990	ER30642	V17702	ER30643	ER30644	-	ER30647	ER30657	V18101	V18403	ER30653	V18814	V18749	V17703	V18878	V17830	V17836	V17837	V17851	V17850	V17901
1991	ER30689	V19002	ER30690	ER30691	-	ER30694	ER30703	V19401	V19703	ER30699	V20114	V20049	V19003	V20178	V19130	V19136	V19137	V19151	V19150	V19201
1992	ER30733	V20302	ER30734	ER30735	-	ER30738	ER30748	V20701	V21003	ER30744	V21420	V21355	V20303	V21484	V20430	V20436	V20437	V20451	V20450	V20501
1993	ER30806	V21602	ER30807	ER30808	-	ER30811	ER30820	V22456	V22809	ER30816	V23276	V23212	V21603	V23323	V21740	V23324	-	V21959	-	V22366
1994	ER33101	ER2002	ER33102	ER33103	-	ER33106	ER33115	ER4017	ER4048	ER33111	ER3944	ER3883	ER4156	ER4140	-	ER4144	ER4145	ER4146	-	ER4147
1995	ER33201	ER5002	ER33202	ER33203	-	ER33206	ER33215	ER6857	ER6888	ER33211	ER6814	ER6753	ER6996	ER6980	-	ER6984	ER6985	ER6986	-	ER6987
1996	ER33301	ER7002	ER33302	ER33303	-	ER33306	ER33315	ER9108	ER9139	ER33311	ER9060	ER8999	ER9247	ER9231	-	ER9235	ER9236	ER9237	-	ER9238
1997	ER33401	ER10002	ER33402	ER33403	-	ER33406	ER33415	ER12085	ER12116	ER33411	ER11848	ER11760	ER12221	ER12080	ER12081	ER12082	-	ER12069	ER12070	ER12071
1999	ER33501	ER13002	ER33502	ER33503	-	ER33506	ER33516	ER13215	ER13727	ER33512	ER15928	ER15836	ER13004	ER16463	ER16464	ER16465	-	ER16452	ER16453	ER16454
2001	ER33601	ER17002	ER33602	ER33603	-	ER33606	ER33616	ER17226	ER17796	ER33612	ER19989	ER19897	ER17004	ER20443	ER18562	ER20447	ER20448	ER20449	-	ER20450
2003	ER33701	ER21002	ER33702	ER33703	-	ER33706	ER33716	-	-	ER33712	ER23426	ER23334	ER21003	ER24116	ER21930	ER24135	ER24136	ER24100	-	ER24101
2005	ER33801	ER25002	ER33802	ER33803	-	ER33806	ER33817	-	-	ER33813	ER27393	ER27297	ER25003	ER27931	ER25911	ER27943	ER27944	ER27953	-	ER28002
2007	ER33901	ER36002	ER33902	ER33903	-	ER33906	ER33917	-	-	ER33913	ER40565	ER40472	ER36003	ER40921	ER36929	ER40933	ER40934	ER40943	-	ER40992
2009	ER34001	ER42002	ER34002	ER34003	-	ER34006	ER34020	-	-	ER34016	ER46543	ER46449	ER42003	ER46829	ER42920	ER46841	ER46842	ER46851	-	ER46900
2011	ER34101	ER47302	ER34102	ER34103	-	ER34106	ER34119	-	-	ER34116	ER51904	ER51810	ER47303	ER52237	ER48242	ER52249	ER52250	ER52259	-	ER52308
2013	ER34201	ER53002	ER34202	ER34203	-	ER34206	ER34230	-	-	ER34216	ER57659	ER57549	ER53003	ER58038	ER53936	ER58050	ER58051	ER58060	-	ER58117
2015	ER34301	ER60002	ER34302	ER34303	-	ER34307	ER34349	-	-	ER34317	ER64810	ER64671	ER60003	ER65216	ER60995	ER65244	-	ER65253	-	ER65314
2017	ER34501	ER66002	ER34502	ER34503	-	ER34506	ER34548	-	-	ER34516	ER70882	ER70744	ER66003	ER71293	ER67047	ER71321	-	ER71330	-	ER71391
2019	ER34701	ER72002	ER34702	ER34703	-	ER34706	ER34752	-	-	ER34716	ER76897	ER76752	ER72003	ER77315	ER73070	ER77343	-	ER77352	-	ER77413

Notes: Int. = interview. Ind. = individual. Fam. = family. Rel. = relation. Ref. = reference. Res. = residence. Accu. = accuracy. Tax. inc. = taxable income.

B.7 Additional Results

B.7.1 All Benchmark Estimates

Parent income definition \ Child income definition	Family labor income (div. number of adults) (30−50)	Family labor income (30−50)	Family taxable income (div. number of adults) (30−50)	Family taxable income (30−50)	Family total income (div. number of adults) (30−50)	Family total income (30−50)	Individual labor incor (30−50)
Individual labor income (sum) (30−50)	0.316 (0.011)	0.363 (0.012)	0.313 (0.012)	0.359 (0.013)	0.282 (0.009)	0.332 (0.011)	0.295 (0.013)
Individual labor income (mean) (30−50)	0.359 (0.012)	0.405 (0.013)	0.355 (0.014)	0.4 (0.015)	0.32 (0.011)	0.369 (0.012)	0.338 (0.015)
Family total income (30−50)	0.531 (0.016)	0.607 (0.017)	0.547 (0.018)	0.622 (0.019)	0.503 (0.014)	0.584 (0.015)	0.487 (0.019)
Family total income (div. number of adults) (30−50)	0.612 (0.018)	0.682 (0.02)	0.628 (0.021)	0.698 (0.023)	0.584 (0.016)	0.659 (0.018)	0.562 (0.023)
Family taxable income (30−50)	0.348 (0.011)	0.4 (0.012)	0.353 (0.013)	0.406 (0.014)	0.319 (0.01)	0.375 (0.011)	0.323 (0.014)
Family taxable income (div. number of adults) (30−50)	0.38 (0.012)	0.431 (0.013)	0.384 (0.014)	0.435 (0.015)	0.348 (0.011)	0.403 (0.012)	0.355 (0.015)
Family labor income (30−50)	0.369 (0.012)	0.425 (0.013)	0.373 (0.013)	0.429 (0.014)	0.333 (0.01)	0.392 (0.012)	0.343 (0.015)
Family labor income (div. number of adults) (30−50)	0.405 (0.013)	0.458 (0.014)	0.408 (0.015)	0.461 (0.016)	0.364 (0.011)	0.421 (0.013)	0.377 (0.016)

Naive standard errors in parenthesis. Number of children varies by income defintion since the number of negative or zero incomes varies.

Figure B.5: Benchmark IGEs for All Income Definitions

Parent income definition \ Child income definition	Family labor income (div. number of adults) (30−50)	Family labor income (30−50)	Family taxable income (div. number of adults) (30−50)	Family taxable income (30−50)	Family total income (div. number of adults) (30−50)	Family total income (30−50)	Individual labor incor (30−50)
Individual labor income (sum) (30−50)	0.465 (0.012)	0.47 (0.012)	0.47 (0.012)	0.474 (0.012)	0.476 (0.012)	0.48 (0.012)	0.371 (0.013)
Individual labor income (mean) (30−50)	0.458 (0.012)	0.453 (0.013)	0.463 (0.012)	0.456 (0.013)	0.47 (0.012)	0.462 (0.012)	0.37 (0.013)
Family total income (30−50)	0.487 (0.012)	0.491 (0.012)	0.495 (0.012)	0.497 (0.012)	0.503 (0.012)	0.504 (0.012)	0.386 (0.013)
Family total income (div. number of adults) (30−50)	0.479 (0.012)	0.471 (0.012)	0.486 (0.012)	0.476 (0.012)	0.495 (0.012)	0.483 (0.012)	0.38 (0.013)
Family taxable income (30−50)	0.491 (0.012)	0.496 (0.012)	0.499 (0.012)	0.502 (0.012)	0.506 (0.012)	0.509 (0.012)	0.39 (0.013)
Family taxable income (div. number of adults) (30−50)	0.485 (0.012)	0.481 (0.012)	0.493 (0.012)	0.487 (0.012)	0.501 (0.012)	0.493 (0.012)	0.386 (0.013)
Family labor income (30−50)	0.48 (0.012)	0.483 (0.012)	0.484 (0.012)	0.486 (0.012)	0.491 (0.012)	0.492 (0.012)	0.385 (0.013)
Family labor income (div. number of adults) (30−50)	0.472 (0.012)	0.465 (0.012)	0.475 (0.012)	0.467 (0.012)	0.482 (0.012)	0.473 (0.012)	0.38 (0.013)

Naive standard errors in parenthesis. Number of children: 5,088.

Figure B.6: Benchmark RRCs for All Income Definitions

B.7.2 Only SRC Sample

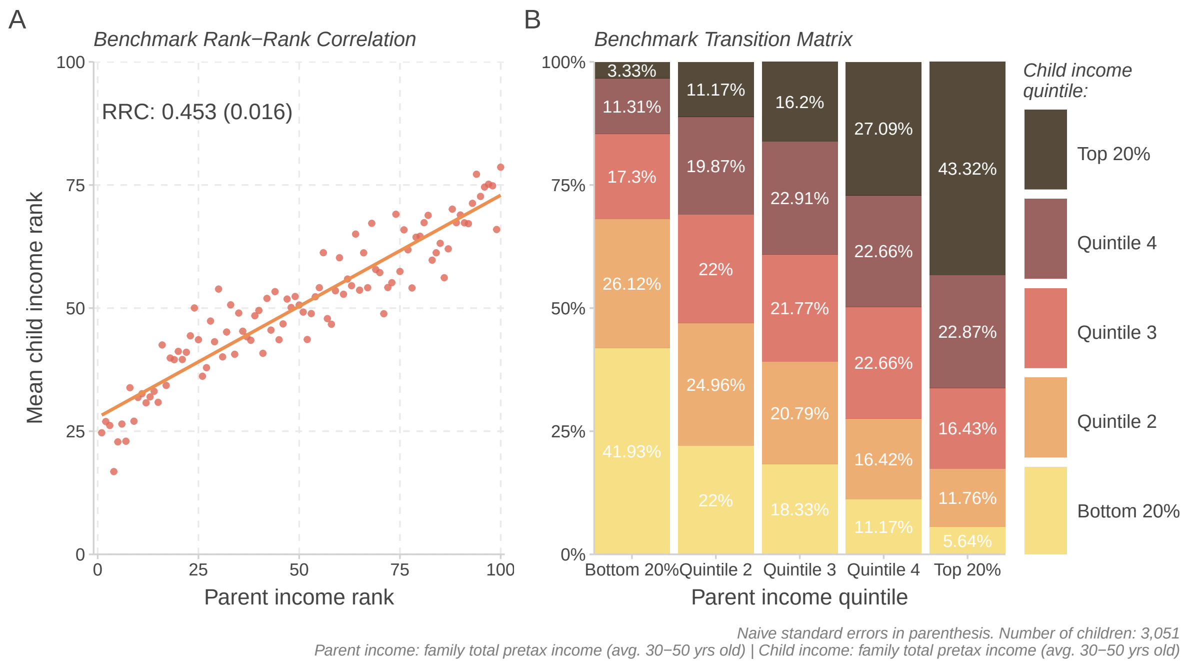


Figure B.7: Benchmark Rank-Rank Correlation and Transition Matrix

Notes: This figure presents the rank-rank correlation (panel A) and the transition matrix (panel B), computed on the Panel Study of Income Dynamics (PSID)'s representative Survey Research Center sample. See Appendix Figure B.1's notes for details on data, sample and income definitions.

B.7.3 Baseline Predictions

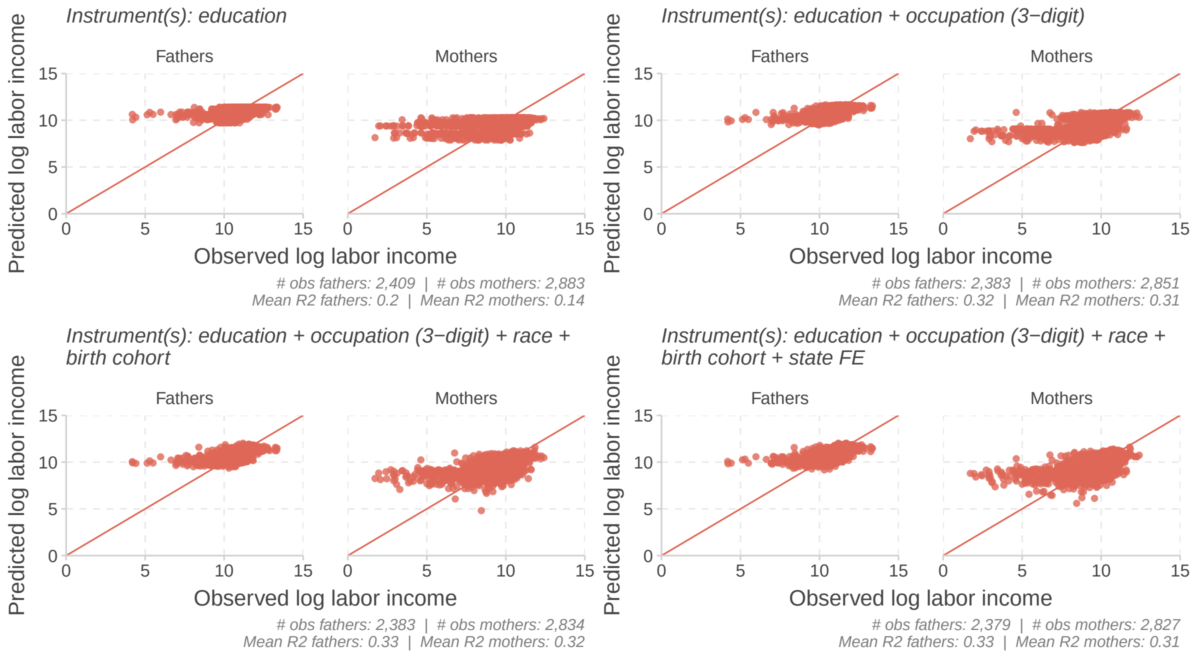


Figure B.8: Observed vs. (out-of-sample) predicted *individual* labor income

Notes: This figure presents observed *individual* log labor income and out-of-sample predicted *individual* log labor income for fathers and mothers depending on variables used in the first-stage prediction. The red line corresponds to the 45 degree line. See Appendix Figure B.3's notes for details on data, sample and income definitions.

Figure B.9: Observed vs. (out-of-sample) predicted *individual* labor income rank

Notes: This figure presents the conditional expectation of out-of-sample predicted *individual* labor income rank, as a function of observed *individual* labor income rank, for fathers and mothers, depending on variables used in the first-stage prediction. The red line corresponds to the 45 degree line, while the red shaded area corresponds to the interquartile range. See Appendix Figure B.3's notes for details on data, sample and income definitions.

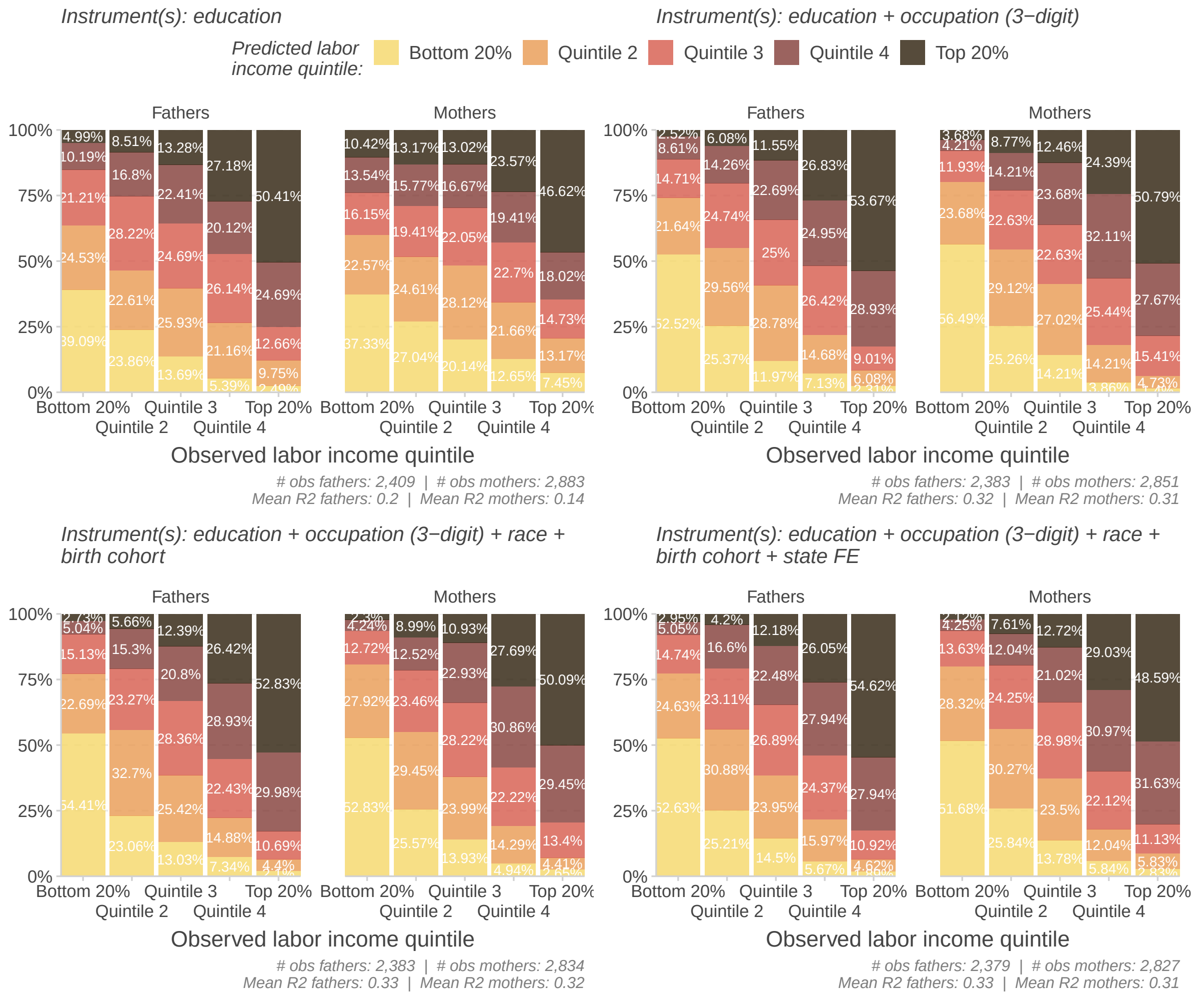


Figure B.10: Observed vs. (out-of-sample) predicted *individual* labor income quintile

Notes: This figure presents the quintile-by-quintile out-of-sample predicted *individual* labor income quintile by observed *individual* labor income quintile, for fathers and mothers, depending on variables used in the first-stage prediction. Each cell documents the share of out-of-sample labor income predictions belonging to the quintile indicated by the color legend among observed labor incomes falling in the quintile indicated on the x-axis. They are computed separately for father and mother, and depending on variables used in the first-stage prediction. See Appendix Figure B.3's notes for details on data, sample and income definitions.

Figure B.11: Observed vs. (out-of-sample) predicted *family* labor income

Notes: This figure presents observed *family* log labor income and out-of-sample predicted *family* log labor income for fathers and mothers depending on variables used in the first-stage prediction. The red line corresponds to the 45 degree line. See Appendix Figure B.3's notes for details on data, sample and income definitions.

Figure B.12: Observed vs. (out-of-sample) predicted *family* labor income rank

Notes: This figure presents the conditional expectation of out-of-sample predicted *family* labor income rank, as a function of observed *family* labor income rank, for fathers and mothers, depending on variables used in the first-stage prediction. The red line corresponds to the 45 degree line, while the red shaded area corresponds to the interquartile range. See Appendix Figure B.3's notes for details on data, sample and income definitions.

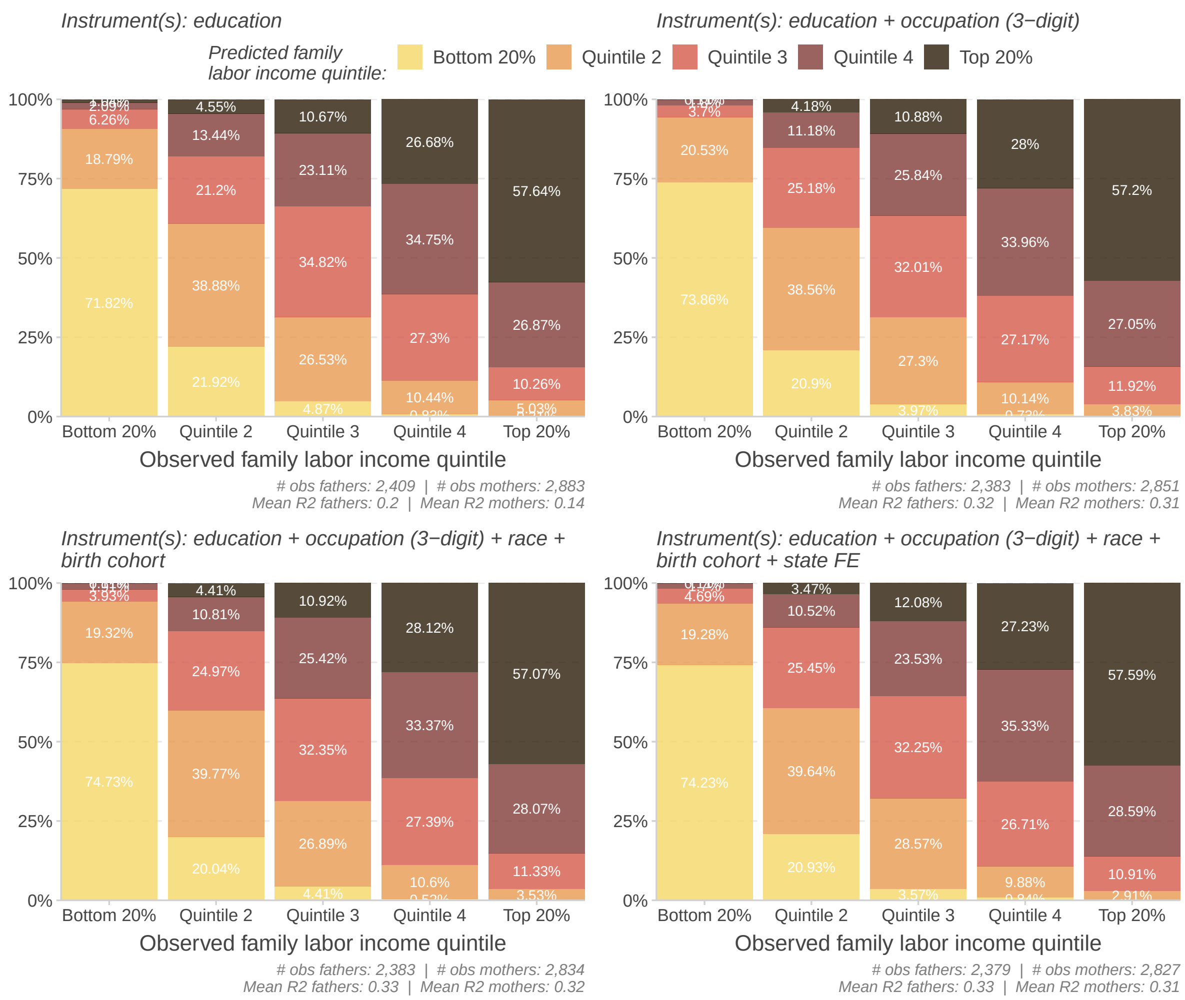


Figure B.13: Observed vs. (out-of-sample) predicted *family* labor income rank

Notes: This figure presents the quintile-by-quintile out-of-sample predicted *family* labor income quintile by observed *family* labor income quintile, for fathers and mothers, depending on variables used in the first-stage prediction. Each cell documents the share of out-of-sample labor income predictions belonging to the quintile indicated by the color legend among observed labor incomes falling in the quintile indicated on the x-axis. They are computed separately for father and mother, and depending on variables used in the first-stage prediction. See Appendix Figure B.3's notes for details on data, sample and income definitions.

B.7.4 Alternative First-Stage Predictors

Table B.4: Comparison for Different Sets of Predictors

	Education (1)	+ occupation (3-digit) (2)	+ race + birth cohort (3)	+ state FE (4)
Panel A. Intergenerational Elasticity (IGE)				
Observed parent income (OLS)	0.335	0.334	0.334	0.334
	(0.011)	(0.011)	(0.011)	(0.011)
Predicted parent income (TSTSLS)	0.464	0.431	0.449	0.445
	(0.017)	(0.014)	(0.014)	(0.014)
Percentage diff. TSTSLS vs OLS	-27.76%	-22.57%	-25.73%	-24.82%
Number of observations	4,805	4,755	4,737	4,730
Panel B. Rank-Rank Correlation (RRC)				
Observed parent income (OLS)	0.476	0.475	0.476	0.476
	(0.013)	(0.013)	(0.013)	(0.013)
Predicted parent income (TSTSLS)	0.43	0.453	0.461	0.459
	(0.013)	(0.013)	(0.013)	(0.013)
Percentage diff. TSTSLS vs OLS	10.53%	4.9%	3.22%	3.85%
Number of observations	4,832	4,780	4,762	4,755
Panel C. Transition Matrix				
P(Bottom 20% ∣ Bottom 20%) (OLS)	43.95%	43.7%	43.84%	43.88%
P(Bottom 20% ∣ Bottom 20%) (TSTSLS)	37.83%	39.77%	40.66%	40.09%
P(Bottom 20% ∣ Top 20%) (OLS)	4.51%	4.46%	4.26%	4.16%
P(Bottom 20% ∣ Top 20%) (TSTSLS)	4.81%	5.18%	5.39%	5.61%
P(Top 20% ∣ Bottom 20%) (OLS)	3.97%	3.92%	3.93%	3.94%
P(Top 20% ∣ Bottom 20%) (TSTSLS)	5.22%	5.83%	5.63%	5.97%
P(Top 20% ∣ Top 20%) (OLS)	45.54%	45.49%	45.53%	45.63%
P(Top 20% ∣ Top 20%) (TSTSLS)	44.22%	43.32%	43.36%	44.49%

Notes:

B.7.5 Alternative Samples and Definitions

In Table B.5 we check the robustness of our baseline results to changes in estimation samples and income definitions. Specifically, we report results for the following changes: (i) using only the nationally-representative Survey Research Center (SRC) sample (col. 2), (ii) restricting the age range over which child and parent incomes are averaged to 35-45 years old in our main analysis, (iii) dropping parent and child income observations equal to zero when computing average incomes[15], (iv) accounting for household size when defining parent and child incomes (see discussion in B.2). Our baseline results are reported in column 1.

[15] According to Mazumder (2016, p.101): "In the PSID, the household head is recorded as having zero labor income if their income was actually zero or if their labor income is missing, so one cannot cleanly distinguish true zeroes with labor income."

Table B.5: Robustness of Baseline Results

	Baseline Estimates (1)	Only SRC Sample (2)	35-45 Income Age Range (3)	Dropping Zero Inc. Obs. (4)	Accounting Household Size (5)
Panel A. National - Intergenerational Elasticity (IGE)					
Observed parent income (OLS)	0.334	0.369	0.363	0.414	0.324
	(0.011)	(0.018)	(0.016)	(0.013)	(0.011)
Predicted parent income (TSTSLS)	0.445	0.418	0.475	0.53	0.485
	(0.014)	(0.022)	(0.021)	(0.017)	(0.016)
Percentage diff. TSTSLS vs OLS	-24.82%	-11.72%	-23.53%	-21.9%	-33.21%
Number of observations	4,730	2,892	2,882	4,732	4,730
Panel B. National - Rank-Rank Correlation (RRC)					
Observed parent income (OLS)	0.476	0.409	0.464	0.47	0.466
	(0.013)	(0.017)	(0.017)	(0.013)	(0.013)
Predicted parent income (TSTSLS)	0.459	0.364	0.448	0.463	0.435
	(0.013)	(0.017)	(0.017)	(0.013)	(0.013)
Percentage diff. TSTSLS vs OLS	3.85%	12.38%	3.56%	1.7%	7.16%
Number of observations	4,755	2,903	2,903	4,732	4,755
Panel C. Region: Midwest					
RRC - OLS	0.528	0.417	0.506	0.523	0.509
	(0.025)	(0.03)	(0.031)	(0.025)	(0.025)
AUM - OLS	39.87	43.94	38.49	39.65	37.19
RRC - TSTSLS	0.535	0.37	0.506	0.521	0.497
	(0.025)	(0.032)	(0.032)	(0.025)	(0.026)
AUM - TSTSLS	41.16	46.85	40.29	41.43	39.01
RRC percentage diff. TSTSLS vs OLS	-1.22%	12.49%	0.02%	0.32%	2.52%
AUM percentage diff. TSTSLS vs OLS	-3.15%	-6.22%	-4.47%	-4.31%	-4.67%
Number of observations	1,283	980	834	1,277	1,283
Panel D. Region: Northeast					
RRC - OLS	0.508	0.429	0.457	0.503	0.52
	(0.035)	(0.04)	(0.048)	(0.035)	(0.036)
AUM - OLS	40.46	41.91	46.3	40.23	44.23
RRC - TSTSLS	0.457	0.35	0.377	0.459	0.46
	(0.034)	(0.039)	(0.045)	(0.033)	(0.035)
AUM - TSTSLS	38.83	41.49	46.9	37.85	42.34
RRC percentage diff. TSTSLS vs OLS	11.25%	22.4%	21.26%	9.6%	13.13%
AUM percentage diff. TSTSLS vs OLS	4.19%	1%	-1.28%	6.3%	4.44%
Number of observations	674	538	406	669	674
Panel E. Region: South					
RRC - OLS	0.417	0.398	0.423	0.414	0.413
	(0.02)	(0.03)	(0.025)	(0.02)	(0.02)
AUM - OLS	36.57	36.19	35.4	37.17	36.71
RRC - TSTSLS	0.41	0.401	0.421	0.42	0.386
	(0.02)	(0.03)	(0.026)	(0.021)	(0.02)
AUM - TSTSLS	37.28	35.01	35.34	37.47	38.12
RRC percentage diff. TSTSLS vs OLS	1.72%	-0.75%	0.47%	-1.5%	7.03%
AUM percentage diff. TSTSLS vs OLS	-1.92%	3.37%	0.15%	-0.8%	-3.71%
Number of observations	2,046	885	1,242	2,036	2,046
Panel F. Region: West					
RRC - OLS	0.371	0.299	0.321	0.357	0.353
	(0.038)	(0.047)	(0.053)	(0.037)	(0.037)
AUM - OLS	42.13	40.03	45.57	42.11	42.68
RRC - TSTSLS	0.328	0.232	0.303	0.344	0.303
	(0.038)	(0.046)	(0.052)	(0.038)	(0.038)
AUM - TSTSLS	43.28	43.43	46.18	42.98	43.66
RRC percentage diff. TSTSLS vs OLS	12.97%	28.89%	5.93%	3.8%	16.65%
AUM percentage diff. TSTSLS vs OLS	-2.66%	-7.84%	-1.31%	-2.02%	-2.23%
Number of observations	676	488	376	674	676

B.7.6 Attenuation Bias

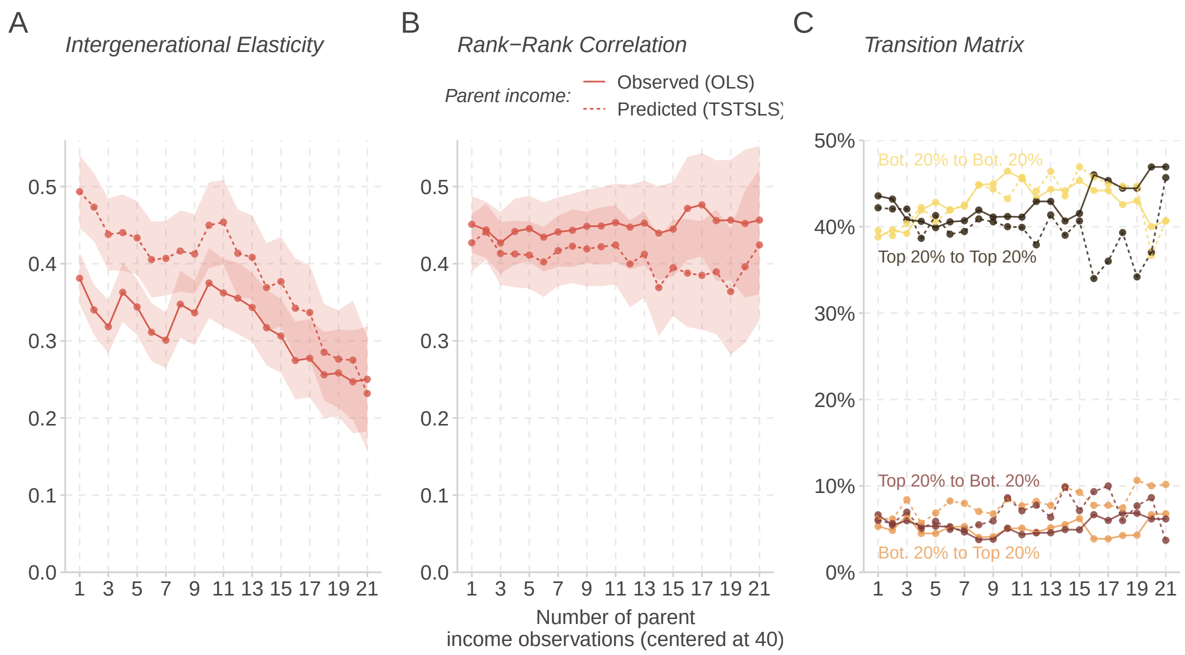


Figure B.14: OLS vs. TSTSLS Estimates - Varying Number of Parent Income Observations

Notes: This figure presents the IGE, RRC and transition matrix cells obtained when parent income is observed (OLS) and when it is predicted using two-sample two-stage least squares (TSTSLS), for different number of parent income observations. To control for the potential effect of lifecycle bias, we center parent incomes around age 40. Thus one observation means parent income is equal to income at age 40, two observations means parent income is equal to income averaged over age 39 and age 41, three observations means parent income is equal to income averaged over age 39 to age 41, etc. See Appendix Figure B.3's notes for details on data, sample and income definitions.

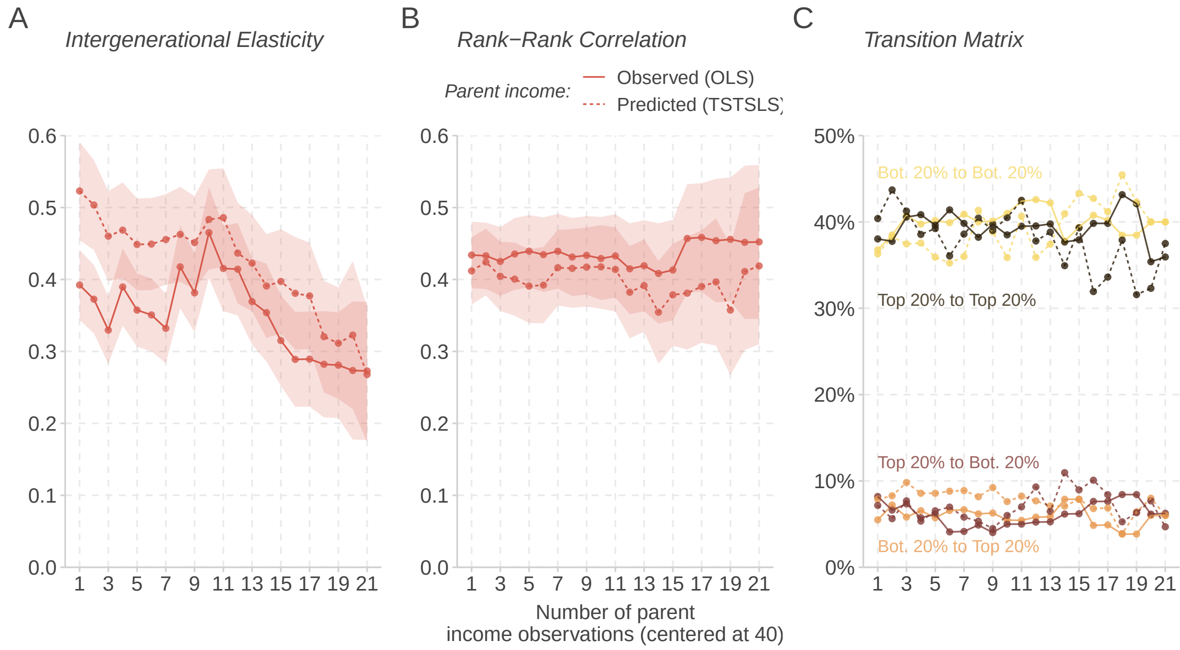


Figure B.15: OLS vs. TSTSLS Estimates - Varying Number of Parent Income Observations - Child Income Mean 37-43

Notes: This figure presents the IGE, RRC and transition matrix cells obtained when parent income is observed (OLS) and when it is predicted using two-sample two-stage least squares (TSTSLS), for different number of parent income observations and when child income is defined over ages 37-43. To control for the potential effect of lifecycle bias, we center parent incomes around age 40. Thus one observation means parent income is equal to income at age 40, two observations means parent income is equal to income averaged over age 39 and age 41, three observations means parent income is equal to income averaged over age 39 to age 41, etc. See Appendix Figure B.3's notes for details on data, sample and income definitions.

B.7.7 Sampling Weights

As is well-known, the PSID is not a nationally representative sample. In particular, the Survey of Economic Opportunity (SEO) component of the PSID oversamples low-income households but suffers from various sampling issues (see footnote 4 in Lee and Solon (2009)). In our baseline results, we opted to use all of the PSID because (i) our goal was to compare OLS to TSTSLS estimates rather than obtain the best OLS estimate, and (ii) the additional sample size allows us to compare OLS and TSTSLS estimates at the regional level. However, one may wish to know how our exercise performs for a nationally-representative sample. Table B.6 compares our baseline results with estimates obtained from four different specifications: (i) using only the PSID's nationally representative Survey Research Center (SRC) sample, (ii) using all of the PSID with three different kinds of weights, all measured in the child's last income observation year: (i) the family longitudinal weights, (ii) the individual longitudinal weights, and (iii) the individual cross-sectional weights (only available from 1997 onwards).

Overall, our baseline estimates have the smallest differences between TSTSLS and OLS. The OLS RRC is roughly 4% larger than the TSTSLS RRC in baseline, while it 12% when using only the SRC sample, 11% when using the family longitudinal weights, 9% when using individual longitudinal weights, and 12% when using the individual cross-sectional weights. Regarding the regional estimates, as with the baseline results, the relative difference between TSTSLS and OLS estimates largely reflect sample size: estimates for the Midwest and the South are quite close across specifications (a bit less so for the Midwest when using the SRC sample), while the differences become more pronounced for the Northeast and the West, for which the sample size is more limited. It should be noted that across regions and specifications, the TSTSLS estimates of the AUM are surprisingly close to their OLS counterparts.

Table B.6: Comparison between Baseline Results and Weighted Results

			Weights in Last Child Income Observation Year		
	Baseline Estimates	Only SRC Sample	Family Longitudinal	Individual Longitudinal	Individual Cross-Sectional
	(1)	(2)	(3)	(4)	(5)
Panel A. National - Intergenerational Elasticity (IGE)					
Observed parent income (OLS)	0.334	0.369	0.372	0.377	0.369
	(0.011)	(0.018)	(0.012)	(0.012)	(0.013)
Predicted parent income (TSTSLS)	0.445	0.418	0.432	0.428	0.411
	(0.014)	(0.022)	(0.014)	(0.014)	(0.015)
Percentage diff. TSTSLS vs OLS	-24.82%	-11.72%	-13.89%	-11.91%	-10.23%
Number of observations	4,730	2,892	4,730	4,730	4,588
Panel B. National - Rank-Rank Correlation (RRC)					
Observed parent income (OLS)	0.476	0.409	0.458	0.456	0.437
	(0.013)	(0.017)	(0.013)	(0.013)	(0.014)
Predicted parent income (TSTSLS)	0.459	0.364	0.413	0.418	0.39
	(0.013)	(0.017)	(0.013)	(0.013)	(0.014)
Percentage diff. TSTSLS vs OLS	3.85%	12.38%	10.91%	9.09%	12.07%
Number of observations	4,755	2,903	4,755	4,755	4,612
Panel C. Region: Midwest					
RRC - OLS	0.528	0.417	0.45	0.456	0.447
	(0.025)	(0.03)	(0.026)	(0.026)	(0.027)
AUM - OLS	39.87	43.94	48.2	48.81	51.49
RRC - TSTSLS	0.535	0.37	0.427	0.433	0.444
	(0.025)	(0.032)	(0.027)	(0.027)	(0.028)
AUM - TSTSLS	41.16	46.85	50.13	50.4	53.01
RRC percentage diff. TSTSLS vs OLS	-1.22%	12.49%	5.48%	5.25%	0.66%
AUM percentage diff. TSTSLS vs OLS	-3.15%	-6.22%	-3.85%	-3.15%	-2.87%
Number of observations	1,283	980	1,283	1,283	1,265
Panel D. Region: Northeast					
RRC - OLS	0.508	0.429	0.497	0.494	0.487
	(0.035)	(0.04)	(0.036)	(0.036)	(0.037)
AUM - OLS	40.46	41.91	37.99	42.71	39.76
RRC - TSTSLS	0.457	0.35	0.443	0.441	0.42
	(0.034)	(0.039)	(0.035)	(0.035)	(0.037)
AUM - TSTSLS	38.83	41.49	37.09	41.41	40.1
RRC percentage diff. TSTSLS vs OLS	11.25%	22.4%	12.24%	11.99%	15.77%
AUM percentage diff. TSTSLS vs OLS	4.19%	1%	2.42%	3.14%	-0.84%
Number of observations	674	538	674	674	650
Panel E. Region: South					
RRC - OLS	0.417	0.398	0.447	0.428	0.421
	(0.02)	(0.03)	(0.019)	(0.019)	(0.02)
AUM - OLS	36.57	36.19	36.69	43.12	40.85
RRC - TSTSLS	0.41	0.401	0.437	0.423	0.406
	(0.02)	(0.03)	(0.019)	(0.019)	(0.02)
AUM - TSTSLS	37.28	35.01	37.31	43.2	41.35
RRC percentage diff. TSTSLS vs OLS	1.72%	-0.75%	2.29%	1.25%	3.79%
AUM percentage diff. TSTSLS vs OLS	-1.92%	3.37%	-1.64%	-0.2%	-1.19%
Number of observations	2,046	885	2,046	2,046	1,970
Panel F. Region: West					
RRC - OLS	0.371	0.299	0.363	0.385	0.336
	(0.038)	(0.047)	(0.038)	(0.039)	(0.04)
AUM - OLS	42.13	40.03	40.98	43.83	47.9
RRC - TSTSLS	0.328	0.232	0.264	0.301	0.231
	(0.038)	(0.046)	(0.036)	(0.038)	(0.038)
AUM - TSTSLS	43.28	43.43	44.76	46.89	52.58
RRC percentage diff. TSTSLS vs OLS	12.97%	28.89%	37.34%	27.62%	45.79%
AUM percentage diff. TSTSLS vs OLS	-2.66%	-7.84%	-8.43%	-6.53%	-8.9%
Number of observations	676	488	676	676	651

C Additional Robustness

This Appendix provides additional robustness checks to those presented in the body of the paper.

C.1 Sensitivity to Data Coverage

C.1.1 Civil Servants

We ensure our results are not affected by the fact that civil servants are only observed from 1988 onwards by estimating the first-stage regression computing synthetic parents' on post-1988 wages only, still restricting to when they are between 35 ad 45 years old. Appendix Figure C.1 displays the results from this check. The results are largely unaffected.

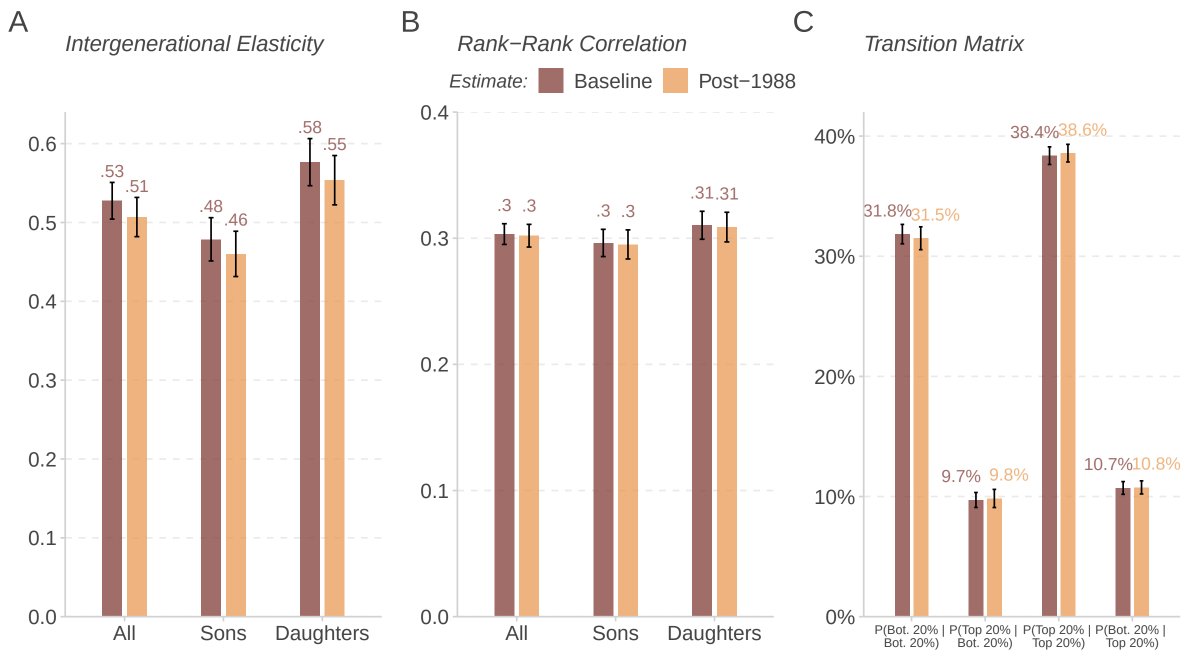


Figure C.1: Robustness of Baseline Estimates to Computing Synthetic Parent Incomes only on Post-1988 Data

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to computing synthetic parents' incomes only on post-1988 data. The All Employee Panel from which synthetic parents' wages are observed did not cover civil servants prior to 1988 (see Appendix Section A for details). The graph presents the baseline estimates (Baseline) to those obtained when synthetic parent incomes are defined as average wage between 35-45 using only post-1988 wages (Post-1988). Vertical lines represent the 95% bootstrapped confidence intervals. All results pertain to parent and child incomes being defined at the household level. The results for the transition matrix correspond to the sample pooling sons and daughters. See Section 3 for details on data, sample and income definitions.

C.1.2 Comparison with Population Statistics

Since the sample selection of the EDP is (virtually) random (individuals born on the first four days of October), we can have a good idea of how our baseline sample compares with the French population by comparing its average characteristics to those of the completely unrestricted EDP sample for the same birth cohorts (1972-1981).

To obtain characteristics on parents (other than from the 1990 census), we rely on individuals' birth-certificates information from the EDP civil registry data. We compare the birth-certificate information (e.g., gender, parents' age at birth, single parenthood, parents' occupation at birth) for all EDP individuals born in 1972-1981 in metropolitan France and for our sample of children. Note that the resulting statistics are subject to the imperfections of birth-certificate data, notably regarding non-random missing information for fathers. Table C.1 displays the statistics for both samples. Overall, our sample of children is very similar to the unrestricted EDP sample, except for a higher probability of being in the fiscal data (91% vs. 100%, by construction) and a lower likelihood of having a father who is a farmer. The household income distributions are very similar.

Characteristic	Population	Sample	Diff.
Females	49.14%	49.77%	0.63
Parent demographics			
Mother age at birth	26	25.89	-0.11
Father age at birth	28.91	28.65	-0.26
Mother born French	90.07%	91.92%	1.85
Father born French	88.12%	90.15%	2.03
Single mothers	4.98%	4.42%	-0.56
Missing parents info.	2.24%	1.75%	-0.49
Father 1-digit occupation at child birth			
Missing father info.	9.4%	8.25%	-1.15
1. Farmers	3.41%	0.64%	-2.77
2. Craftsmen, salespeople, and heads of businesses	3.95%	3.96%	0.01
3. Managerial and professional occupations	7.14%	5.98%	-1.16
4. Intermediate professions	13.58%	14.63%	1.05
5. Employees	14.58%	16%	1.42
6. Blue collar workers	46.46%	49.4%	2.94
7. Retirees	0.03%	0.02%	-0.01
8. Other with no professional activity	1.45%	1.11%	-0.34
Mother 1-digit occupation at child birth			
Missing mother info.	5.34%	4.69%	-0.65
1. Farmers	0.83%	0.11%	-0.72
2. Craftsmen, salespeople, and heads of businesses	0.91%	0.85%	-0.06
3. Managerial and professional occupations	2.08%	1.62%	-0.46
4. Intermediate professions	8.92%	9.19%	0.27
5. Employees	26.19%	28.33%	2.14
6. Blue collar workers	11.2%	12%	0.8
7. Retirees	0.02%	0.02%	0
8. Other with no professional activity	44.51%	43.2%	-1.31
All Employee Panel (AEP) information in adulthood, 1968-2015, age 35-45			
Observed in AEP	72.83%	78.67%	5.84
Mean number of obs. in AEP	2.9	3.15	0.25
Q1 individual wage (AEP)	12,671	13,179	508
Mean individual wage (AEP)	21,538	21,666	128
Med. individual wage (AEP)	19,528	19,726	198
Q3 individual wage (AEP)	26,623	26,723	100
Tax information in adulthood, 2010-2016, age 35-45			
Observed in tax data	90.92%	100%	9.08
Mean number of obs. in tax data	4.23	4.65	0.42
Q1 household income (tax)	27,339	27,696	357
Mean household income (tax)	46,858	46,598	-260
Med. household income (tax)	41,220	41,418	198
Q3 household income (tax)	56,630	56,481	-149
N	83,009	64,571	

Notes: Comparison of birth-certificate information on the full EDP sample vs. the study sample. See Section 3.2 for details on construction of the study sample.

Table C.1: Average Characteristics of Overall Population vs. Sample

C.2 Alternative First-Stage Estimation

The parent income predictions we use to palliate French data limitations are central to our analysis. It is of primary importance that the first stage of the two-step strategy we rely on is reliable. We make sure that this first stage does not spuriously drive the results in one way or another by evaluating its sensitivity to varying the set of instruments and to relaxing parametric assumptions.

C.2.1 Set of First-Stage Predictors

The most important dimension to consider is the set of variables included in the first stage, notably because it has been shown that inadequate instruments could yield inconsistent estimates (Jerrim et al., 2016). Appendix Figure C.2 documents the sensitivity of IGE, RRC and transition matrix estimates to the set of predictors used in the first-stage estimation. We estimate them when adding each of the following predictors sequentially (all measured in 1990): education (8 categories), 2-digit occupation (42 cat.), a group of demographic characteristics (age, French nationality dummy, country of birth (6 cat.), and household structure (6 cat.)) and a group of municipality-level characteristics (unemployment rate, share of single mothers, share of foreigners, population, and population density). Since relying on a single variable with less than 100 categories induces some income values to span over several percentiles, parents with a given predicted income are attributed the average rank of individuals earning that level of income. Lastly, we also report the adjusted R^2, computed as the average from 5-fold cross-validation.

We find that the IGE is 0.68 when using only education as the first-stage predictor, consistent with a point already made in the literature that using only education as a predictor is likely to yield inflated estimates of the IGE. Once 2-digit occupation is included in the first-stage, adding other demographic or municipality-level characteristics has no effect on the estimates. Indeed, as can be seen from the R^2, most of the predictive power actually comes from the 2-digit occupation variable. The RRC appears remarkably unchanged by the set of first-stage predictors used, at 0.28 with only education and 0.30 with all variables. This appears once more to be a strength of the RRC in the TSTSLS context.

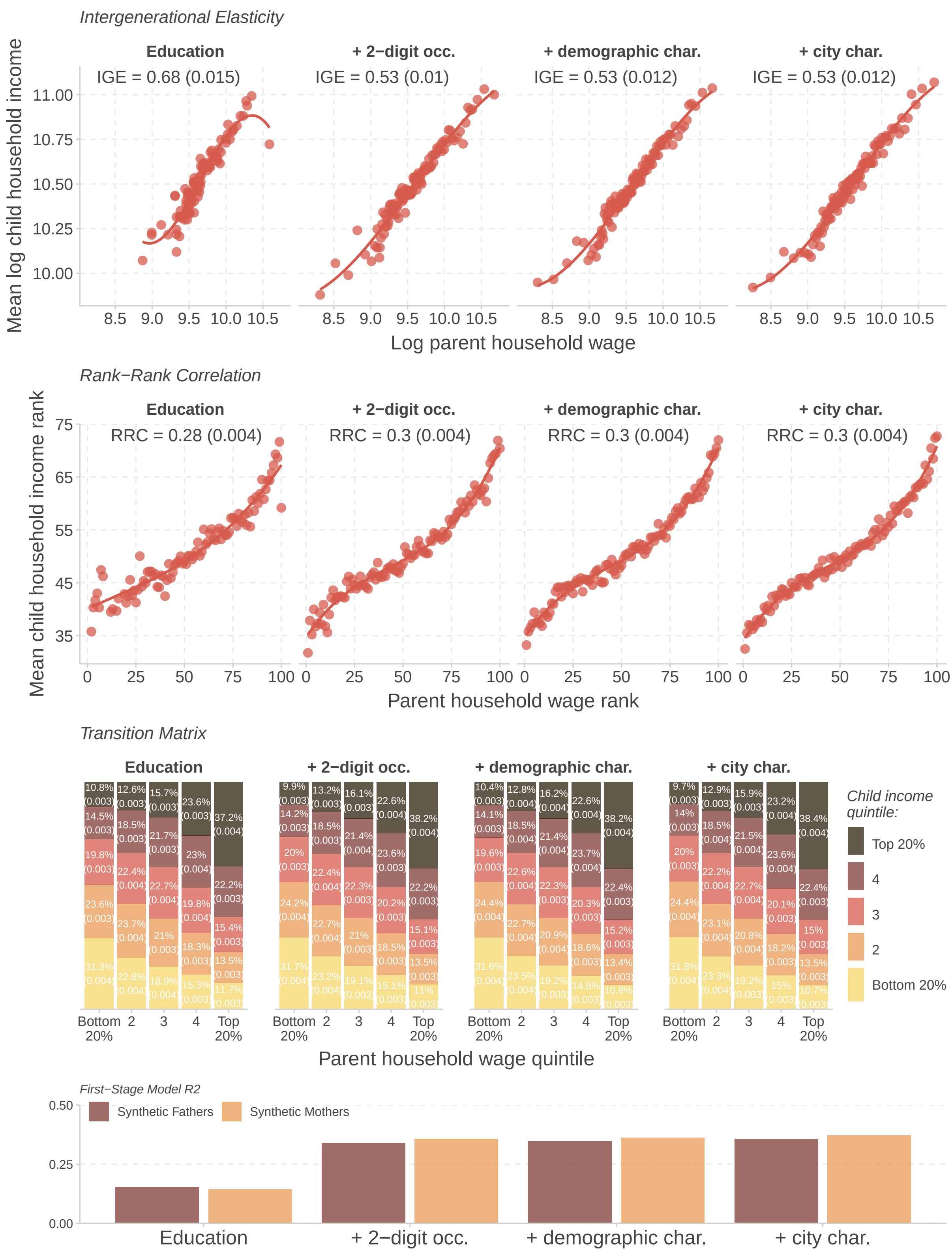


Figure C.2: Robustness of Baseline Estimates to Different First-Stage Predictors

Notes: This figure assesses the robustness of our baseline IGE, RRC and transition matrix estimates to variations in the set of first-stage predictors. Parent income is predicted separately for fathers and mothers using a set a of instruments that vary along the x-axis. We report the corresponding CEFs, along with the point estimates and the bootstrapped standard error in parenthesis. The bottom panel of the figure reports separately for synthetic fathers and mothers the R^2 associated with each first stage. See Figure 3's notes for details on data, sample and income definitions.

C.2.2 Flexible Models

We make use of semi- and non-parametric models to elicit potential misspecifications in the first stage. The baseline specification of the first stage is of the form $y = \beta X + \varepsilon$, where y is the log of parent lifetime income and X is a set of k predictors. OLS would not account for interactions between predictors nor for non-linearities in the relationship between X and y unless they are explicitly modeled. Fully non-parametric methods of the form $y = m(X) + \varepsilon$ would capture both interactions and non-linearities that may help reduce the out-of-sample MSE. Obtaining a lower MSE and significantly different second-stage estimates with non-parametric models than with OLS would suggest that non-modeled non-linearities, interactions, or both, influence the resulting intergenerational mobility estimates.

We implement this test using three machine learning methods: (i) a generalized additive model (GAM) of the form $y = m_1(x_1) + m_2(x_2) + ... + m_k(x_k) + \varepsilon$ which accounts for non-linearities but not for interactions unless explicitly specified, (ii) a gradient boosted regression tree, that is a high-dimensional combination of sequentially grown regression trees, and (iii) the ensemble method, which consists in taking the average of the predictions from each model weighted in a way that minimizes the out-of-sample MSE.

Appendix Figure C.3 compares the intergenerational mobility estimates and out-of-sample MSE resulting from these three methods using our baseline child and parent income definitions. We do not observe significant differences in MSE between the different prediction methods. The resulting mobility estimates are virtually the same for OLS, GAM and the ensemble method, and slightly smaller for boosted trees. This suggests that conditional on the set of predictors we use, using more flexible estimation methods does not lead to better income predictions and different estimates than using an additive OLS specification.

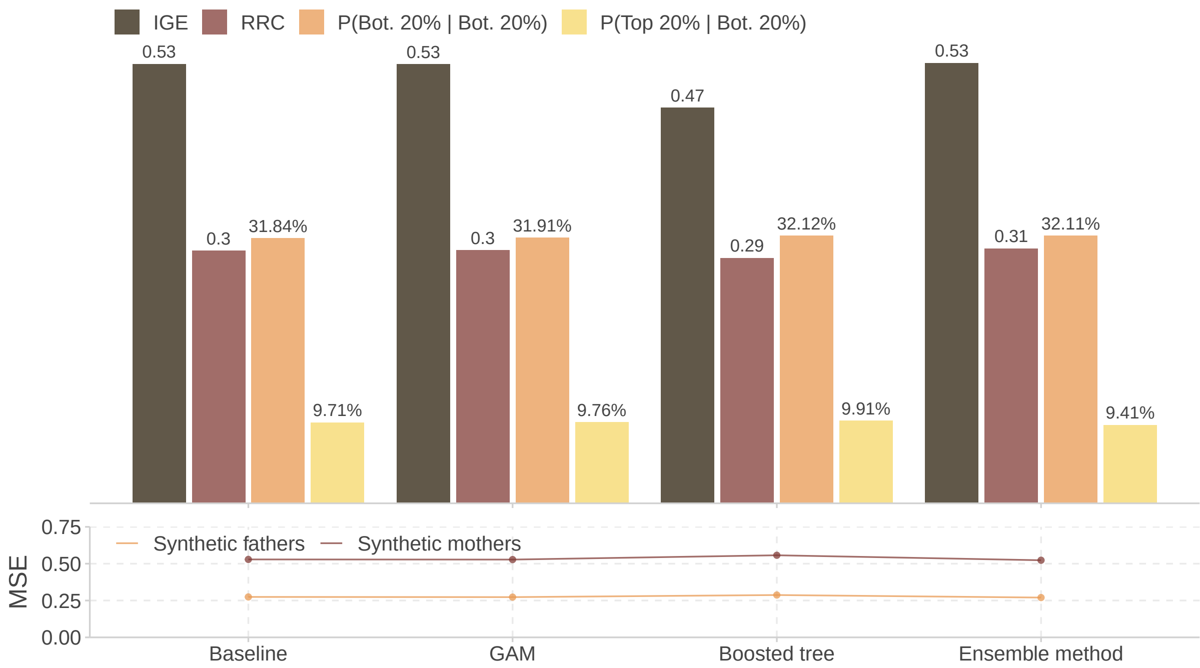


Figure C.3: Robustness to Machine Learning Prediction

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to increasingly flexible first-stage prediction models. Each bar represents the magnitude of the estimate of the corresponding color estimated using the first-stage model indicated on the x-axis. The first set of estimates are the baseline estimates obtained using OLS. The three other sets are obtained using increasingly flexible models: generalized additive models (GAM), gradient boosted regression trees, and the ensemble method. The connected dots represent the average out-of-sample MSEs of the associated prediction models, estimated using 5-fold cross-validation. See Figure 3's notes for details on data, sample and income definitions.

C.3 Lifecycle and Attenuation Bias

C.3.1 Child Lifecycle Bias - Constant Sample of Children

To overcome the issue related to changes in Figure 6's underlying sample of children, we reproduce the individual wage estimates using the All Employee Panel keeping the sample of children constant. To do so we restrict to children born in 1972 and 1974[16] for whom wages are observed every year between 25 and 43 years old and 25 and 41 years old respectively. Appendix Figure C.4 displays the results. Since the sample is kept constant throughout, the coefficients can be compared to one another and the change in magnitude can only be driven by the age at which child income is measured rather than sample composition. As in Figure 6, we find that measuring child income prior to the mid-thirties seriously underestimates the IGE (panel A) and RRC (panel B), and overestimates (underestimates) bottom or top mobility (persistence) (panel C).

[16]We cannot include the 1973 cohort as the All Employee Panel income data are only available for individuals born an even year before 2001. This choice of cohorts is done to be able to measure their incomes after they are 40 years old.

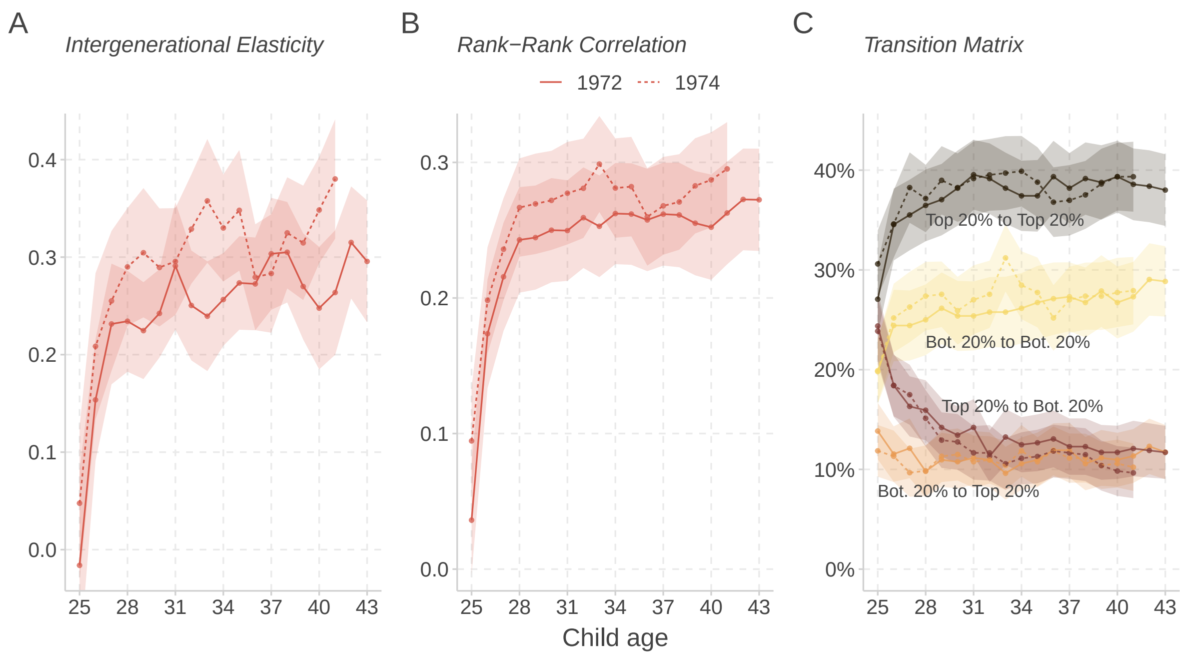
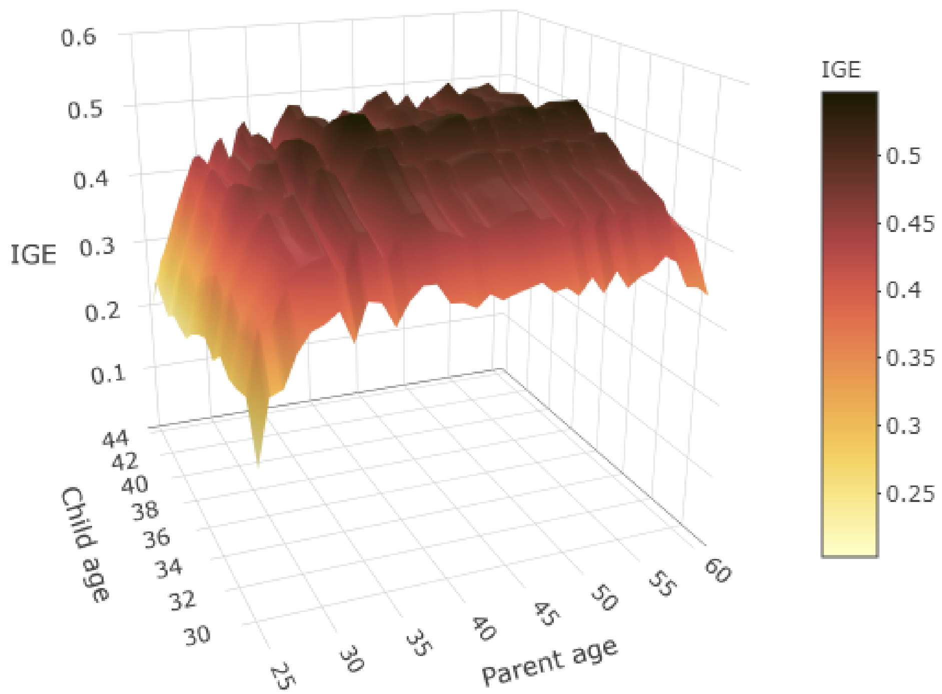


Figure C.4: Child Lifecycle Bias - 1972 and 1974 Cohorts (Constant Sample)

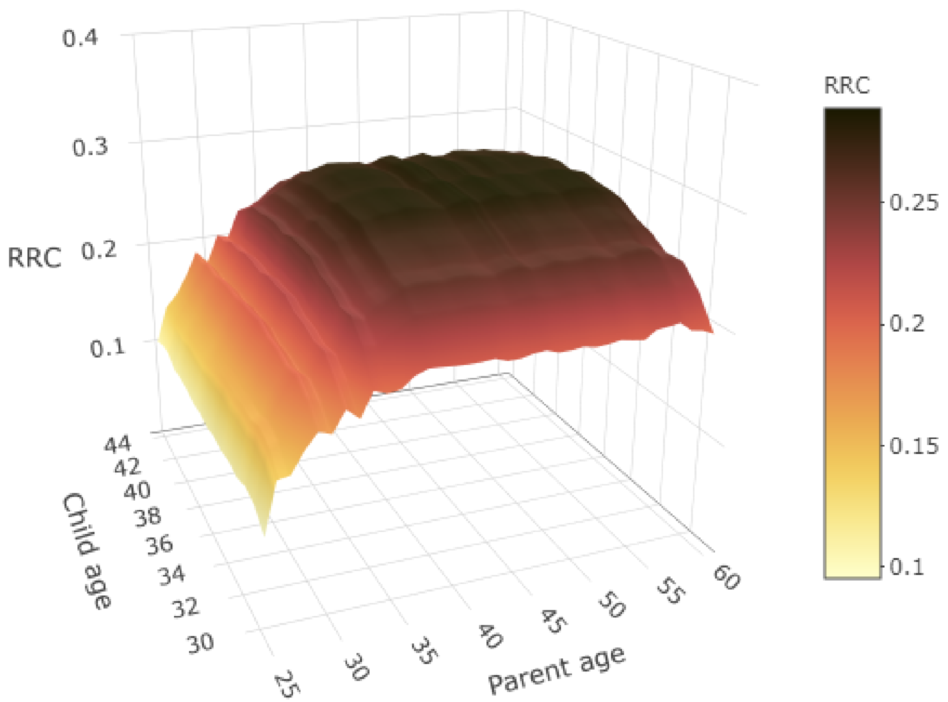
Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figures 3 and 4 to changes in the age at which child income is measured, for children born in 1972 (solid line) and 1974 (dashed line). For both birth cohorts the sample is kept constant, that is only children with wages observed in the All Employee Panel at each age between 25 and 43 years old are retained. Shaded areas represent the 95% bootstrapped confidence interval. See Sections 3 and 4.4 for for details on data, sample and income definitions.

C.3.2 Child and Parent Lifecycle Bias Jointly

Child and parent lifecycle bias are typically assessed independently, as we do in the main body of the article. Yet they influence one another and it is instructive to estimate our measures of intergenerational persistence for each possible combination of synthetic parent and child age. Appendix Figure C.5 shows such estimates when child income is measured between ages 30 and 44, and synthetic parent income between ages 28 and 60.



(a) Intergenerational Elasticity



(b) Rank-Rank Correlation

Figure C.5: Child and Parent Lifecycle Bias

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figure 3 to changes in the age at which child and synthetic parent incomes are measured. The sample of children and synthetic parents varies across ages. See Sections Figure 3's notes for for details on data, sample and income definitions.

C.3.3 Parent Attenuation Bias

Figure C.6 plots estimates of our persistence measures varying the number of synthetic parent income observations used in the first-stage regression from 1 to 11. To control for the potential effect of lifecycle bias we center the age at which synthetic parent income is measured at 40 years old. In other words, one income observation corresponds to income at age 40, two income observations corresponds to average income at ages 39 and 41, three income observations to average income between 39 and 41, and so on. Therefore, 11 income observations corresponds to the average between 35 and 45 years old. The sample of synthetic parents over which income is predicted varies for each estimate depending on how many synthetic parents had incomes observed each year in the required age range. We report results both for parent household wage and father wage.

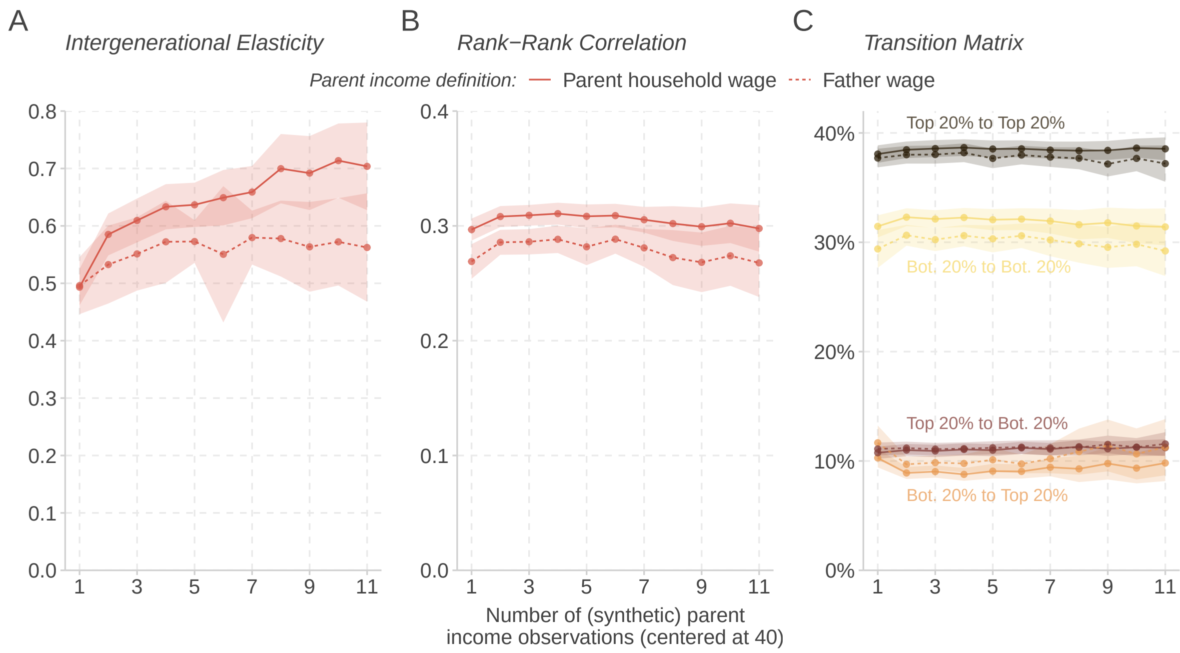


Figure C.6: Attenuation Bias

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to the number of income observations used to predict parent income. While varying the number of parent income observations, we center the age range at 40 to control for lifecycle bias. Shaded areas represent the 95% bootstrapped confidence interval. See Figure 3's notes for details on data, sample and income definitions.

These results suggest that attenuation bias might affect our baseline IGE (panel A) but not our other estimates of intergenerational mobility. Indeed when defining parent income at the household level, the IGE increases from 0.5 when using only one income observation to 0.7 when averaging over 11 income observations (i.e., between 35 and 45). It is important to highlight that almost all of this change is driven by how mothers' incomes are predicted. Indeed when looking at the father-child IGE, the estimate does not increase so markedly and stabilizes around 2 or 3 income observations, consistent with the idea that the two-stage procedure employed drastically shrinks the transitory component of annual income, and in large contrast with what is typically found when parent income is actually observed (Mazumder, 2005). Indeed, since we are already predicting parent income based on observable characteristics, and thus in a sense reducing year-to-year income volatility, averaging over more years does not affect the estimate much.

How one interprets the results based on parent household wage depends on one's prior as to how to best predict mothers' incomes. Our view is that predicting mothers' incomes only on the subsample of synthetic mothers with observed wages in all years between 35 and 45 years old biases the underlying sample considering the uneven labor force participation of women at the time. We believe our choice of restricting our sample of synthetic parents to those with at least two income observations between ages 35 and 45 is reasonable.

Constant Sample. We check whether the lack of change in intergenerational mobility measures with the number of synthetic parent income observations observed in Figure C.6 could be due to the fact that the sample of synthetic parents varies throughout. We replicate those estimates restricting the sample of synthetic parents to those with all 11 income observations between 35 and 45 years old and estimating the intergenerational mobility measures by varying the number of income observations averaged in the first-stage regression (centered around 40 years old again). To do so, we impute wages in 1981, 1983 and 1990, for which the data are not available,[17] using the average wage between the previous and subsequent year only if both wages are observed. This enables us to have a consistent sample and increase the number of synthetic parents on which the predictions can be done.

Appendix Figure C.7 displays the results from this sensitivity analysis. The increase in the parent household wage IGE is much less marked, increasing from 0.636 when using one income observation to 0.704 when using all 11 observations (panel A). Our interpretation of this relatively modest increase is that averaging over at least 2 income observations as we do for our baseline estimate should suffice to not suffer from attenuation bias. Note that what matters in this figures is not how different the estimates are from our baseline estimate but rather the extent to which they vary with the number of synthetic parent income observations used. Indeed, the difference between our baseline IGE estimate and the estimates obtained are driven by the fact that the sample of synthetic parents for whom we observe all incomes between 35 and 45 years old is a highly non-representative sample, especially when it comes to mothers. In fact, we do not find any attenuation bias when restricting our analysis to fathers, suggesting all the variation in the IGE can be accounted for by changes in mothers' incomes predictions. As with the varying synthetic parent sample estimates, rank-based intergenerational mobility measures are significantly less sensitive to averaging over more income years, and the estimates found are very close to our baseline ones (panels B and C).

[17] As explained in Appendix A, the 1982 and 1990 population censuses generated an extra workload which prevented INSEE from compiling the All Employee Panel data for these years.

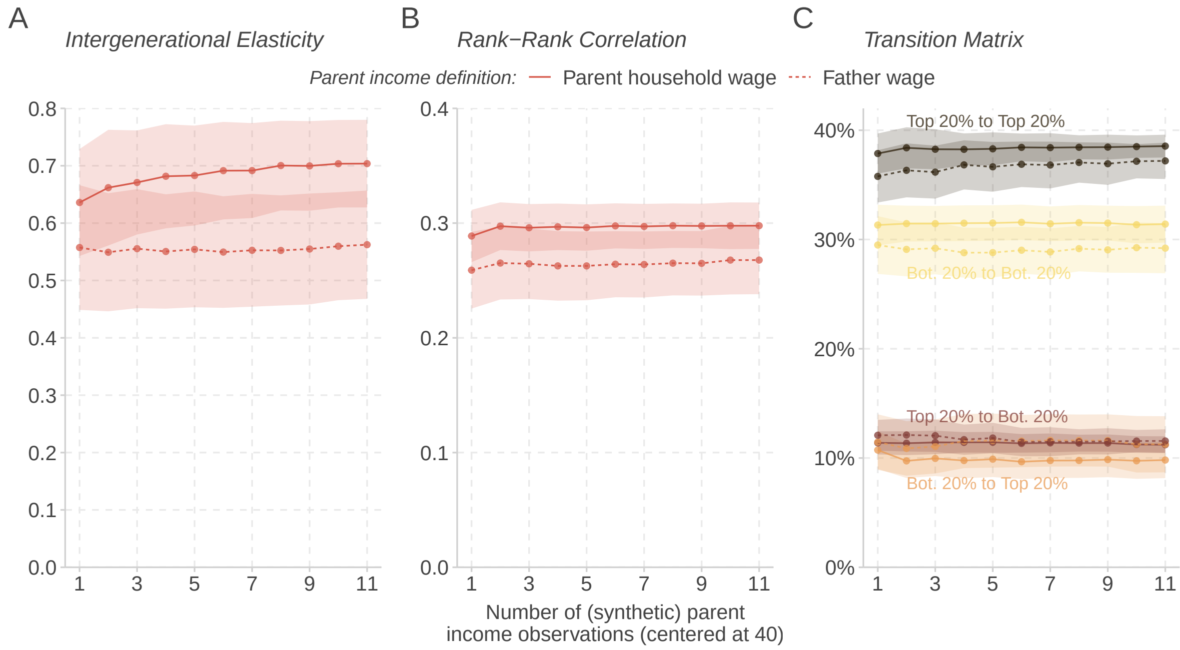


Figure C.7: Parent Attenuation Bias (Constant Sample of Synthetic Parents)

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to the number of income observations used to predict parent income, keeping the sample of synthetic parents constant. The sample of synthetic parents is thus restricted to those with all 11 income observations between 35 and 45 years old. While varying the number of parent income observations, we center the age range at 40 to control for lifecycle bias. Shaded areas represent the 95% bootstrapped confidence interval. See Figure 3's notes for details on data, sample and income definitions.

C.3.4 Child Attenuation "Bias"

Appendix Figure C.8 plots estimates of our persistence measures varying the number of child income observations from 1 to 7 between 35 and 45 years old, keeping the sample of children constant[18] (i.e. keeping only children with 7 household income observations). Due to this restriction only cohorts born between 1972 and 1975 are kept. Without this restriction, the value reported for 1 income observation would correspond to our baseline estimate. In the same way as for parents, we control for lifecycle bias by centering the year in which child income is measured to 2013. In other words, one child income observation corresponds to income measured in 2013, two income observations corresponds to the average between 2012 incomes and 2014 incomes, three to average income between 2012 and 2014, etc. The results suggest that estimates are largely unaffected by increasing the number of child income observations.

[18]The sample varies ever so slightly for the IGE due to the number of negative or 0 incomes changing between years.

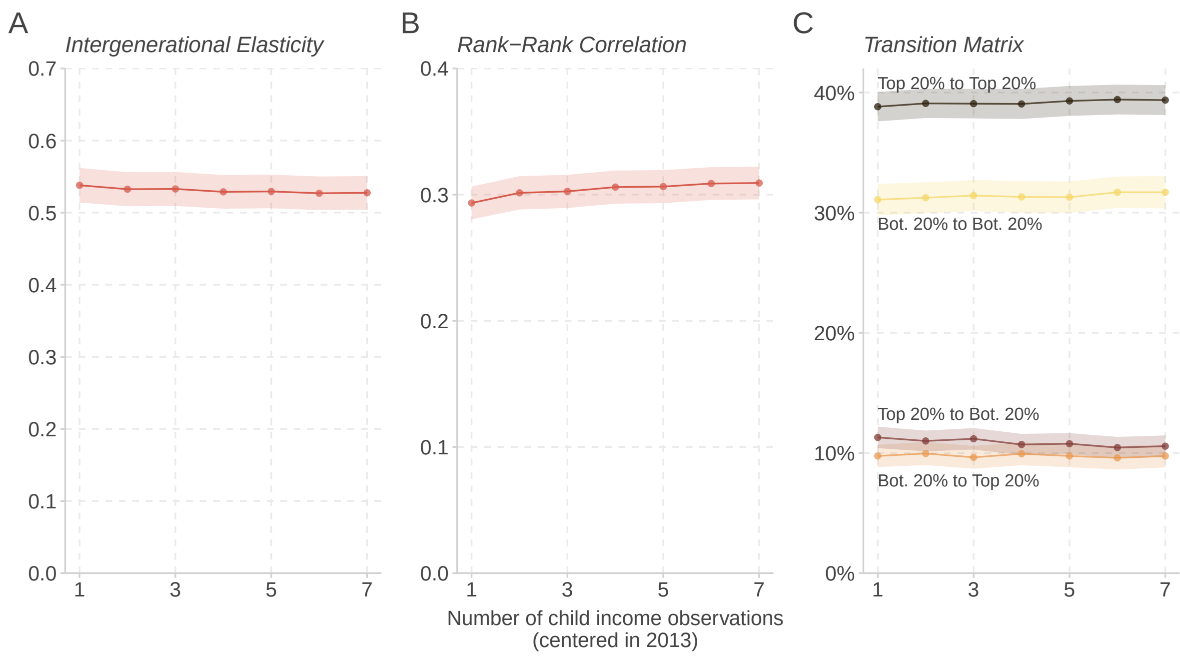


Figure C.8: Sensitivity to Number of Child Income Observations (Constant Sample)

Notes: This figure presents estimates of our persistence measures varying the number of child income observations from 1 to 7 between 35 and 45 years old, keeping the sample of children constant, i.e. keeping only children with 7 household income observations. (The sample varies ever so slightly for the IGE due to the number of negative or 0 incomes varying between years.) Due to this restriction only cohorts born between 1972 and 1975 are kept. Without this restriction, the value reported for 1 income observation would equal our baseline estimate. We control for lifecycle bias by centering the year in which child income is measured to 2013. In other words, one child income observation corresponds to income measured in 2013, two income observations corresponds to the average of 2012 and 2014, three to average income between 2012 and 2014, etc. Shaded areas represent the 95% bootstrapped confidence interval. See Figure 3's notes for for details on data, sample and income definitions.

C.4 Sensitivity to Income Distribution Tails.

Our baseline estimates may be sensitive to two main sample selection choices when it comes to the income distributions of parent and children: (i) how children with negative or zero incomes are treated; and (ii) how the top and bottom tails of both the parent and child income distributions are dealt with.

C.4.1 Treatment of Zeros

The first issue is particularly salient for the estimation of the intergenerational income elasticity due to the impossibility of taking the log of zero.[19] Many researchers simply discard such observations since they are likely not representative of lifetime income. Though this may potentially be the case if only short income time spans are available, we nonetheless evaluate how our baseline estimates of both the IGE and the RRC when replacing negative or zero child income values by 1 or 1,000 euros.

Appendix Figure C.9 shows estimates for the IGE and RRC when replacing income of children reporting negative or zero incomes by 1 euro or 1,000 euros, for different child income

[19]Various methods have been proposed to overcome this issue. Bellégo et al. (2021) describe such methods and propose a novel solution that can be applied to a variety of cases.

definitions. For our primary child income definition, household income, the estimates do not change due to there being very few children with negative or zero household income. However, for child income defined at the individual-level, for which the share of negative or zero incomes is more important, the IGE becomes highly sensitive to the recoding of such observations while the RRC remains unchanged. For example, for individual child income, the IGE is 0.46 when zeros are dropped and 0.82 when they are recoded to 1 and 0.55 when recoded to 1,000. The RRC is entirely insensitive to such recoding as ranks are not altered by it.

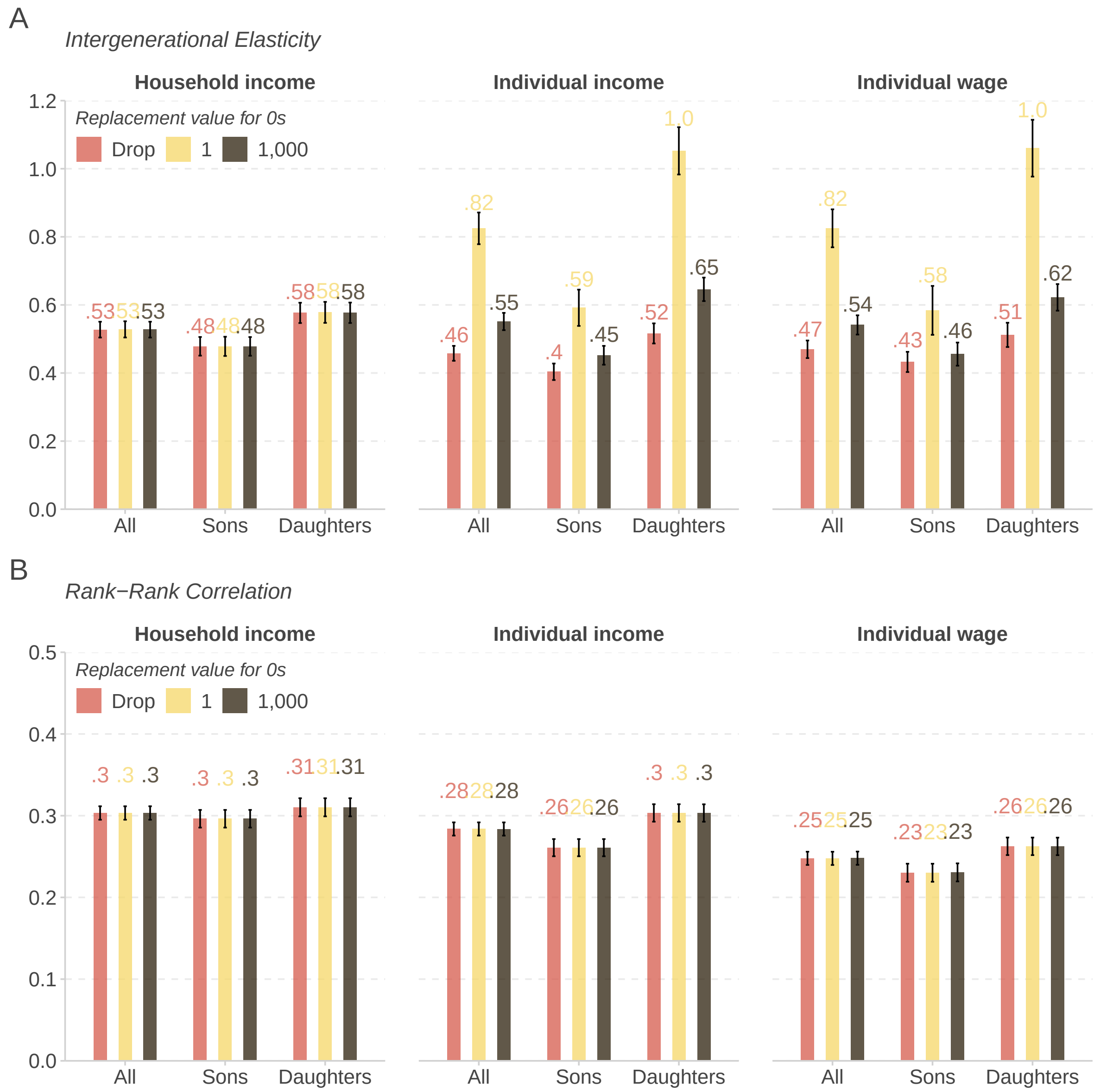


Figure C.9: Sensitivity to Different Zero Child Income Replacement Values

Notes: This figure assesses the robustness of our baseline IGE and RRC estimates to replacing incomes of children reporting negative or zero incomes by 1 euro or 1,000 euros, for different child income definitions. Vertical lines represent the 95% bootstrapped confidence intervals. See Section 3 for for details on data, sample and income definitions.

C.4.2 Top and Bottom Trimming

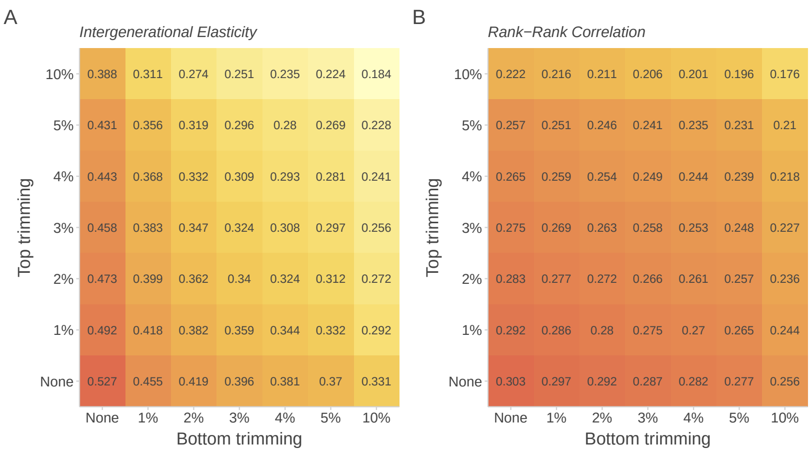
The second issue relates to the treatment of top and bottom earners in both the parent and child income distributions. For the parent income distribution the choice can both be made in the prediction stage and in the second stage. Specifically, we assess how the IGE and RRC vary when trimming the top and/or bottom 1% to 5% and 10%. Appendix Figure C.10 displays the results of this sensitivity check. There are three main takeaways.

First, the IGE is significantly more sensitive to small changes in parent or child income distributions while the RRC remains relatively stable. For example, removing the top and bottom 1% of child incomes decreases the IGE from 0.527 to 0.418 while the RRC only decreases from 0.303 to 0.286. It does not seem desirable that a measure of intergenerational mobility should be so sensitive to excluding just 2% of children. Mathematically it can be linked to changes in the dispersion of the distribution of child incomes but conceptually it seems difficult to defend such responsiveness to minor sample changes.

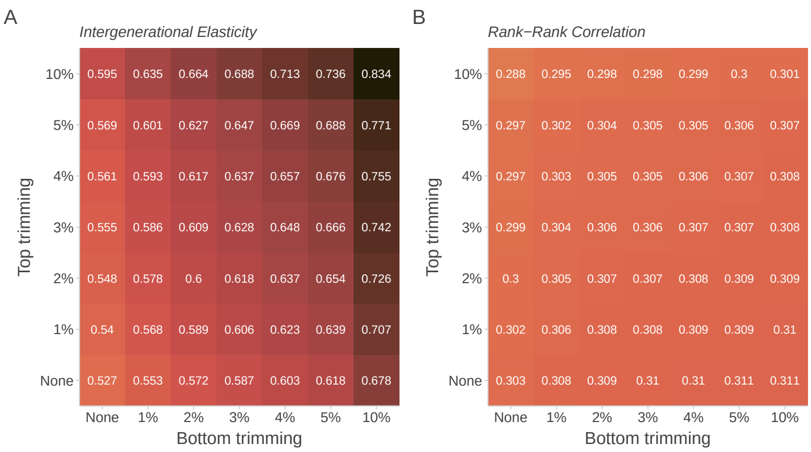
Second, the IGE is quite strongly influenced by minor trimming in the first-stage prediction sample. For example, excluding the bottom and top 2% of synthetic parent incomes leads to an IGE of 0.6. Such exclusions are not uncommon in the literature though their relevance is unclear.[20] Meanwhile the RRC is once more remarkably untouched by first-stage parent income exclusions. In fact excluding the bottom and top 10% of synthetic parent incomes decreases the RRC to 0.301 (from 0.303). This appears to be an additional benefit of estimating the RRC when using with the TSTSLS method. Note that trimming the first-stage prediction sample does lead to increased out-of-sample MSE, as shown in Appendix Figure C.11.

Third, for second-stage parent income trimming, the effects are relatively mild for both intergenerational mobility measures. This is very likely a consequence of the two-stage procedure which reduces the variance in parent incomes.

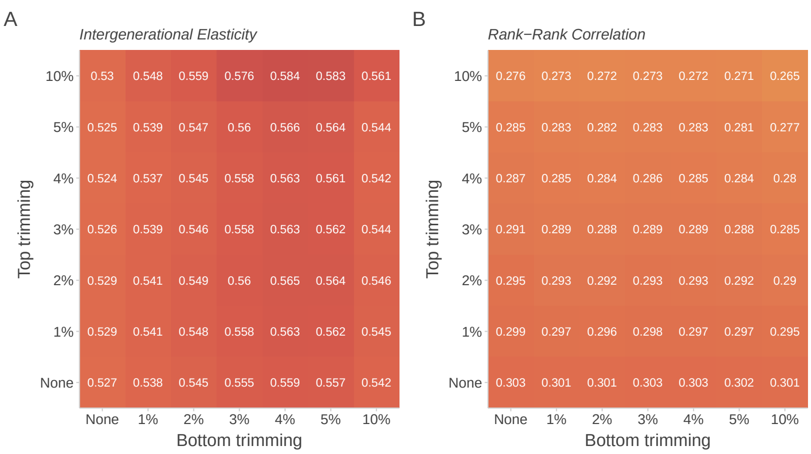
[20] For example, Barbieri et al. (2020) exclude the top and bottom 1% of their sons and synthetic fathers' incomes.



(a) Child Income Trimming



(b) First-Stage Synthetic Parent Income Trimming



(c) Second-Stage Parent Income Trimming

Figure C.10: Sensitivity to Child and Parent Income Distributions Trimming

Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented to trimming the tails of the parent and child income distributions. Each cell displays the value of the corresponding intergenerational mobility measure obtained after trimming the income distribution of the corresponding sample by the fraction indicated on the x-axis at the bottom and by that indicated on the y-axis at the top. See Figure 3's notes for details on data, sample and income definitions.

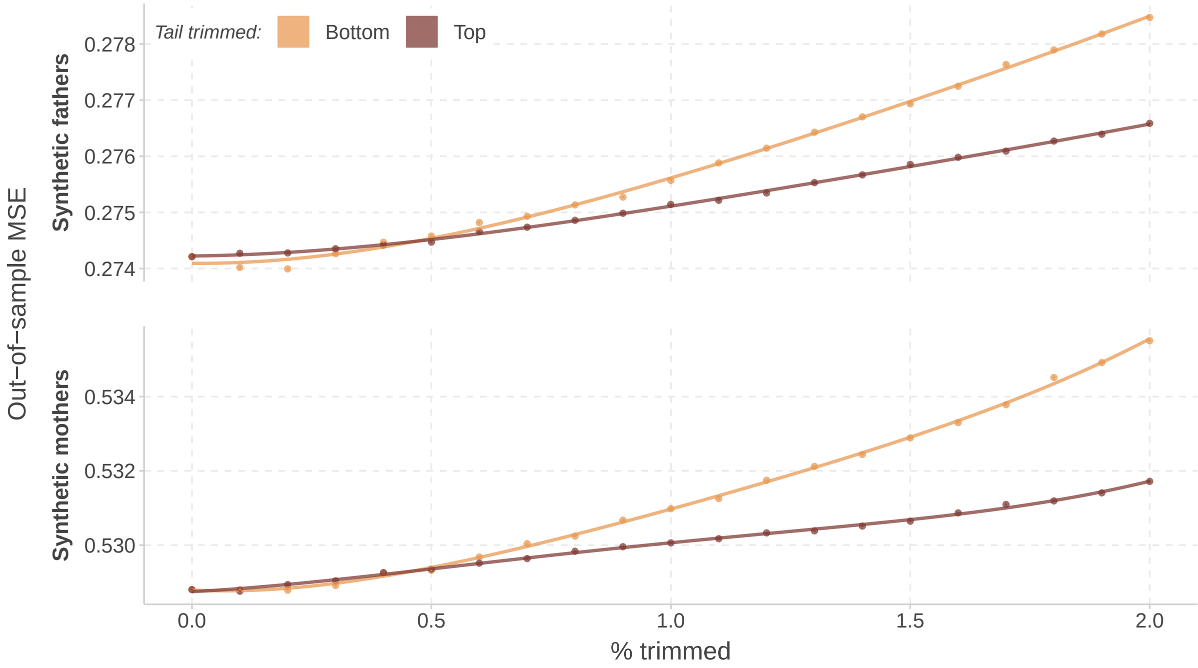


Figure C.11: Out-of-sample MSE as a Function of Top and Bottom Trimming

Notes: This figure plots the out-of-sample MSE as a function of trimming various shares of the tails of synthetic parents' income distribution. Our-of-sample MSEs correspond to the average MSE obtained from 5-fold cross-validation. See Sections 3.1 and 3.2 for details on the exact model being estimates and sample construction.

C.5 Transition Probabilities at the Top

To analyze persistence at the top of the parent income distribution, we estimate transition matrices for the top 10%, top 5% and top 2% of parent incomes and compare our results with those from the United States.[21] We estimate the likelihood of remaining in the top 10% to be about 28% in France close to the United States figure of 26%. This statistic is almost 3 times larger than would be observed in a world where child income is unrelated to parent income (i.e., 10%). This persistence at the top gets stronger as we zoom into the top 5% (22% remaining in top 5%) and top 2% (14% remaining in top 2%). The ratio of observed persistence to counterfactual world with no link between incomes increases with parent income rank in the distribution. This suggests that mechanisms of intergenerational persistence at the top of the parent income distribution might differ from those at play for the rest of the distribution.

[21] We use the detailed percentile-by-percentile estimates provided in the online appendix of Chetty et al. (2014).

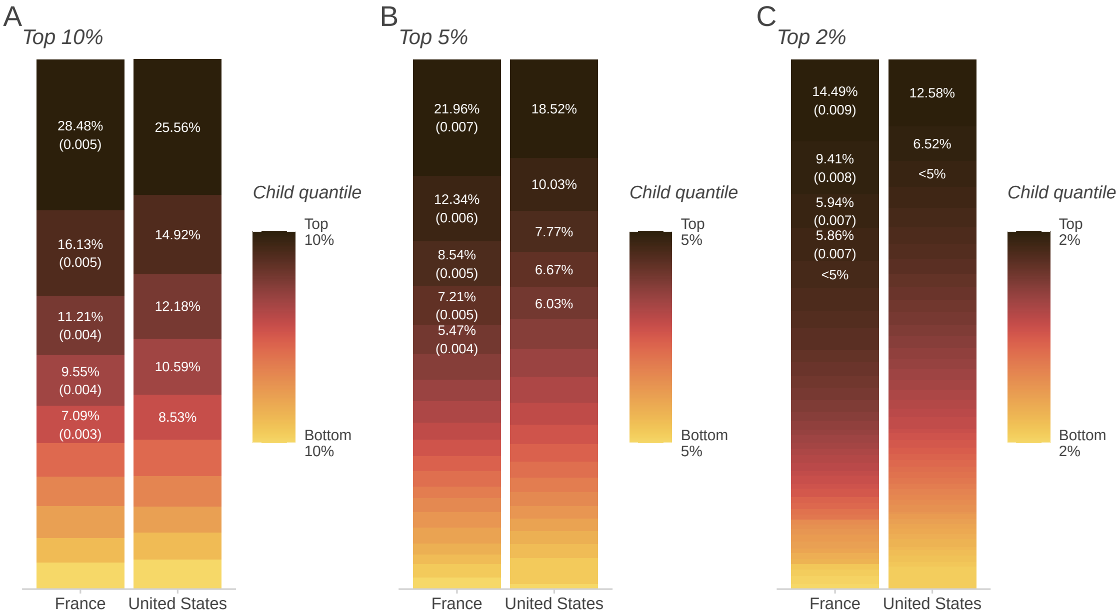


Figure C.12: Top Parent Income Quantiles Transition Matrices in France and United States

Notes: This figure presents intergenerational transition matrix estimates for children coming from families in the top 10% (panel A), top 5% (panel B) and top 2% (panel C) of the parent income distribution, with bootstrapped standard errors in parentheses. We compare the transition probabilities we obtain for France with those computed by Chetty et al. (2014) for the United States. Each cell corresponds to the percentage of children in a given income quantile who have parents in a given parent income quantile. See Section 3 for for details on data, sample and income definitions.

D Correlation with Local Characteristics

D.1 Definitions and Data Sources

Appendix Table D.1 displays the variables used in the correlational analysis presented in Section 6 (subsection *Correlation with Local Characteristics*). We measure these variables as close to 1990 as possible so as to reflect the environment individuals grew up in.

Variable	Definition	Source
Demographic		
Density	Log number of inhabitants per square meter	1990 BDCOM[1]
% Foreigner	Share without French nationality	1990 Census
% Single mothers	Share of single mothers in the adult population ($\geq$ 18)	1990 Census
Economic		
Mean wage	Log average wage	1996 DADS Panel
% Unemployed	Unemployment rate	1990 Census
Inequality		
Gini index	Gini index of wage inequality	1996 DADS Panel
Theil index	Theil index of spatial wage segregation	1996 DADS Panel
Share top 1%	Share of total wage accrued by the top 1% of wage earners	1996 DADS Panel
Education		
# HEI	Number of higher education institutions	2007 BPE[2]
Distance to HEI	Average distance to the closest public higher education institution	2007 BPE[2]
% HS graduates	Share of high-school graduates in adult population ($\geq$ 18)	1990 Census
Social capital		
Cultural amenities	Number of cinemas and museums per capita	2007 BPE[2], Min. de la Culture
Crime	Number of offenses and crimes per capita	Min. de l'Intérieur
% Voters	Participation rate to the first round of the 1995 presidential election	Min. de l'Intérieur

Notes:
[1] Base de données communales du recensement de la population (BDCOM) - 1990, INSEE (producteur), ADISP (diffuseur) - doi:10.13144/lil-0363
[2] Base permanente des équipements (BPE) - 2007, INSEE (producteur), PROGEDO-ADISP (diffuseur) - doi:10.13144/lil-0423

Table D.1: Definitions and Sources of Department Characteristics

D.2 Simple Regression Analysis

We start by regressing department-level intergenerational mobility estimates on each of these variables in separate regressions. Both the department intergenerational mobility estimates and the characteristics are standardized, implying that the coefficients can be interpreted as correlations. Results are presented in Appendix Tables D.2 to D.4 and summarized in Figure 9. Note that for the IGE and RRC, a positive coefficient implies the characteristic is positively correlated with intergenerational *persistence* (i.e., negatively correlated with intergenerational *mobility*), while for absolute upward mobility a positive coefficient implies the characteristic is positively correlated with higher incomes for children born to low-income families.

Appendix Figure D.1 provides a potential explanation for the results of the correlational analysis by documenting the correlation between all department characteristics. The 14 variables considered are for the most part quite strongly correlated with one another, both within and between variable groups. For instance, the Gini index is highly correlated with other inequality measures, but also with population density and the share of high school graduates, two variables whose relationship with absolute upward mobility is positive.

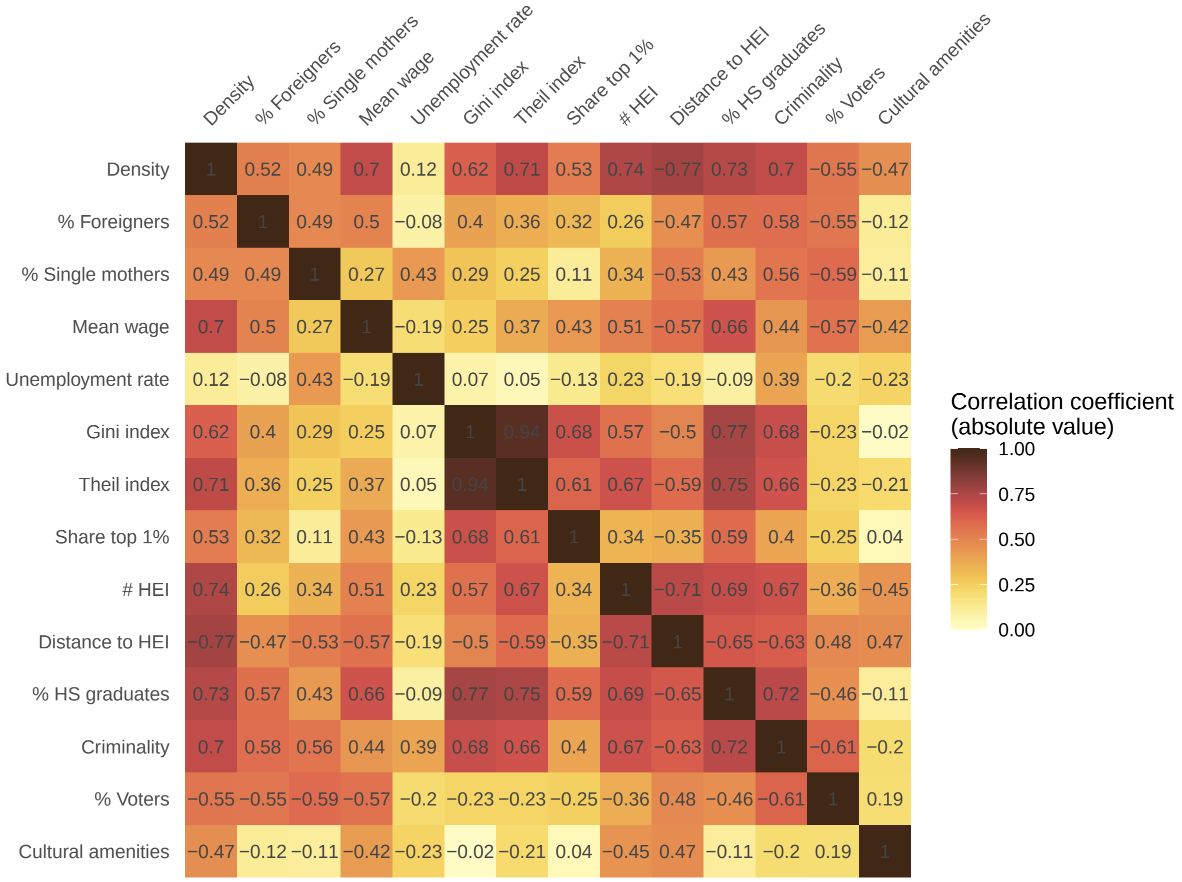


Figure D.1: Correlation Between Department Characteristics

Notes: This figure presents the correlation coefficient between all department characteristics considered, defined as follows. See Appendix Table D.1 for definitions and sources of the department characteristics.

D.3 Lasso Analysis

Considering the strong correlation across department characteristics, we estimate lasso regressions in order to identify the characteristics that are most strongly associated with intergenerational mobility. The result of this analysis is presented in Appendix Figure D.2.

The lasso analysis does not change the picture much. For the IGE, only the unemployment rate is picked up, as was the case in the univariate setting. For the RRC, the lasso maintains some demographic characteristics (% of single mothers and % foreigners), the unemployment rate, all three education variables, and two measures of social capital (cultural amenities and % voters). Again, these results are largely in line with what was observed in the univariate regressions. Lastly, for absolute upward mobility roughly the same characteristics that were significant in the simple regression analysis are kept except importantly for the Gini index.

Though the relationships we document between intergenerational mobility and department characteristics are overall pretty intuitive, these descriptive relationships cannot distinguish a potential causal effect of place from sorting. We leave this causal assessment to future studies.

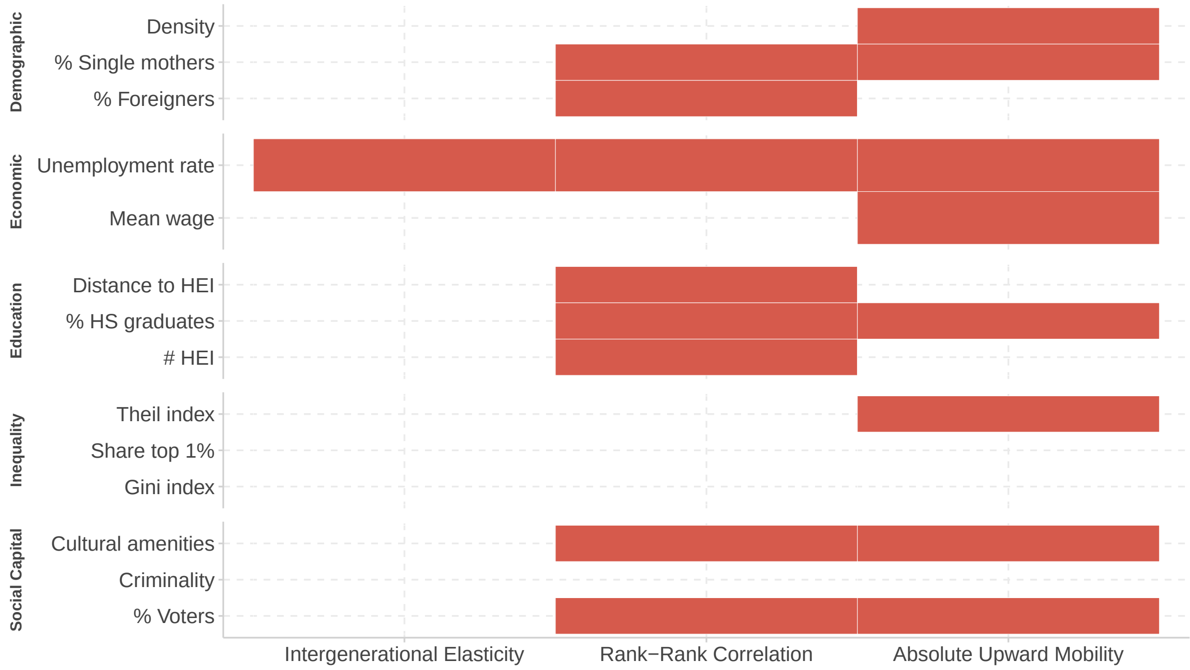


Figure D.2: Department Characteristics Kept by Lasso

Notes: This figure presents the department characteristics kept by the lasso regression. See Appendix Table D.1 for definitions and sources of the department characteristics.

	Dependent variable: Intergenerational Elasticity													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	−0.022 (0.110)													
% Single mothers		0.016 (0.110)												
% Foreigners			−0.026 (0.110)											
Unemployment rate				0.306*** (0.104)										
Mean wage					−0.157 (0.108)									
Distance to HEI						−0.100 (0.109)								
% HS graduates							−0.114 (0.109)							
# HEI								−0.022 (0.110)						
Theil index									−0.024 (0.110)					
Share top 1%										−0.092 (0.109)				
Gini index											0.007 (0.110)			
Cultural amenities												−0.090 (0.109)		
Crime													0.086 (0.109)	
% Voters														0.042 (0.110)
Intercept	4.124*** (0.135)	4.008*** (0.893)	4.183*** (0.213)	2.512*** (0.565)	18.162* (9.654)	4.374*** (0.277)	4.554*** (0.412)	4.218*** (0.406)	4.397*** (1.169)	4.954*** (0.977)	4.000* (2.287)	4.386*** (0.318)	3.911*** (0.312)	2.943 (3.152)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.0005	0.0003	0.001	0.094	0.025	0.010	0.013	0.0005	0.001	0.008	0.00005	0.008	0.007	0.002

Notes: All variables are standardized such that the regression coefficient corresponds to the correlation. See Appendix Table D.1 for variable definitions and data sources. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table D.2: Correlation Between Intergenerational Elasticity and Department Characteristics

	Dependent variable: Rank-Rank Correlation													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	−0.083													
	(0.109)													
% Single mothers		−0.168												
		(0.108)												
% Foreigners			−0.255**											
			(0.106)											
Unemployment rate				0.181*										
				(0.108)										
Mean wage					−0.131									
					(0.109)									
Distance to HEI						−0.078								
						(0.109)								
% HS graduates							−0.131							
							(0.109)							
# HEI								0.099						
								(0.109)						
Theil index									0.055					
									(0.110)					
Share top 1%										−0.105				
										(0.109)				
Gini index											0.008			
											(0.110)			
Cultural amenities												−0.141		
												(0.109)		
Crime													−0.049	
													(0.110)	
% Voters														0.239**
														(0.107)
Intercept	5.197***	6.617***	5.683***	4.294***	16.940*	5.440***	5.736***	4.906***	4.675***	6.186***	5.097**	5.642***	5.389***	−1.612
	(0.135)	(0.881)	(0.206)	(0.583)	(9.692)	(0.278)	(0.411)	(0.404)	(1.167)	(0.976)	(2.287)	(0.316)	(0.312)	(3.064)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.007	0.028	0.065	0.033	0.017	0.006	0.017	0.010	0.003	0.011	0.0001	0.020	0.002	0.057

Notes: All variables are standardized such that the regression coefficient corresponds to the correlation. See Appendix Table D.1 for variable definitions and data sources. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table D.3: Correlation Between Rank-Rank Correlation and Department Characteristics

	Dependent variable: Absolute Upward Mobility													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	0.491^{***}													
	(0.096)													
% Single mothers		0.218^{**}												
		(0.107)												
% Foreigners			0.478^{***}											
			(0.096)											
Unemployment rate				-0.563^{***}										
				(0.091)										
Mean wage					0.591^{***}									
					(0.089)									
Distance to HEI						-0.293^{***}								
						(0.105)								
% HS graduates							0.530^{***}							
							(0.093)							
# HEI								0.274^{**}						
								(0.106)						
Theil index									0.337^{***}					
									(0.103)					
Share top 1%										0.406^{***}				
										(0.100)				
Gini index											0.288^{***}			
											(0.105)			
Cultural amenities												0.015		
												(0.110)		
Crime													0.221^{**}	
													(0.107)	
% Voters														-0.405^{***}
														(0.100)
Intercept	13.426^{***}	11.305^{***}	12.268^{***}	16.055^{***}	-39.580^{***}	13.751^{***}	11.138^{***}	12.092^{***}	9.488^{***}	9.459^{***}	7.068^{***}	13.026^{***}	12.476^{***}	24.709^{***}
	(0.118)	(0.872)	(0.187)	(0.490)	(7.885)	(0.266)	(0.352)	(0.390)	(1.100)	(0.896)	(2.190)	(0.319)	(0.305)	(2.884)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.241	0.048	0.229	0.317	0.349	0.086	0.280	0.075	0.114	0.165	0.083	0.0002	0.049	0.164

Notes: All variables are standardized such that the regression coefficient corresponds to the correlation. See Appendix Table D.1 for variable definitions and data sources. Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table D.4: Correlation Between Absolute Upward Mobility and Department Characteristics

E Additional Figures

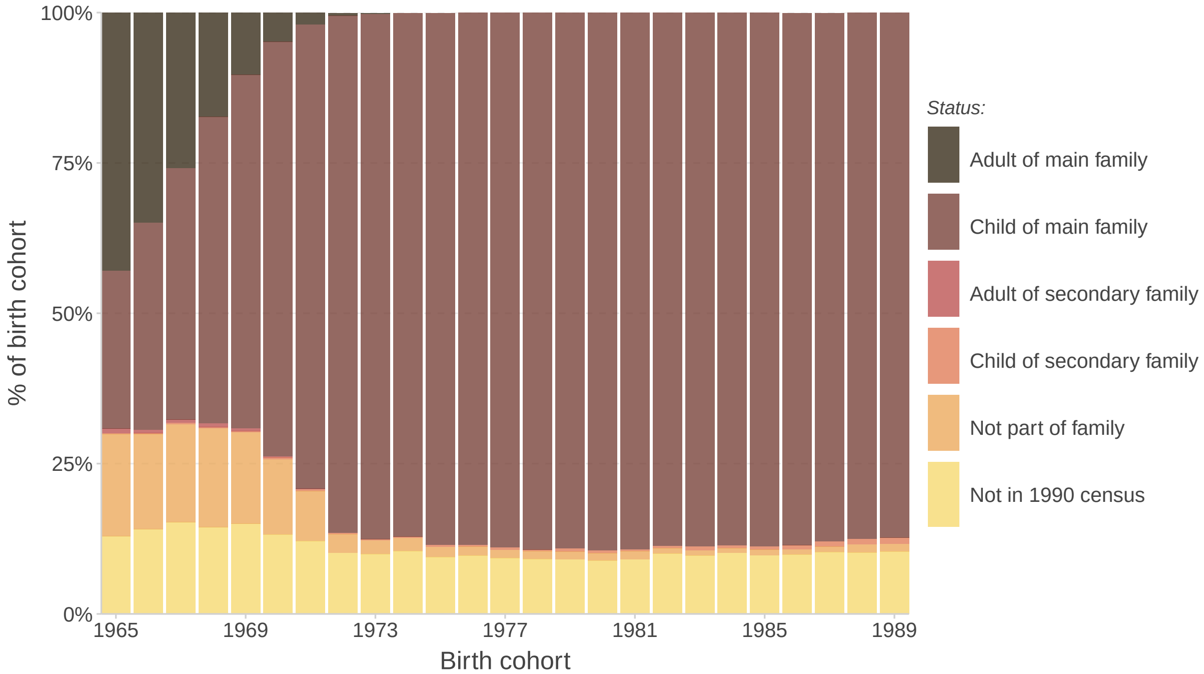


Figure E.1: Family Position in 1990 Census by Child Birth Cohort

Notes: This figure presents the family position of EDP individuals in the 1990 census by birth cohort. The sample is restricted to EDP individuals born in metropolitan France.

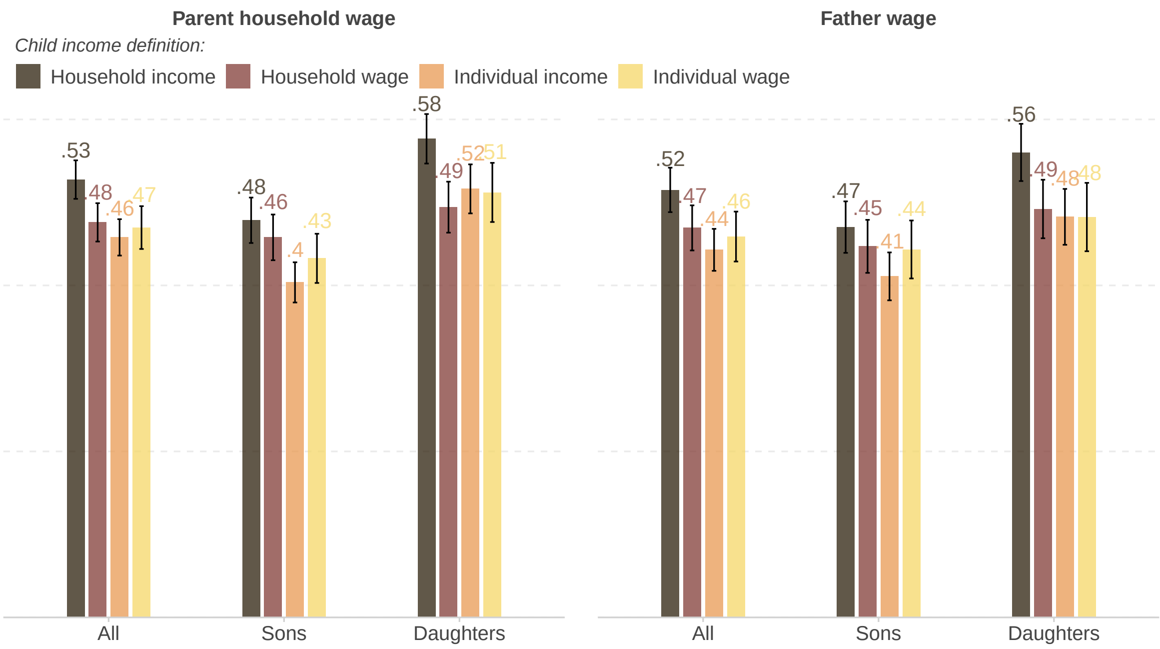


Figure E.2: Baseline IGE Estimates for Different Child and Parent Income Definitions

Notes: This figure presents our baseline intergenerational income elasticity estimates for various parent and child income definitions. Each bar represents the coefficient of an OLS regression of child income on parent income, for the entire sample (All) and for sons and daughters separately. Error bars represent the 95% bootstrapped confidence intervals. See Section 3 for for details on data, sample and income definitions.

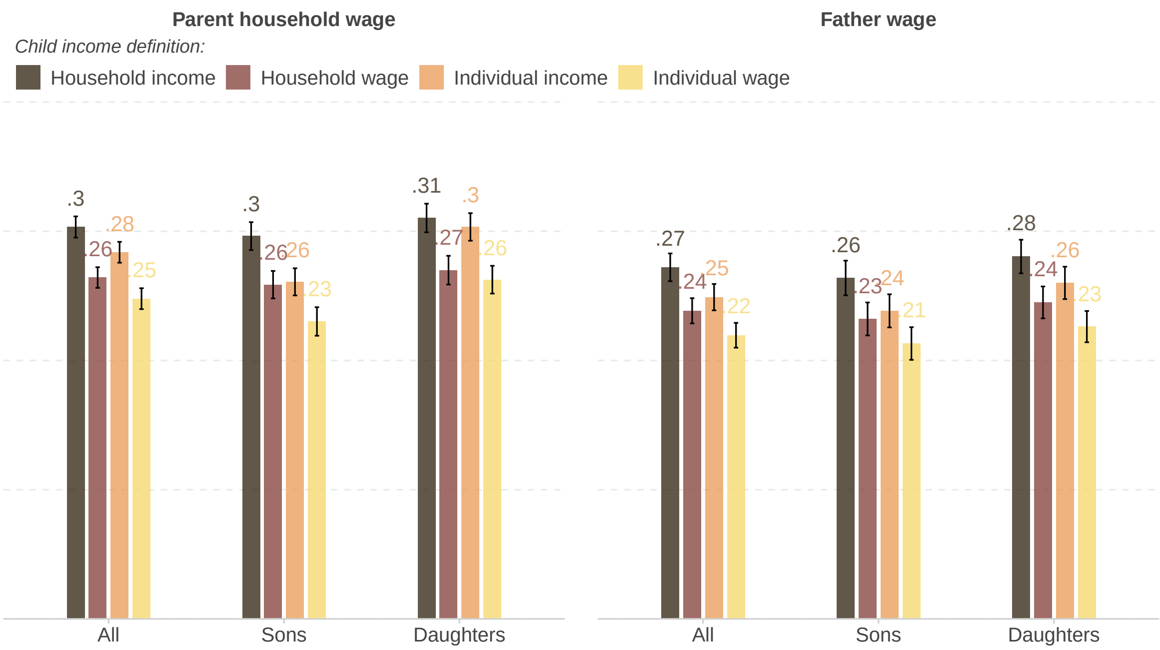


Figure E.3: Baseline RRC Estimates for Different Child and Parent Income Definitions

Notes: This figure presents our baseline intergenerational rank-rank correlation estimates for various parent and child income definitions. Each bar represents the coefficient of an OLS regression of child income rank on parent income rank, for the entire sample (All) and for sons and daughters separately. Error bars represent the 95% bootstrapped confidence intervals. See Section 3 for for details on data, sample and income definitions.

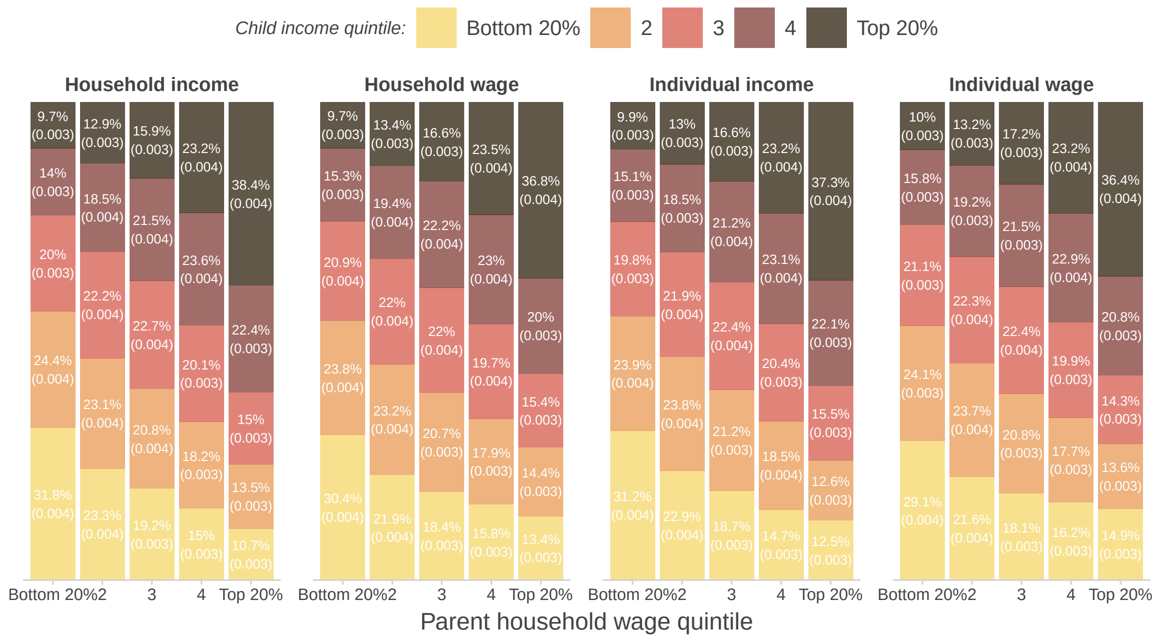


Figure E.4: Baseline Quintile Transition Matrix for Different Child Income Definitions

Notes: This figure presents our baseline intergenerational transition matrix estimates for various child income definitions, with bootstrapped standard errors in parentheses. Each cell corresponds to the percentage of children in a given income quintile among children who have parents in a given parent income quintile. See Section 3 for for details on data, sample and income definitions.

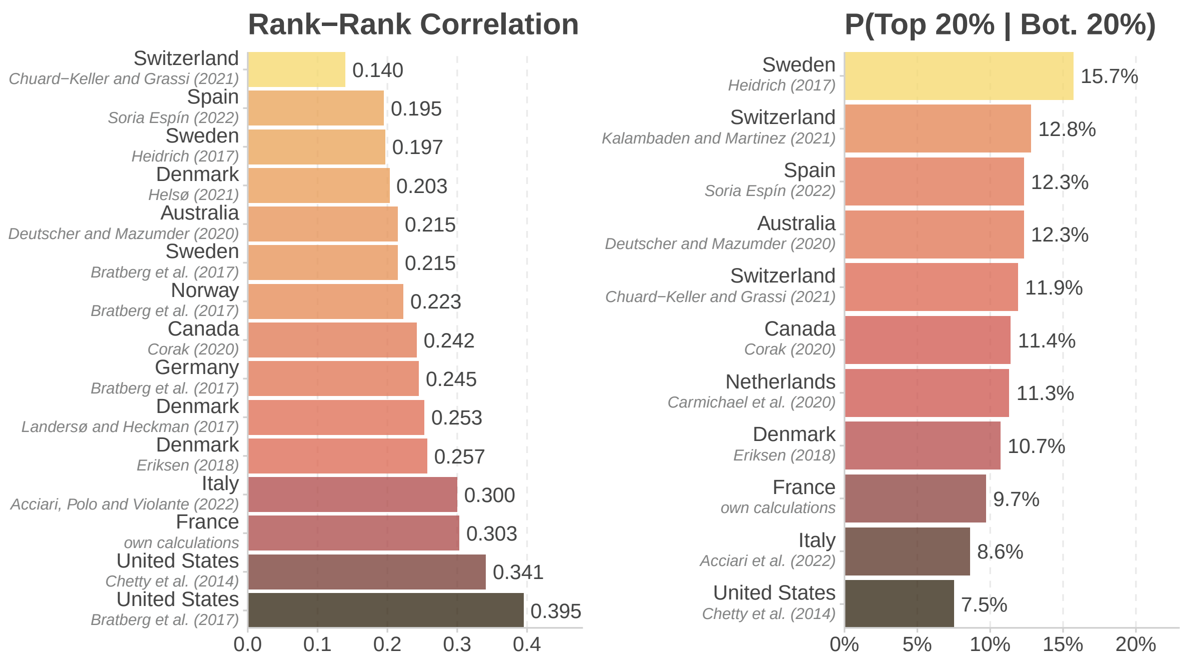


Figure E.5: Rank-Rank Correlation and Upward Mobility in International Comparison

Notes: This figure represents the international comparisons in rank-rank correlation and transition matrix cells presented in Tables 1 and 2.

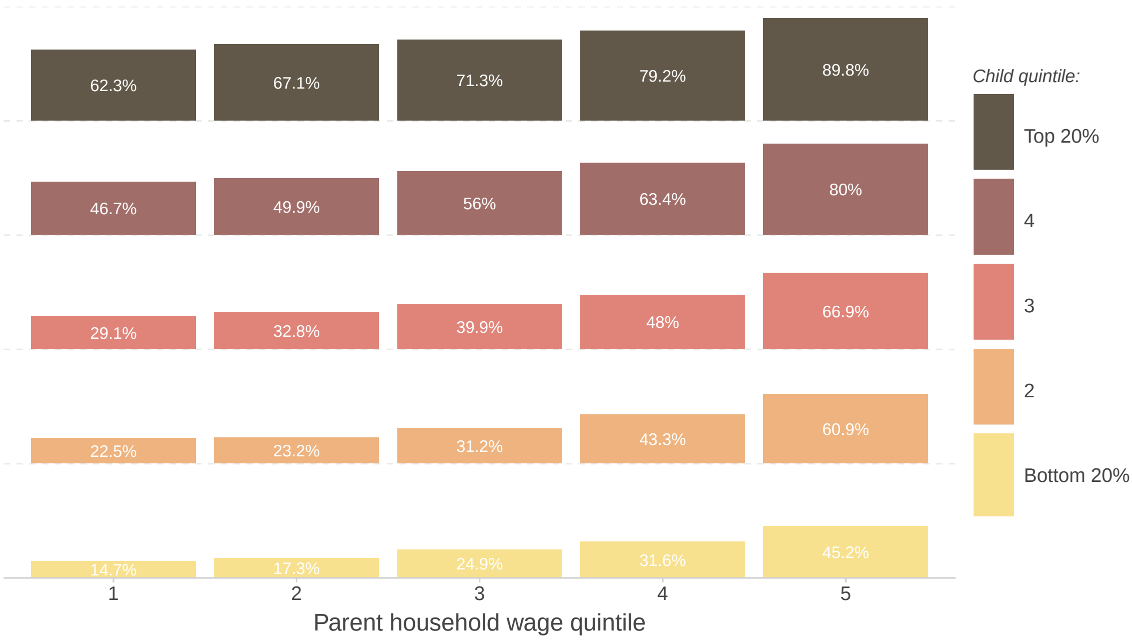


Figure E.6: Higher Education Graduation by Quintile Transition Matrix Cell

Notes: This figure presents the percentage of children graduating from higher education in each cell of the quintile transition matrix. Each cell corresponds to the percentage of children in a given income quintile coming from a family in a given parent income quintile who have a higher education diploma. See Sections 3 and 4.4 for for details on data, sample and income definitions. In this figure parent income ranks are computed without parent education in the set of first-stage predictors to avoid capturing the effect of parent education independent from that of parent income.

Figure E.7: French Departments

Notes: This figure represents the 96 metropolitan French departments. The borders of these departments have not changed over the study period. For convenience, we treat Corsica (*Haute Corse* and *Corse du Sude*) as a single department.

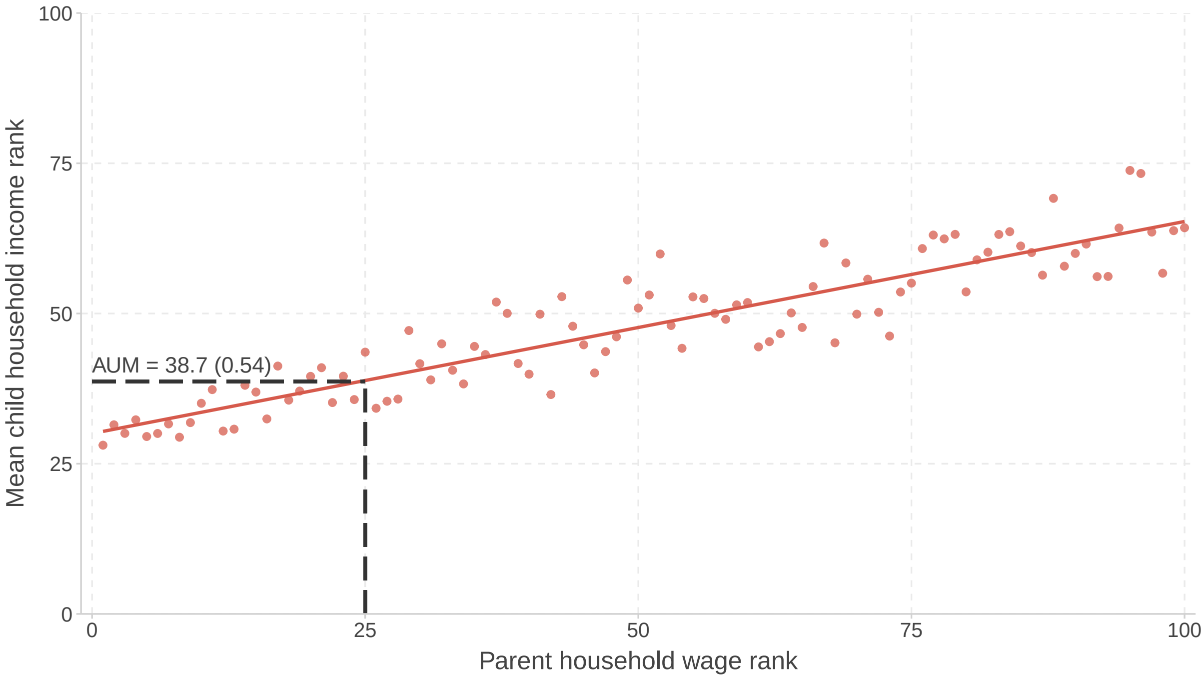


Figure E.8: Illustration of Absolute Upward Mobility for the *Nord* Department

Notes: This figure presents a non-parametric binned scatter plot of the relationship between child income rank and parent income rank for individuals who grew up in the *Nord* department. The dashed line shows the expected income rank, here 38.7 (bootstrapped standard error = 0.54), for children whose parents locate at the 25th percentile. The orange line is a linear regression fit through the conditional expectation. See Figure 3's notes for details on data, sample and income definitions.

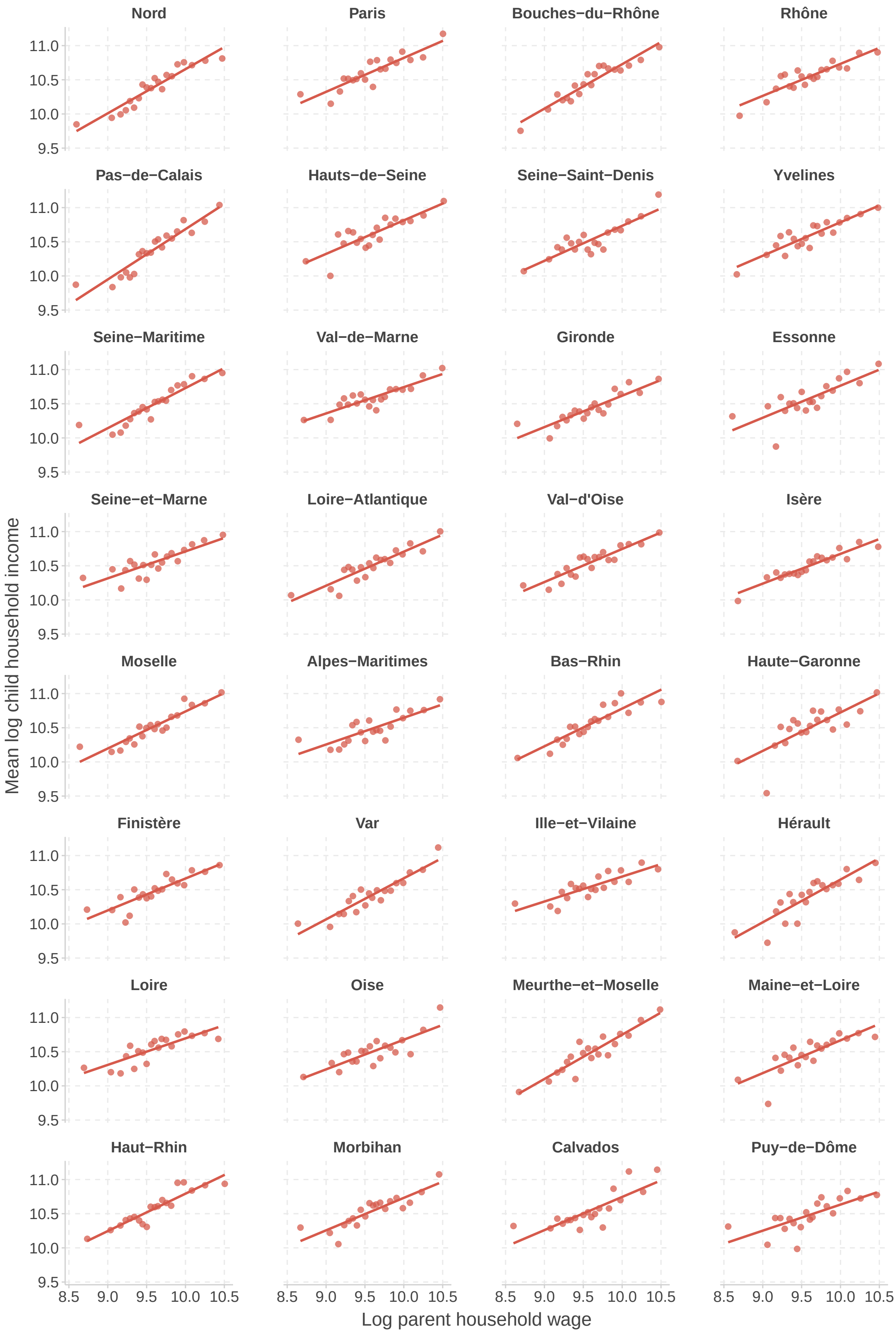


Figure E.9: Department-Level Log-Log Relationships

Notes: This figure presents the non-parametric binned scatter plot of the relationship between child log income and parent log income separately for each childhood department. The childhood department is that observed in 1990, i.e., when individuals were between 9 and 18 years old. See Figure 3's notes for details on data, sample and income definitions.

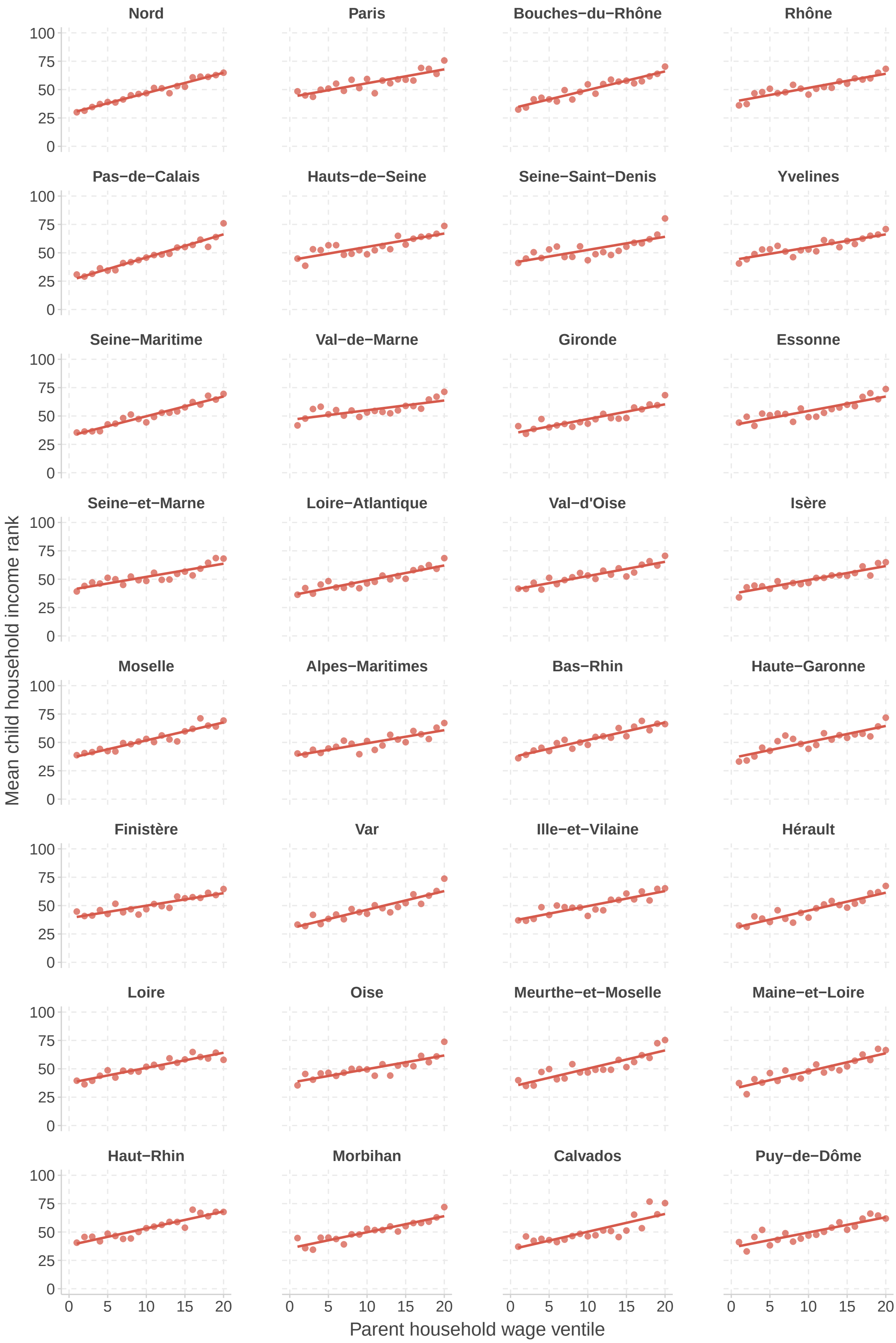


Figure E.10: Department-Level Rank-Rank Relationships

Notes: This figure presents the non-parametric binned scatter plot of the relationship between child income rank and parent income rank separately for each childhood department. The childhood department is that observed in 1990, i.e., when individuals were between 9 and 18 years old. See Figure 3's notes for details on data, sample and income definitions.

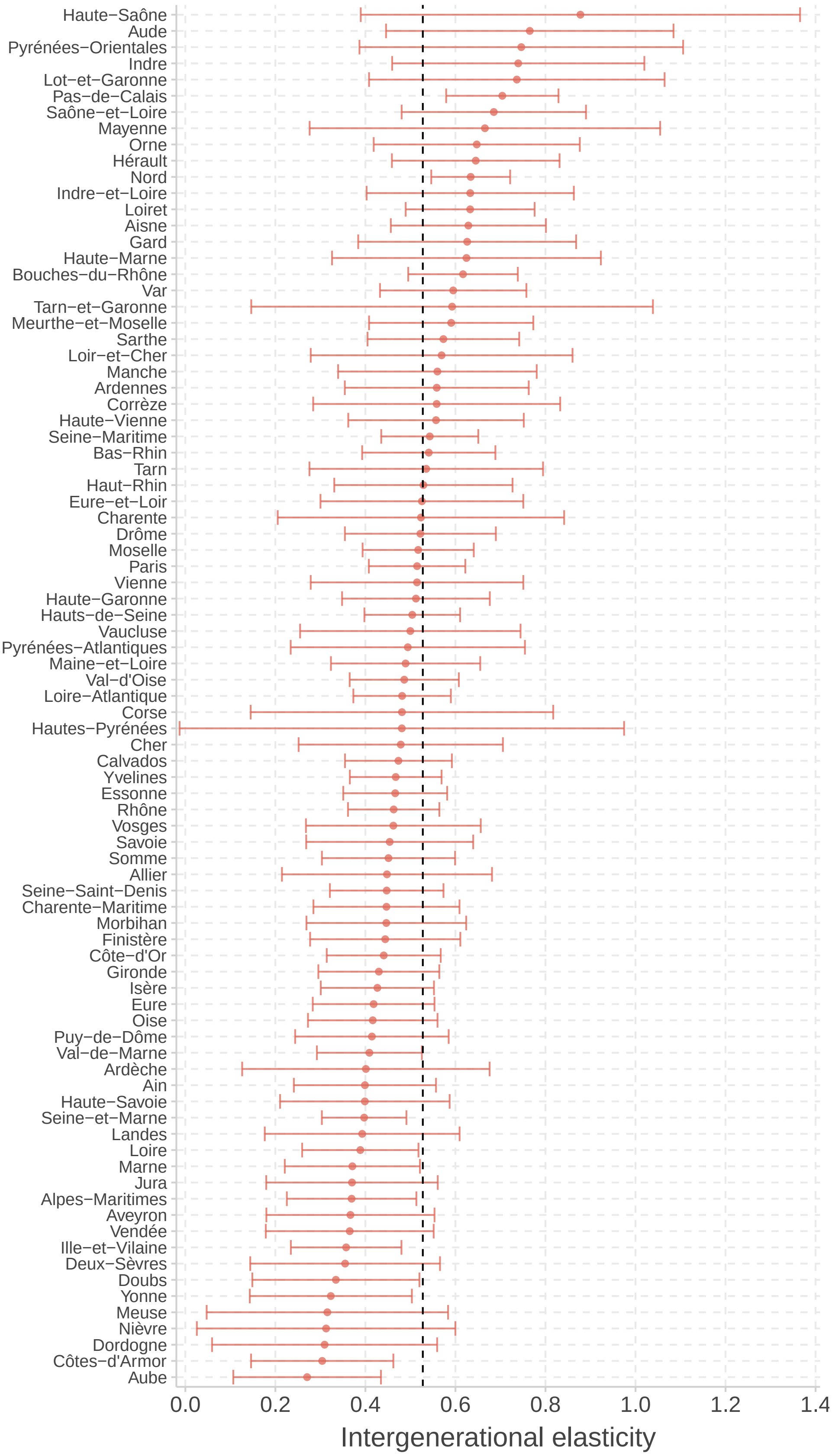


Figure E.11: Department-Level Intergenerational Elasticities

Notes: This figure presents the intergenerational elasticity in household income and its 95% bootstrapped confidence interval, estimated separately for each childhood department with more than 200 observations. The childhood department is that observed in 1990, i.e., when individuals were between 9 and 18 years old. The dashed black line represents the national estimate. See Figure 3's notes for details on data, sample and income definitions.

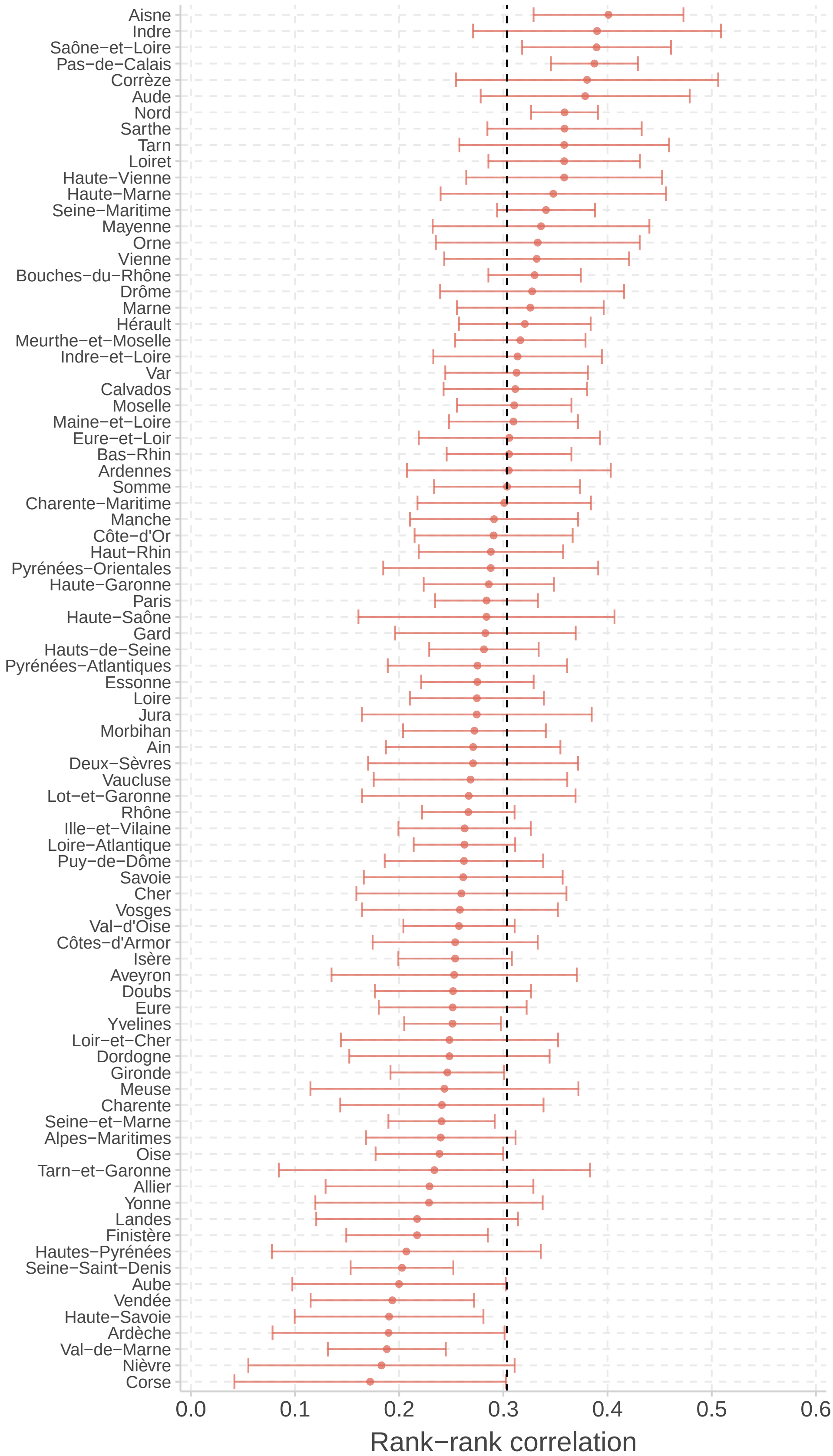


Figure E.12: Department-Level Rank-Rank Correlations

Notes: This figure presents the rank-rank correlation in household income and its 95% bootstrapped confidence interval, estimated separately for each childhood department with more than 200 observations. The childhood department is that observed in 1990, i.e., when individuals were between 9 and 18 years old. The dashed black line represents the national estimate. See Figure 3's notes for details on data, sample and income definitions.

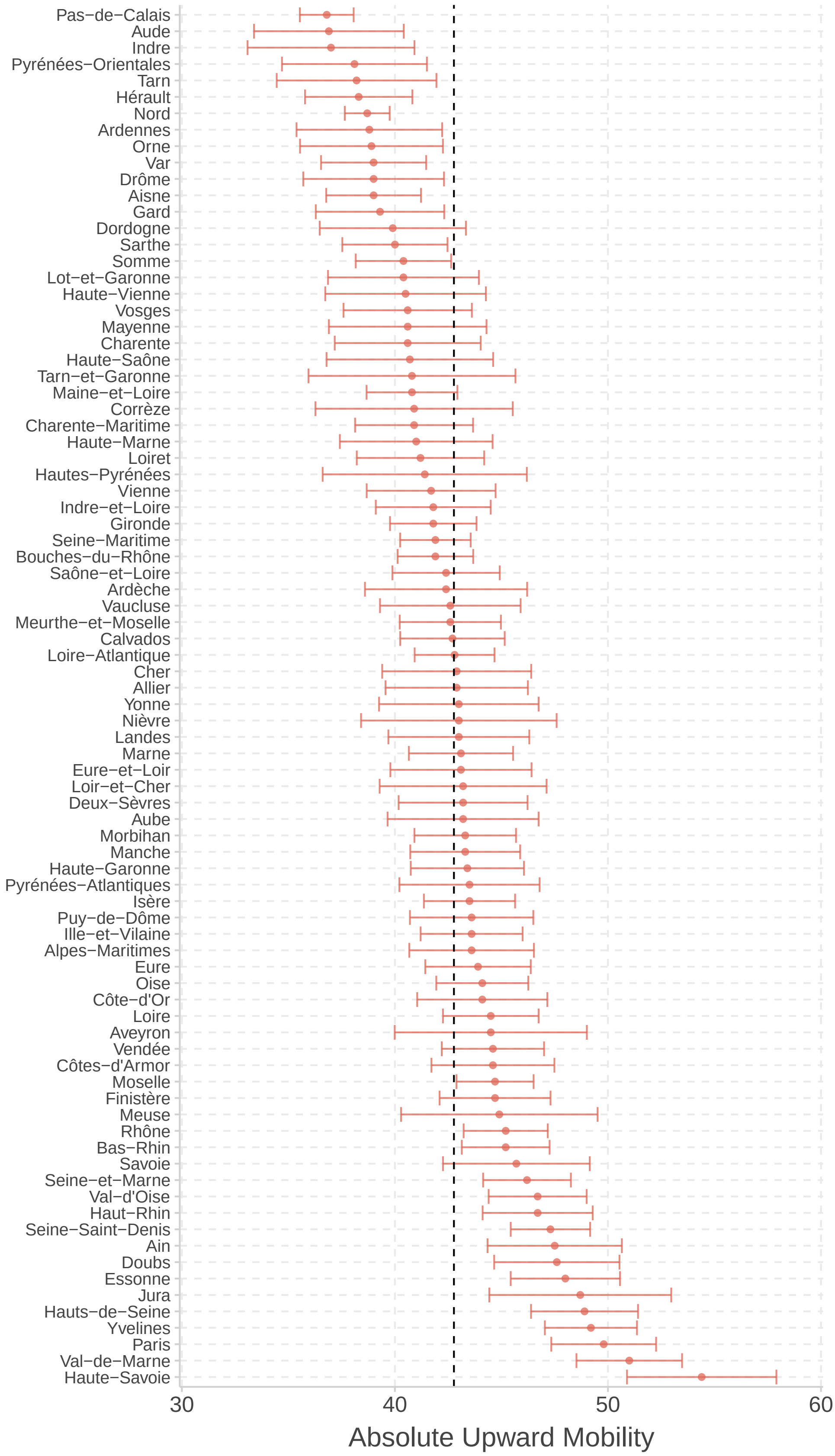


Figure E.13: Department-Level Absolute Upward Mobility

Notes: This figure presents the absolute upward mobility in household income ranks and its 95% bootstrapped confidence interval, estimated separately for each childhood department with more than 200 observations. The childhood department is that observed in 1990, i.e., when individuals were between 9 and 18 years old. The dashed black line represents the national estimate. See Figure 3's notes for details on data, sample and income definitions.

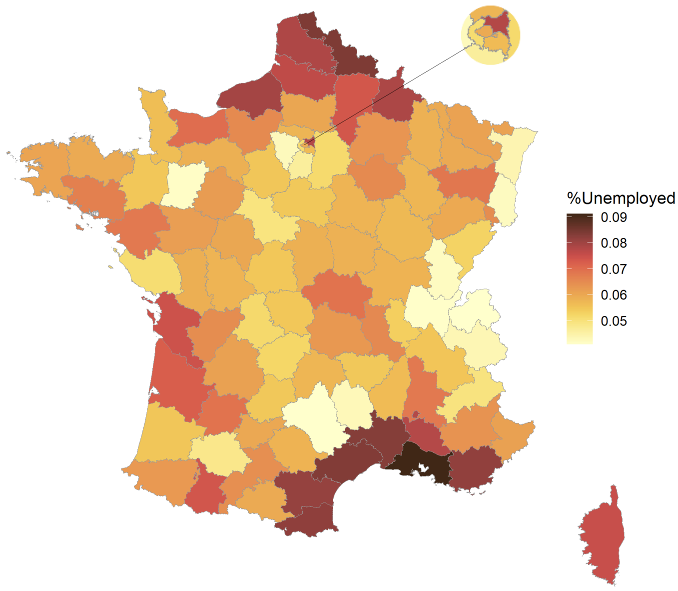


Figure E.14: Department-Level Unemployment Rate in 1990

Figure E.15: Geographic Mobility by Parent Household Wage Rank

Notes: This figure presents the percentage of movers by parent income rank. Movers are defined as individuals whose adulthood department of residence is different from that of their childhood. See Figure 3 and 10's notes for details on data, sample and income definitions.

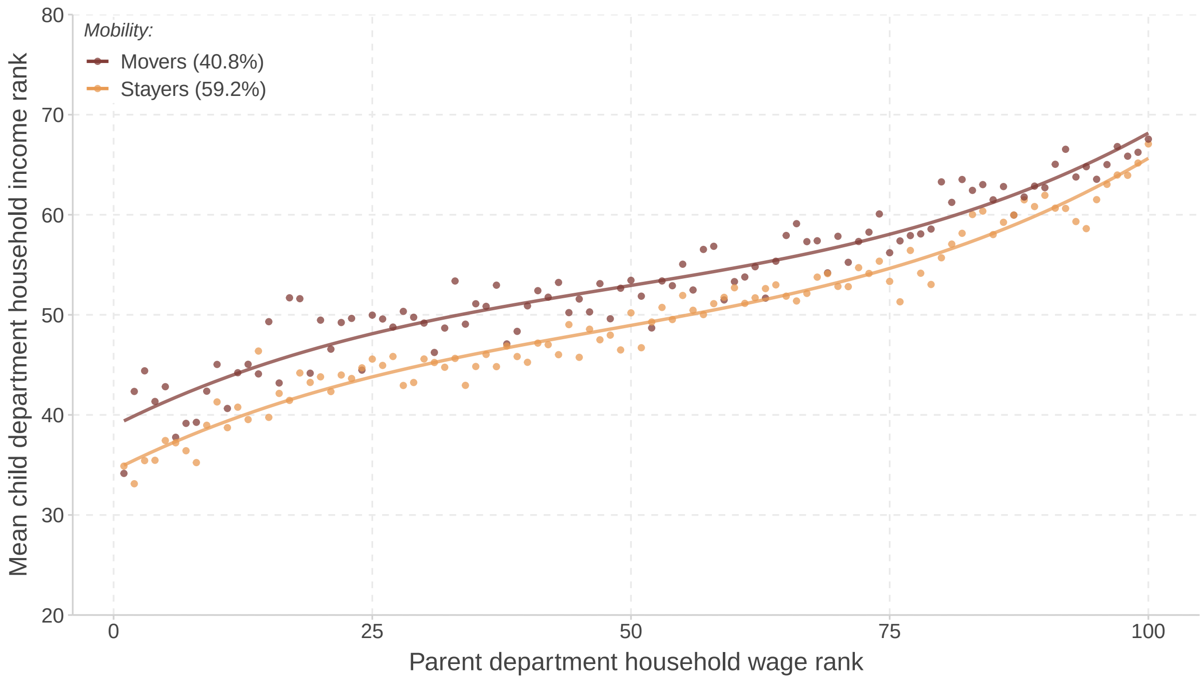


Figure E.16: Intergenerational Mobility and Geographic Mobility - Department Ranks

Notes: This figure represents the conditional expectation function of child household income rank with respect to parent household wage rank separately for individuals whose adulthood department of residence is different or not from their childhood department of residence. Percentile ranks are computed according to the local department income distribution. Parents are ranked within their department of residence in 1990 while children are ranked within their adulthood department. See Figures 3 and 10's notes for details on data, sample and income definitions.

F Additional Tables

Birth Cohort	Born in Metropolitan France	+ Live with parents in 1990 census	+ At least one obs. in tax returns data (each inc. def.)	+ At least one obs. in tax returns data at 35-45	+ No parent in occupation 1 or 31
1972	9,083	7,946	7,515	7,515	7,015
1973	8,647	7,670	7,263	7,263	6,726
1974	8,704	7,713	7,294	7,294	6,758
1975	7,334	6,565	6,230	6,230	5,818
1976	7,762	6,963	6,567	6,547	6,100
1977	7,972	7,175	6,823	6,763	6,319
1978	7,755	7,000	6,691	6,585	6,136
1979	8,473	7,620	7,280	7,102	6,644
1980	8,822	7,965	7,559	7,239	6,774
1981	8,457	7,631	7,267	6,716	6,304
1972-1981	83,009	74,248	70,489	69,254	64,594

Notes: This table displays the number of observation for each child birth year cohort and the entire sample, at each sample restriction. Note that since parent income cannot be predicted for 23 children because one of their parents have an occupation not represented in the sample of synthetic parents of the corresponding gender, the actual sample size on which estimates are computed when using parent household wage as the parent income definition is 64,571 (i.e., 64,594 - 23).

Table F.1: Child Sample Construction

Characteristic	Synthetic Parents	Actual Parents
Females	53.42%	52.26%
Age in 1990	41.22%	40.74%
Born French	89.95%	88.36%
1-digit occupation		
1. Farmers	3.72%	3.47%
2. Craftsmen, salespeople, and heads of businesses	6.98%	6.77%
3. Managerial and professional occupations	9.76%	9.35%
4. Intermediate professions	15.48%	15.35%
5. Employees	20.76%	20.39%
6. Blue collar workers	23.19%	24.6%
7. Retirees	1.30%	1.32%
8. Other with no professional activity	18.81%	18.76%
Education		
No diploma	22.45%	23.8%
Primary education	19.38%	18.93%
BEPC	7.99%	8.18%
CAP	20.76%	19.91%
BEP	4.95%	5.00%
High school diploma	11.64%	11.47%
Bachelor or technical degree	6.08%	6.18%
Masters or PhD	6.75%	6.52%
Country of birth		
France	86.18%	84.81%
Maghreb	6.62%	8.03%
Other Africa	0.55%	0.73%
South Europe	3.32%	3.33%
Other Europe	2.33%	2.17%
Rest of the world	1.00%	0.94%
Family structure		
Single fathers	0.93%	0.72%
Single mothers	5.58%	5.25%
Both spouses active	58.73%	58.28%
Mother inactive	31.35%	32.32%
Father inactive	1.38%	1.38%
Both spouses inactive	2.03%	2.06%
Municipality characteristics		
Log population	7.83	7.85
Log density	0.46	0.49
% foreigners	2.31%	2.33%
Unemployment rate	6.22%	6.25%
% single mothers	6.3%	6.4%
N	134,572	140,136

Notes: See Section 3.2 for details on construction of samples. These statistics are computed before applying any income observation restrictions.

Table F.2: Average Characteristics of Actual and Synthetic Parents

2-digit occupation	Synthetic Parents	Actual Parents
Farmers with small farm	0.92%	0.84%
Farmers with medium-sized farm	1.22%	1.19%
Farmers with large farm	1.58%	1.44%
Craftsmen	3.62%	3.57%
Trade workers and related	2.62%	2.50%
Heads of company with $\geq$ 10 employees	0.73%	0.70%
Liberal profession	1.38%	1.32%
Public sector executives	1.07%	1.05%
Professors and scientific professions	2.12%	1.97%
Information, arts, and entertainment professions	0.32%	0.31%
Administrative executives and sales representatives	2.72%	2.66%
Engineers, technical executives	2.16%	2.05%
Teachers and related	2.64%	2.57%
Intermediate health and social work professions	2.48%	2.62%
Clerk, religious	0.01%	0.01%
Intermediate administrative professions of the public sector	1.54%	1.41%
Intermediate administrative professions and salesmen	4.06%	4.03%
Technicians	2.30%	2.29%
Foremen, supervisors	2.44%	2.42%
Civil servants	6.74%	6.69%
Police and military officers	1.27%	1.35%
Company administrative employees	6.92%	6.70%
Trade employees	2.24%	2.16%
Personal service workers	3.58%	3.49%
Skilled industrial workers	5.82%	6.14%
Skilled crafts workers	4.60%	4.83%
Drivers	2.19%	2.39%
Skilled handling, storing and transport workers	1.41%	1.47%
Unskilled industrial workers	6.19%	6.67%
Unskilled crafts workers	2.32%	2.42%
Agricultural workers	0.66%	0.69%
Former farmers	0.09%	0.07%
Former craftsmen, salespeople, and heads of businesses	0.10%	0.08%
Former managerial and professional occupation	0.09%	0.10%
Former intermediate professions	0.19%	0.17%
Former employees	0.33%	0.30%
Former blue collar workers	0.51%	0.60%
Unemployed who have never worked	0.36%	0.38%
Military contingent	0.00%	0.00%
Students $\geq$ 15 yrs old	0.10%	0.04%
Other inactive $\leq$ 60 yrs old	18.24%	18.20%
Other inactive $\geq$ 60 yrs old	0.10%	0.12%
N	$134,572$	$140,136$

Notes: See Table F.2's notes for sample construction.

Table F.3: Share of Actual and Synthetic Parents by 2-Digit Occupation

Gender	At least one child born in Metrop. France 1972-1981	+ Observed in 1990 Census	+ Born even year	+ At least two obs. at 35-45 in All Employee Panel	+ Not in occupation 1 or 31
Fathers	49,746	43,851	22,227	16,699	16,450
Mothers	52,904	48,097	24,297	15,104	14,973
All	102,650	91,948	46,524	31,803	31,423

Table F.4: Synthetic Parents Sample Construction

Child income definition	Parent income definition	Number of observations	0 child incomes (N.)	0 child incomes (%)	Negative child incomes (N.)	Negative child incomes (%)
Household income	Family income	64,571	0	0	41	0.06
Household income	Father income	57,902	0	0	35	0.06
Household wage	Family income	64,571	1976	3.06	0	0
Household wage	Father income	57,902	1690	2.92	0	0
Individual income	Family income	64,571	2479	3.84	68	0.11
Individual income	Father income	57,902	2162	3.73	60	0.1
Labor income	Family income	64,571	4990	7.73	0	0
Labor income	Father income	57,902	4376	7.56	0	0

Table F.5: Number of Observations by Child and Parent Income Definitions

	N	Missing (%)	Mean	Std. Dev.	25^{th} pctile	Median	75^{th} pctile
Synthetic Parents							
Synthetic father income (35-45 yrs old)	16,450	0	25,902	17,265	16,251	21,966	30,427
Number of syn. father income observations	16,450	0	7.66	2.42	6	8	9
Synthetic mother income (35-45 yrs old)	14,973	0	15,167	10,143	7,496	14,140	21,027
Number of syn. mother income observations	14,973	0	6.95	2.84	5	7	9
Parents							
Fraction single parents in 1990	11.72%						
Fraction female among single parents	88.3%						
Father age at child's birth	64,594	10.35	28.48	6.08	24	28	31
Mother age at child's birth	64,594	1.37	25.89	5.15	22	25	29
Father age in 1990	64,594	10.35	41.98	6.61	38	41	45
Mother age in 1990	64,594	1.37	39.42	5.81	35	39	43
Children							
Household income (average 2010-16)	64,594	0	46,599	38,371	27,696	41,417	56,481
Household wage (average 2010-16)	64,594	0	38,460	30,184	20,812	35,205	50,096
Individual income (average 2010-16)	64,594	0	23,512	20,471	14,375	21,159	28,737
Labor income (average 2010-16)	64,594	0	21092	19,120	10,067	19,877	27,487
Fraction female	49.77%						

Notes: See Sections 3.2 and 3.3 for details on sample construction and income definitions.

Table F.6: Descriptive Statistics

	Intergenerational Elasticity	First-Stage Instruments	Data	Income Definitions	Child Age
Lefranc and Trannoy (2005)	0.4-0.438[1]	Education (8 cat.) + occupation (7 cat.)	FQP	labor earnings (excl. UI)[2]	30-40
Lefranc (2018)	0.577[3]	Education (6 cat.)	FQP	labor earnings (excl. UI)[2]	28-32
EqualChances.org	0.357	Education (3 cat.) + occupation (9 cat.)	Synthetic fathers: ECHP Sons: EU-SILC	net personal employee income	-
Our estimate	**0.443**				

Notes: FQP = Formation-Qualification-Profession; ECHP = European Community Household Panel; EU-SILC = European Union Statistics on Income and Living Conditions

[1] Estimates taken from Table I, Panel A, cols. (1)-(4), p.65.

[2] Only salaried workers.

[3] Estimates taken from Table 2, 1971-75, col. (2), p.823.

Table F.7: Comparison with Existing Father-Son IGE Estimates for France

	IGE	RRC	AUM
First-stage MSE	-0.160	-0.088	1.400
	(0.127)	(0.056)	(3.487)
Constant	0.565^{***}	0.318^{***}	42.370^{***}
	(0.053)	(0.024)	(1.465)
Observations	85	85	85
R^2	0.019	0.029	0.002

Notes: Standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table F.8: Department-Level MSEs and Measures of Intergenerational Income Mobility

	Department	Observations	IGE	RRC	AUM
01	Ain	535	0.4	0.27	47.5
02	Aisne	735	0.63	0.4	39
03	Allier	365	0.45	0.23	42.9
04	Alpes-de-Haute-Provence	141	*	*	*
05	Hautes-Alpes	112	*	*	*
06	Alpse-Maritimes	773	0.37	0.24	43.6
07	Ardèche	313	0.4	0.19	42.4
08	Ardennes	376	0.56	0.31	38.8
09	Ariège	121	*	*	*
10	Aube	361	0.27	0.2	43.2
11	Aude	274	0.77	0.38	36.9
12	Aveyron	243	0.37	0.25	44.5
13	Bouches-du-Rhône	1,795	0.62	0.33	41.9
14	Calvados	781	0.47	0.31	42.7
15	Cantal	164	*	*	*
16	Charente	374	0.52	0.24	40.6
17	Charente-Maritime	559	0.45	0.3	40.9
18	Cher	370	0.48	0.26	42.9
19	Corrèze	219	0.56	0.38	40.9
20	Corse	236	0.48	0.17	45.6
21	Côte-d'Or	549	0.44	0.29	44.1
22	Côtes-d'Armor	590	0.3	0.25	44.6
23	Creuse	102	*	*	*
24	Dordogne	337	0.31	0.25	39.9
25	Doubs	635	0.33	0.25	47.6
26	Drôme	435	0.52	0.33	39
27	Eure	738	0.42	0.25	43.9
28	Eure-et-Loire	505	0.53	0.31	43.1
29	Finistère	979	0.44	0.22	44.7
30	Gard	577	0.63	0.28	39.3
31	Haute-Garonne	949	0.51	0.29	43.4
32	Gers	136	*	*	*
33	Gironde	1,304	0.43	0.25	41.8
34	Hérault	788	0.65	0.32	38.3
35	Ille-et-Vilaine	1,036	0.36	0.26	43.6
36	Indre	235	0.74	0.39	37
37	Indre-et-Loire	597	0.63	0.31	41.8
38	Isère	1,217	0.43	0.25	43.5
39	Jura	269	0.37	0.27	48.7
40	Landes	326	0.39	0.22	43
41	Loir-et-Cher	357	0.57	0.25	43.2
42	Loire	901	0.39	0.27	44.5
43	Haute-Loire	194	*	*	*
44	Loire-Atlantique	1,467	0.48	0.26	42.8

Notes: * Insufficient number of observations (< 200).

Table F.9: Department-Level Intergenerational Mobility Estimates

	Department	Observations	IGE	RRC	AUM
45	Loiret	706	0.63	0.36	41.2
46	Lot	137	*	*	*
47	Lot-et-Garonne	319	0.74	0.27	40.4
48	Lozère	63	*	*	*
49	Maine-et-Loire	931	0.49	0.31	40.8
50	Manche	566	0.56	0.29	43.3
51	Marne	676	0.37	0.33	43.1
52	Haute-Marne	263	0.62	0.35	41
53	Mayenne	329	0.67	0.34	40.6
54	Meurthe-et-Moselle	862	0.59	0.32	42.6
55	Meuse	238	0.32	0.24	44.9
56	Morbihan	778	0.45	0.27	43.3
57	Moselle	1,274	0.52	0.31	44.7
58	Nièvre	251	0.31	0.18	43
59	Nord	3,668	0.63	0.36	38.7
60	Oise	1,008	0.42	0.24	44.1
61	Orne	357	0.65	0.33	38.9
62	Pas-de-Calais	2,145	0.7	0.39	36.8
63	Puy-de-Dôme	664	0.41	0.26	43.6
64	Pyrénées-Atlantiques	571	0.49	0.28	43.5
65	Hautes-Pyrénées	209	0.48	0.21	41.4
66	Pyrénées-Orientales	356	0.75	0.29	38.1
67	Bas-Rhin	1,033	0.54	0.31	45.2
68	Haut-Rhin	792	0.53	0.29	46.7
69	Rhône	1,583	0.46	0.27	45.2
70	Haute-Saône	273	0.88	0.28	40.7
71	Saône-et-Loire	661	0.69	0.39	42.4
72	Sarthe	635	0.57	0.36	40
73	Savoie	430	0.45	0.26	45.7
74	Haute-Savoie	629	0.4	0.19	54.4
75	Paris	1,352	0.51	0.28	49.8
76	Seine-Maritime	1,547	0.54	0.34	41.9
77	Seine-et-Marne	1,529	0.4	0.24	46.2
78	Yvelines	1,645	0.47	0.25	49.2
79	Deux-Sèvres	376	0.35	0.27	43.2
80	Somme	737	0.45	0.3	40.4
81	Tarn	354	0.54	0.36	38.2
82	Tarn-et-Garonne	202	0.59	0.23	40.8
83	Var	773	0.59	0.31	39
84	Vaucluse	468	0.5	0.27	42.6
85	Vendée	627	0.37	0.19	44.6
86	Vienne	464	0.51	0.33	41.7
87	Haute-Vienne	357	0.56	0.36	40.5
88	Vosges	504	0.46	0.26	40.6

Notes: * Insufficient number of observations ($<$ 200).

Table F.9: Department-Level Intergenerational Mobility Estimates *(continued)*

	Department	Observations	IGE	RRC	AUM
89	Yonne	388	0.32	0.23	43
90	Territoire de Belfort	172	*	*	*
91	Essonne	1,302	0.47	0.28	48
92	Hauts-de-Seine	1,248	0.5	0.28	48.9
93	Seine-Saint-Denis	1,495	0.45	0.2	47.3
94	Val-de-Marne	1,188	0.41	0.19	51
95	Val-d'Oise	1,366	0.49	0.26	46.7

Notes: * Insufficient number of observations ($<$ 200).

Table F.9: Department-Level Intergenerational Mobility Estimates *(continued)*

Child income definition	IGE-RRC	RRC-AUM	IGE-AUM
Household income	0.65	−0.57	−0.55
Individual income	0.72	−0.55	−0.45
Individual wage	0.70	−0.41	−0.26

Notes: See Figure 8 for corresponding maps.

Table F.10: Correlation Between Department-Level Intergenerational Mobility Measures

	Dependent variable: Child household income rank				
	(1)	(2)	(3)	(4)	(5)
Parents income rank	0.259***	0.259***	0.258***	0.162***	0.135***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.012)
Mover ($\hat{\gamma}$)	4.572***	4.591***	4.897***	4.926***	4.883***
	(0.472)	(0.471)	(0.478)	(0.477)	(0.478)
Parents income rank × Mover ($\hat{\delta}$)	−0.014*	−0.014*	−0.016**	−0.026***	−0.027***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	36.401***	36.137***	35.574***	26.815***	28.162***
	(0.265)	(0.279)	(1.125)	(1.570)	(1.620)
Birth cohort	✓	✓	✓	✓	✓
Gender		✓	✓	✓	✓
Department FE			✓	✓	✓
Parents' education				✓	✓
Parents' 2-digit occupation					✓
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_{p,i}] = \hat{\gamma} + \hat{\delta} \times 50.5$	3.87	3.88	4.09	3.61	3.52
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_p \vert p_p = 25]$	4.22	4.24	4.50	4.28	4.21
$\mathbb{E}[\hat{\gamma} + \hat{\delta} p_p \vert p_p = 75]$	3.52	3.54	3.70	2.98	2.86
Observations	64,571	64,571	64,571	64,571	64,571
Adjusted R^2	0.074	0.074	0.077	0.089	0.095

Notes: Bootstrapped standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table F.11: Intergenerational & Geographic Mobility - Department Ranks

	Dependent variable: Child household income rank				
	(1)	(2)	(3)	(4)	(5)
Parents income rank	0.278***	0.278***	0.271***	0.171***	0.153***
	(0.005)	(0.005)	(0.005)	(0.008)	(0.017)
Destination department (ref.: stayers)					
Low-income	0.902	0.923	1.046*	0.852	0.685
	(0.618)	(0.618)	(0.616)	(0.616)	(0.616)
Medium-income	11.355***	11.373***	10.846***	11.045***	11.027***
	(0.951)	(0.952)	(0.945)	(0.948)	(0.950)
High-income	18.819***	18.839***	18.265***	18.465***	18.567***
	(1.224)	(1.224)	(1.247)	(1.260)	(1.258)
Parents income rank × Low-income	−0.019*	−0.019*	−0.017	−0.020*	−0.018
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Parents income rank × Medium-income inc	−0.042***	−0.042***	−0.038***	−0.051***	−0.052***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Parents income rank × High-income	−0.035**	−0.035**	−0.035**	−0.054***	−0.058***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Constant	34.143***	33.860***	37.460***	28.392***	29.369***
	(0.261)	(0.277)	(1.213)	(1.655)	(1.779)
Birth cohort	✓	✓	✓	✓	✓
Gender		✓	✓	✓	✓
Department FE			✓	✓	✓
Parents' education				✓	✓
Parents' 2-digit occupation					✓
Observations	64,571	64,571	64,571	64,571	64,571
Adjusted R^2	0.118	0.118	0.124	0.135	0.142

Notes: Bootstrapped standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table F.12: Intergenerational Mobility and Income Level in the Destination Department