SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

THE ANATOMY OF INTERGENERATIONAL INCOME MOBILITY IN FRANCE AND ITS SPATIAL VARIATIONS

Gustave Kenedi and Louis Sirugue

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2021-09

The Anatomy of Intergenerational Income Mobility in France and its Spatial Variations*

Gustave Kenedi†
Sciences Po

Louis Sirugue‡
Paris School of Economics

This Version: November 2021

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 procedure. At the national level, every measure of intergenerational income persistence (intergenerational elasticities, rank-rank correlations, and transition matrices) suggests that France is characterized by relatively strong persistence relative to other developed countries. Children born to parents in the bottom 20% of their income distribution have a 10.1% probability of reaching the top 20% as adults. This probability is of 39.1% for children born to parents in the top 20%. At the local level, we find substantial spatial variations in intergenerational mobility. It is higher in the West of France and particularly low in the North and in the South. We uncover significant relationships between absolute upward mobility and characteristics of the environment an individual grew up in, such as the unemployment rate, population density, and income inequality.

JEL Codes: J62, R23, C18

Keywords: intergenerational mobility, measurement, spatial variations, France

*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), and the French Collaborative Institute on Migration, coordinated by the CNRS (reference ANR-17-CONV-0001). This paper benefited from helpful feedback and suggestions from Pierre Cahuc, Pierre-Philippe Combes, Clément Dherbécourt, Thomas Piketty, Benoît Schmutz, Patrick Simon, Yann Thommen and Alain Trannoy as well as participants at the 2021 Augustin Cournot Doctoral Days, the 69th Congress of the French Economic Association, and seminars at CREST, PSE, and Sciences Po.

†Email: gustavekenedi@gmail.com

‡Email: louis.sirugue@psemail.eu - PSE, CNRS (UMR 8545), EHESS, CIMigration

1 Introduction

To what extent are individuals' incomes related to those of their parents? Over the past two decades, this question has seen renewed interest, both in the general public and in academia, as discussions about income inequality inevitably raise concerns about equality of opportunity. Examining this link is essential to understand whether the same opportunities are afforded to children from different socio-economic backgrounds, as well as for economic efficiency as high persistence across generations may reflect inefficient allocation of talents (so-called "Lost Einsteins"). There are now numerous studies providing estimates of persistence for a number of advanced economies and conducting cross-country comparisons to uncover the potential mechanisms at play. Our aim is to provide comparable estimates for France.

The most recent literature on intergenerational income persistence has differed in several respects compared to older studies. First, it has expanded the set of measures used to characterize the extent of intergenerational income mobility. In addition to the traditional intergenerational income elasticity (IGE), which measures the expected percentage change in child income associated with a 1% increase in parent income, the rank-rank correlation (RRC), measuring the correlation between child and parent income percentile ranks, has been proposed. The RRC has been shown to be less sensitive to various statistical considerations and more robust to idiosyncratic sample selection choices. Additionally, transition matrices have gained in popularity due to their ease of interpretation and their ability to capture non-linearities in intergenerational mobility along the parent income distribution. We estimate all three measures.

Second, the new literature has shifted its attention to household-level income measures as opposed to the historic father-son labor earnings analyses, thanks to the availability of rich administrative data. These income measures provide a better depiction of one's economic resources and importantly enable the analysis of mothers and daughters who had been largely ignored. Third, pioneered by Chetty et al. (2014), it has analyzed intergenerational mobility at subnational levels, highlighting significant spatial variations in intergenerational mobility across local areas. These analyses are replacing cross-country comparisons to help shed light on the mechanisms that may underlie income persistence across generations and hint at the types of policies likely to remediate such intergenerational inequalities. We implement these improvements and analyze spatial variations in intergenerational persistence.

Despite these developments in the measurement of intergenerational mobility for many countries, much remains to be known for France, a country with relatively modest post-tax and transfers income inequality in international comparison due to progressive taxation and substantial social transfers. Existing studies only estimate the intergenerational income elasticity and assess its evolution over time (Lefranc and

Trannoy, 2005; Lefranc, 2018). Using the Permanent Demographic Sample (*Échantillon Démographique Permanent*), a rich administrative dataset on individuals born on the first four days of October, we characterize the extent of income persistence for almost 65,000 children born between 1972 and 1981 using three statistics: (i) the intergenerational income elasticity; (ii) the rank-rank correlation; and (iii) quintile-by-quintile transition matrices. The richness of the data allows us to implement the improvements of the most recent literature and to convincingly address concerns related to the well-established lifecycle and attenuation biases (Haider and Solon, 2006; Black and Devereux, 2011; Nybom and Stuhler, 2017).

Since parents' incomes are not observed, we use the two-sample two-stage least squares estimation introduced by Björklund and Jäntti (1997) and previously employed by Lefranc and Trannoy (2005) and Lefranc (2018) in the French context.[1] It consists in a two-step procedure: (i) estimating a prediction model using parents drawn from the same population as our actual parents but for whom income is observed (we call this sample "synthetic parents"), based on variables that are also available for the actual parents, and (ii) predicting actual parents' incomes based on this model. Our set of predictors for (i) includes demographic characteristics of parents in 1990 (age, French nationality dummy, country of birth (6 categories), 2-digit occupation (42 cat.), education (8 cat.), 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). We perform robustness checks to ensure our results are not driven by the choice of first-stage predictors. Parent income is defined as the average of predicted pretax wage[2] over ages 35-45 for both parents, and child income as average pretax household income[3] within the same age range between 2010 and 2016. These two income definitions represent the most comprehensive household-level income definitions possible for either generation.

National Results. Our baseline estimate of the intergenerational elasticity in household income is 0.515, suggesting that, on average, a 10% increase in parent income is associated with a 5.15%[4] increase in child income. Another way to illustrate this is that, on average, if one's parents earn 10% more than the average of parents' incomes, then one is expected to preserve roughly 50% of that advantage relative to the average of one's cohort's average incomes. This confirms that France exhibits relatively strong intergenerational income persistence compared other developed countries. Our com-

[1] As well as in many other countries, see Jerrim et al. (2016, Table A1).

[2] Self-employment income is not observed and therefore not included in our parent income measure.

[3] Defined as the sum of labor earnings (wages + self-employment income), taxable capital income and predicted non-taxable capital income, unemployment insurance, retirement and alimony. Social benefits such as family allowances, social minima (e.g. RSA, disability) and housing benefits are not included in this definition. See Section 4.2 for details.

[4] The exact expected change is equal to $(1.1^{0.515} - 1) \times 100 \approx 5.03\%$.

parable father-son wage IGE is around 0.4, between Germany (0.32) and the United States (0.47), and far from Scandinavian countries (around 0.2) (Corak, 2016).

Moving to the rank-rank relationship, we find that the conditional expectation of child income percentile rank (relative to other children in the same birth cohort) with respect to parent income percentile rank (relative to other parents with children in the same birth cohort) is linear for most of the parent income distribution, with tighter relationships at the tails. Our baseline estimate of the rank-rank correlation is 0.337, implying that a 10 percentile increase in parent income rank is associated, on average, with a 3.37 percentile increase in child income rank. This estimate is of similar magnitude to that found for the United States (0.341; Chetty et al. (2014)), but significantly higher 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)).[5] Consistent with the literature, our rank-based estimates are remarkably robust to various sample selection choices, making them potentially more reliable than IGEs in contexts where parents' incomes are not observed.

Intergenerational persistence is strongest at the tails of the parent income distribution: 10.1% of children born to parents in 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% (39.1%). In comparison, in the US, only 7.5% of children born to families in the bottom 20% reach the top 20% as adults (Chetty et al., 2014), while this statistic is 12.3% in Australia (Deutscher and Mazumder, 2020). Moreover, persistence at the top becomes stronger and stronger as we zoom in the right tail of the parent income distribution. Once more, our transition matrices highlight that in international comparison, France appears to have some of the strongest persistence in income positions for children from the bottom and top of the income distribution.

We assess the robustness of our baseline results to a number of statistical biases. Foremost, we evaluate how sensitive they are to the lifecycle and attenuation biases by varying the ages at which child and (synthetic) parent incomes are measured as well as the number of (synthetic) parent income observations used. We find that our baseline results do not appear to under- nor over-estimate intergenerational mobility due to measuring child and/or parent income too early or too late in the lifecycle or because of averaging incomes over too few years. Moreover, we check whether using machine learning algorithms and varying the set of first-stage predictors influences our estimates. Slightly improved prediction from using flexible machine learning algorithms does not quantitatively alter our estimates, while the set of first-stage predictors appears to only matter if 2-digit occupation is not included. Lastly, we evaluate whether trimming the tails of child or parent income distributions significantly affects our estimates and find that it matters significantly for the IGE but not that much for the RRC.

[5] See Table 2 for a comparison with all existing RRC estimates.

Subnational Results. We uncover large spatial variations in intergenerational mobility at the department level.[6] Children's location is defined as their department of residence in 1990, when they are between 9 and 18 years old. The IGE and the RRC vary across French departments about as much as they vary across countries. Higher levels of mobility are typically found in the West of France, and lower levels in the North and in the South. While the IGEs range from 0.28 to 0.40 in departments in Brittany (West), they range from 0.46 to 0.71 in departments in Hauts-de-France (North). The distribution of department-level RRC is tighter than that of the IGE, 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 obtained from the fitted values of the local rank-rank regression (Chetty et al., 2014). Rates of absolute upward mobility range from the 36$^{\text{th}}$ percentile in Pas-de-Calais (North) to the 54$^{\text{th}}$ in Haute-Savoie (East). Haute-Savoie, as well as the other departments that share a border with Switzerland, tend to exhibit both a high level of relative *and* absolute upward mobility. But relative and absolute mobility do not always coincide. Even though the Paris department exhibits around average intergenerational persistence levels in terms of IGE (0.53) and RRC (0.33), it stands out in terms of AUM (50). Such a discrepancy is also observed in the "empty diagonal",[7] where relative persistence tends to be quite high but absolute upward mobility quite low. This highlights the importance of using a variety of intergenerational mobility measures to characterize a country's income persistence across generations (Mazumder and Deutscher, 2021). The cross-department correlation between the IGE and RRC is only 0.65, and -0.47 with AUM.

Investigating further the potential sources of spatial heterogeneity, it appears that population density, the share of high school graduates, and income inequality are among the characteristics of departments most strongly correlated with absolute upward mobility. We also document the relationship between the Gini index of income inequality and intergenerational mobility, in the light of the "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) across countries (Corak, 2013). We find no evidence of a significant relationship between the Gini index and the IGE (or RRC) across French departments, though we find a significant negative relationship with AUM (once controlling for other characteristics), in line with what is observed in Italy (Acciari et al., 2021) and in North America (Chetty et al., 2014; Corak, 2020).

[6]There are 96 departments in metropolitan France.

[7]The empty diagonal (*diagonale du vide*) is an area of low-density population that stretches from the southwest to the northeast of France.

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 typically subject to. Section 3 outlines our two-stage estimation procedure. The data, as well as the sample and variable definitions are presented in Section 4. Section 5 reports our baseline estimates at the national level, while Section 6 assesses their robustness to various sources of bias. In Section 7, we investigate the spatial variations in intergenerational income mobility and attempt to uncover which local characteristics correlate with intergenerational mobility. Section 8 concludes.

2 Measuring Intergenerational Income Mobility

Intergenerational income dependence can be characterized using a variety of statistics, each with its advantages and downsides.[8] 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 mobility measures primarily aim to characterize the joint distribution of lifetime incomes of children and their parents with a parsimonious set of practical statistics. Specifically, we base our characterization of intergenerational persistence on the following types of statistics.

Intergenerational income elasticity (IGE). The traditional measure used to quantify intergenerational persistence is the intergenerational income elasticity. It is obtained by regressing the log lifetime income of children on the log lifetime income of their parents. The resulting elasticity has a very intuitive interpretation: an IGE of 0.4, for example, implies one would expect to earn 4% more than the mean of one's income distribution if one's parents had earned 10% more than the mean of their income distribution. A noteworthy property of this estimator is its sensitivity to variations in inequality across generations. This can be seen in the following equation, where y_p is parent log lifetime income and y_c is child log lifetime income:

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

To have a sense of the possible magnitudes of IGEs, existing estimates of father-

[8] 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), Mazumder (2016), Nybom and Stuhler (2017), Mogstad and Torsvik (2021) and Mazumder and Deutscher (2021).

son labor earnings IGEs range between 0.15 in Denmark[9] and 0.67 in Peru, though comparisons across studies are problematic due to differences in income definitions, ages at which incomes are measured and sample selection criteria (Corak, 2016). In addition to being affected by the lifecycle and attenuation biases discussed below, the empirical literature has also highlighted that IGEs are particularly sensitive to sample selection criteria, non-linearities along the parent income distribution, differences in income definitions[10], and to the treatment of negative/zero incomes (Couch and Lillard, 1998; Chetty et al., 2014; Landersø and Heckman, 2017; Pekkarinen et al., 2017).

Rank-rank correlation (RRC). While the IGE depends on the two components of the joint distribution, i.e., the marginal distributions and the copula, the correlation between the percentile rank of children and parents in their respective income distributions purely depends on the latter. Although these two estimators are suited to answer complementary questions, the rank-rank slope (or equivalently, the Spearman rank correlation) has gained in popularity lately, notably because of 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). This estimate corresponds to the slope in the regression of the percentile rank of child lifetime income on the percentile rank of parent lifetime income. A RRC of 0.4, for example, means that a 10 percentile increase in parent income rank is associated, on average, with a 4 percentile increase in child income rank. With p_p being parent percentile rank and p_c child percentile rank in their respective lifetime income distribution, it formally writes:

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

Because parents and children percentile income ranks both follow by construction a uniform distribution from 1 to 100, their standard deviations factor out and the regression slope equals the correlation coefficient. This illustrates that while the IGE is sensitive to relative inequality across generations, the RRC is not. Consequently, the greater the degree of inequality in the children generation relative to the parent generation, the greater the IGE relative to the RRC. As such, 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, implying greater income improvements from moving up ranks.

In terms of magnitude, it appears so far that there is less between-country variation in RRC than in IGE. According to our compilation of existing estimates based

[9] This estimate is likely to understate intergenerational income persistence in Denmark and be closer to 0.25, see Helsø (2021).

[10] Helsø (2021, Table 4) shows using Danish register data that depending on the income definition used the IGE could be between 0.23 and 0.35.

on household income definitions and pooling sons and daughters together, the RRC ranges from 0.197 in Sweden (Heidrich, 2017) to 0.341 in the United States (Chetty et al., 2014).[11]

Rank-rank regressions also enable computing a measure of upward mobility proposed by Chetty et al. (2014) - absolute upward mobility (AUM) -, which is defined as the expected income rank of children whose parents locate at the 25$^{\text{th}}$ percentile of the income distribution. This measure therefore combines the intercept and the slope.

Transition Matrices. To get a finer picture, one can also use so-called 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 across the parent income distribution.

2.2 Main sources of bias

All measures of intergenerational income mobility aim at characterizing the relationship between *lifetime* incomes across generations. Yet the vast majority of currently available data sources do not cover the whole lifetime of children's or parents' incomes, let alone both, leading researchers to approximate lifetime income based on shorter time spans. The literature has identified two fundamental biases inherent to this data limitation, that any intergenerational mobility estimator is, to a varying extent, subject to.

Attenuation bias. As first highlighted by Solon (1992) and Zimmerman (1992), and further documented by Mazumder (2005, 2016) and Nybom and Stuhler (2017), a direct implication of relying on a limited number of income observations to approximate parent lifetime earnings is the standard attenuation bias arising from classical measurement error. This leads to downward-biased estimates of intergenerational mobility. Considering parents' income averaged over few years as a noisy measurement of parents' lifetime income, it follows that the larger the variance of the noise, the stronger the downward pressure on the estimate. Mazumder (2016) and Nybom and Stuhler (2017) find that the attenuation bias can be very large for the IGE but affects the RRC only mildly. O'Neill et al. (2007) shows that the corner elements of the transition matrix are most sensitive to attenuation bias. The common "solution" to overcome attenuation bias is to average parent income over as many years as possible. Note that a limited number of child income observations theoretically reduces the precision of the coefficient, but does not threaten unbiasedness.

[11] See Table 2 for more details.

Lifecycle bias. The second common bias to intergenerational persistence measurement relates to the age at which child and parent incomes are observed in the lifecycle (Grawe, 2006; Haider and Solon, 2006; Nilsen et al., 2012; Nybom and Stuhler, 2016, 2017). In particular, lifecycle bias arises in the presence of heterogeneous age-income profiles, which is observed empirically as high income individuals tend to experience steeper earnings profiles than low 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 moderately so long as incomes are measured at least in the late 20s/early 30s. As for the attenuation bias, the corner elements of the transition are most sensitive to lifecycle bias. The general rule to minimize (though not eliminate fully) lifecycle bias is to measure child and parent incomes in their mid to late 30s (Grawe, 2006; Haider and Solon, 2006; Chetty et al., 2014; Nybom and Stuhler, 2017).

We conduct extensive sensitivity analyses to ensure our estimates are not affected by these biases.

3 Parental Income Prediction

The various measures of intergenerational mobility laid out in Section 2.1 cannot be estimated directly with our data since we do not observed parents' incomes. We therefore rely on the two-sample two-stage least squares (TSTSLS) strategy introduced by Björklund and Jäntti (1997).[12] This method consists in using a sample of parents drawn from the same population as the actual parents but for whom incomes are observed - so-called synthetic parents - to estimate a prediction model, which is then applied to actual parents. Appendix Figure A4 provides an illustration of the methodology.

Let Z_i denote a set of socio-economic characteristics observed for both parents and synthetic parents. Parents' and synthetic parents' log lifetime income can be expressed as

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

where $\varepsilon_{i,t}$ is the component of lifetime income not captured by the set of predictors and assumed uncorrelated with Z_i. We estimate β from equation (3) on our sample of synthetic parents, and parents' lifetime incomes are then predicted using the resulting $\hat{\beta}$ as

[12] This method has been used in the French context by Lefranc and Trannoy (2005) and Lefranc (2018) and for many other countries where child and parent incomes cannot be observed simultaneously.

$$\hat{y}_i = \hat{\beta} Z_i. \tag{4}$$

The prediction model is estimated by ordinary least squares introducing linearly demographic characteristics in 1990 (age, French nationality dummy, country of birth (6 categories), 2-digit occupation (42 cat.), education (8 cat.), 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). Synthetic parents' income is defined as average pretax wage between 35 and 45, with at least 2 income observations over this age range. The model is estimated separately on synthetic mothers and fathers, whose log income distribution is shown in Appendix Figure A5. Even though the distribution is left-skewed, Appendix Figure A6 shows that trimming the bottom (or the top) of the distribution has virtually no effect on the out-of-sample mean squared error. The prediction model is reported in Appendix Table A1. The adjusted R^2 for synthetic fathers is 0.36 and it is 0.37 for synthetic mothers.

Method validity. Despite the extensive use of the TSTSLS method for estimating the IGE, relatively little is known about the consistency of this estimator, let alone for measures of intergenerational persistence other than the IGE. Jerrim et al. (2016) perform a validation exercise using the US' Panel Study of Income Dynamics (PSID). They exploit the fact that in the PSID parents' income is observed and compare estimates obtained with TSTSLS (i.e. using in-sample predictions) to those obtained using the parents' observed incomes. When averaging parents' incomes over ages 30 to 60 (with at least 5 income observations) and using race, education, 3-digit occupation and 3-digit industry as predictors, they find an IGE of 0.644 when it should actually be 0.570. This represents an upward bias of about 13%. Thus, we remain somewhat cautious about the exact magnitude of our IGEs as they may potentially be upward biased. Jerrim et al. (2016) do not perform this exercise for the RRC or transition matrices. In Appendix B we perform our own tentative simulation exercise and find that our IGE estimates may potentially be moderately upward biased.

4 Data

We use data from the Permanent Demographic Sample, a socio-demographic panel combining several administrative data sources on individuals born on the first four days of October.[13] We refer to individuals born on one of these days as *EDP individ-*

[13] See Jugnot (2014) and INSEE (2021) for a detailed description of the dataset. All the documentation of the EDP can be found at https://utiledp.site.ined.fr/fr/variables/variables-de-l-edp/. The EDP selection criterion has progressively widened to include individuals born on the first days of January, April, and July.

uals. We use four of the EDP's administrative data sources: (i) the civil registers; (ii) the 1990 population census; (iii) the *All Employee Panel*; and (iv) the tax returns data. We match these datasets using a unique individual identifier. We describe the most relevant details for each data source below and provide additional technicalities in Appendix A.

Civil registers. The civil registers contain information from birth certificates, death certificates, adoptions, marriages and birth certificates of children of EDP individuals. They are available from 1968 onwards, and typically include information on the EDP individual's gender, date and place of birth, and her parents' date and place of birth, nationality and occupation among other variables.

1990 census. The 1990 census contains socio-demographic information about EDP individuals, as well as, though to a lesser extent, about their family and household members. Importantly, if the EDP individual is a child in 1990, it contains information about her parents' education level, occupation, and other demographic information such as nationality, age, marital status, etc.

All Employee Panel. The All Employee Panel combines two sources of data: the annual declarations of social data (*déclarations annuelles des données sociales* - DADS), which covers the private sector, and data on central government employees (*fichiers de paie des agents de l'état* - FPE), and is available from 1967 onwards (only for individuals born on an even year prior to 2001). The All Employee Panel data are reported at the worker-year level. They cover all private sector employees in metropolitan France, except those in the agricultural sectors, and public sector workers.[14] Lastly, because of increased workload due to the population censuses of 1982 and 1990, the All Employee Panel data were not compiled by INSEE in 1981, 1983 and 1990.

Tax returns data. The tax returns data is compiled using housing and income tax forms filed for incomes earned from 2010 to 2016. 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 are married or in a civil union, but also for couples who live together, an increasingly common arrangement.[15]

[14] See Appendix A for details on the evolution of the coverage of the All Employee Panel.

[15] In France, families headed by non married couples are as frequent as single-parent families (INSEE, ed, 2020).

4.1 Sample Definitions

Sample of children. Our sample of children consists of EDP individuals who are (i) born between 1972 and 1981 in metropolitan France, (ii) observed as children in the 1990 census, (iii) whose parents are neither farmers nor in a liberal profession[16], and (iv) observed at least once in the tax returns data between 35 and 45 years old.[17] The reason for restriction (i) is to ensure we observe them as children in the 1990 census[18] and have a sufficiently large number of observations for the subnational analysis. Restriction (ii) enables us to retrieve characteristics on their parents, and (iii) is due to that fact that farmers and liberal professions are not covered by the All Employee Panel[19] from which we obtain synthetic parent income. Restriction (iii) accounts for the recommendations of the literature regarding lifecycle bias. The final sample contains 65,632 children.[20]

Sample of synthetic parents. Our sample of synthetic parents is constructed as to ensure our sample of actual parents and synthetic parents come from the same overarching population of parents. It consists in 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, and (iv) have at least two wage observations between 35 and 45 years old.[21] We observe their pretax wages in the All Employee Panel. Wage data is only available for individuals born in even years prior to 2001, therefore our sample consists in individuals born in an even year. The final sample contains 31,423 synthetic parents.[22]

4.2 Variable Definitions

The variables we use in our main analysis are constructed as follows and are expressed in 2015 euros using the consumer price index published by INSEE.

Parent income. We define parent income as mean predicted pretax wage over ages 35

[16] 4.64% of EDP individuals satisfying (i) and (ii) have at least one parent who is a farmer and 2.1% have at least one parent who is in a liberal profession.

[17] 5.23% of EDP individuals satisfying (i) and (ii) are not observed in the tax returns data between 35 and 45 years old.

[18] See Appendix Figure A7 for the position in the family in the 1990 census by child birth cohort.

[19] The exact profession codes are: for farmers, 1-digit occupation 1 - *agriculteurs exploitants*; for liberal professions, 2-digit occupation 31 - *professions libérales*.

[20] See Appendix Table A2 for the number of observations at each additional restriction.

[21] In Appendix Table A3 we compare average characteristics of parents and synthetic parents. To ensure appropriate comparability of the two samples, no restriction on income observations for synthetic parents or children is applied. Average characteristics are remarkably similar for most variables, even for 2-digit occupation (Appendix Table A4), which confirms the assumption that actual and synthetic parents are random subsets of the same population.

[22] See Appendix Table A5 for the number of observations at each additional restriction.

to 45. This income is not observed but predicted according to the methodology described in Section 3. We use our sample of synthetic parents for whom pretax wage is observed and regress pretax wage averaged over ages 35-45 on demographic characteristics in 1990 (age, French nationality dummy, country of birth (6 categories), 2-digit occupation (42 cat.), education (8 cat.), 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), separately for synthetic fathers and mothers. Averaging synthetic parent incomes over ages 35-45 is done to reduce the potential for lifecycle and attenuation bias. We then predict actual fathers' and mothers' incomes, and compute mean income at the household level (regardless of marital status) by taking the average incomes of the father and the mother if the child is observed with both parents in 1990,[23] and income of the only parent otherwise. We refer to this income definition as parent household wage. We also report results when using only father predicted income, which we refer to as father wage.

Child income. Our main measure of child income is total pretax household income which we call household income. Specifically, total pretax household income is equal to the sum of labor earnings (wages + self-employment income), taxable and predicted non-taxable capital income[24], unemployment insurance, retirement and alimony. 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. The reason for this exclusion is to ensure a more appropriate comparison with the way parent income is defined. All incomes are measured before taxes but after the deduction of employer- and employee-level payroll taxes. A household is defined as individuals living in the same dwelling, identified by the housing tax declaration, and is therefore independent of marital status just as for parents.

To mitigate the potential for lifecycle bias, we average child income over 2010-16 only for incomes declared when the individual is between 35 and 45 years old.[25] We then divide child household income by the number of household adults based on the family structure (1 for single-headed households, and 2 for couples and complex households). We also report results using the following alternative child income definitions: (i) individual income, which we define as the sum of all individual-level

[23] The parents observed in the same household as the child in 1990 do not necessarily correspond to the biological parents. Since we are interested in the relationship between the economic environment in which the child grew up and the child's own economic outcomes, the biological link is not relevant.

[24] Financial incomes not subject to any tax reporting are predicted by 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, ...).

[25] Therefore, the 1981 birth cohort will only have at most one income observation, that for 2016 when they are 35; the 1980 cohort will have at most two income observations (2015 and 2016 when they are 35 and 36), etc.

incomes: labor earnings (wages + self-employment income), unemployment benefits, retirement and alimony; and (ii) wages, which is only equal to wage earnings.

Income definition discussion. Our parent and child incomes definitions are not identical as they 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 in which individuals grew up in, and (ii) enable the analysis of daughters, whose labor incomes alone may be a poor proxy for their economic outcomes. The extent to which this may matter is unclear since existing studies define incomes in the same way for parents and for children (as the income variables come from the same dataset as opposed to our case).

Percentile ranks. For analyses based on individuals' ranks, we proceed as follows. Children are ranked within their birth cohort, and parents are ranked relative to other parents with children in the same birth cohort. Children with negative or zero incomes are assigned a rank equal to the ceiling of the percentage of such earners in their birth cohort divided by 2.[26] No parent has negative or zero parent income.

4.3 Descriptive Statistics

Table 1 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. Unsurprisingly synthetic fathers' average pretax wages are higher than the average for synthetic mothers, and displays greater variance. As for children, household income, on average, is greater than either individual income measure.

[26] For example, if there are 3.65% of children with negative or zero incomes, they are assigned a rank of $\lceil 3.65/2 \rceil = 2$. Depending on the child income definition the percentage of children with negative or zero incomes varies. At most they represent 8% of the child sample, see Appendix Table A6 for the exact figures.

Table 1: Descriptive Statistics

	N	Missing (%)	Mean	Std. Dev.	25th pctile	Median	75th 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	11.64%						
Fraction female among single parents	88.44%						
Father age at child's birth	65,632	0.06	28.7	5.77	25	28	32
Mother age at child's birth	65,632	0.02	25.91	5.01	22	25	29
Father age in 1990	65,632	0.1	41.97	6.61	38	41	45
Mother age in 1990	65,632	0.01	39.41	5.81	35	39	43
Children							
Household income (average 2010-16)	65,632	0.01	25,655	19,402	17,049	22,772	30,173
Individual income (average 2010-16)	65,632	0	20,410	17,855	10,012	19,227	26,662
Labor income (average 2010-16)	65,632	0	22,726	19,005	13,951	20,593	27,865
Fraction female	49.59%						

Notes: See Sections 4.1 and 4.2 for details on sample construction and income definitions.

5 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 average predicted pretax wage at ages 35-45 and child income as average pretax household income in 2010-2016 only when aged between 35-45. We include child birth cohort fixed effects in the log-log and rank-rank regressions.[27]

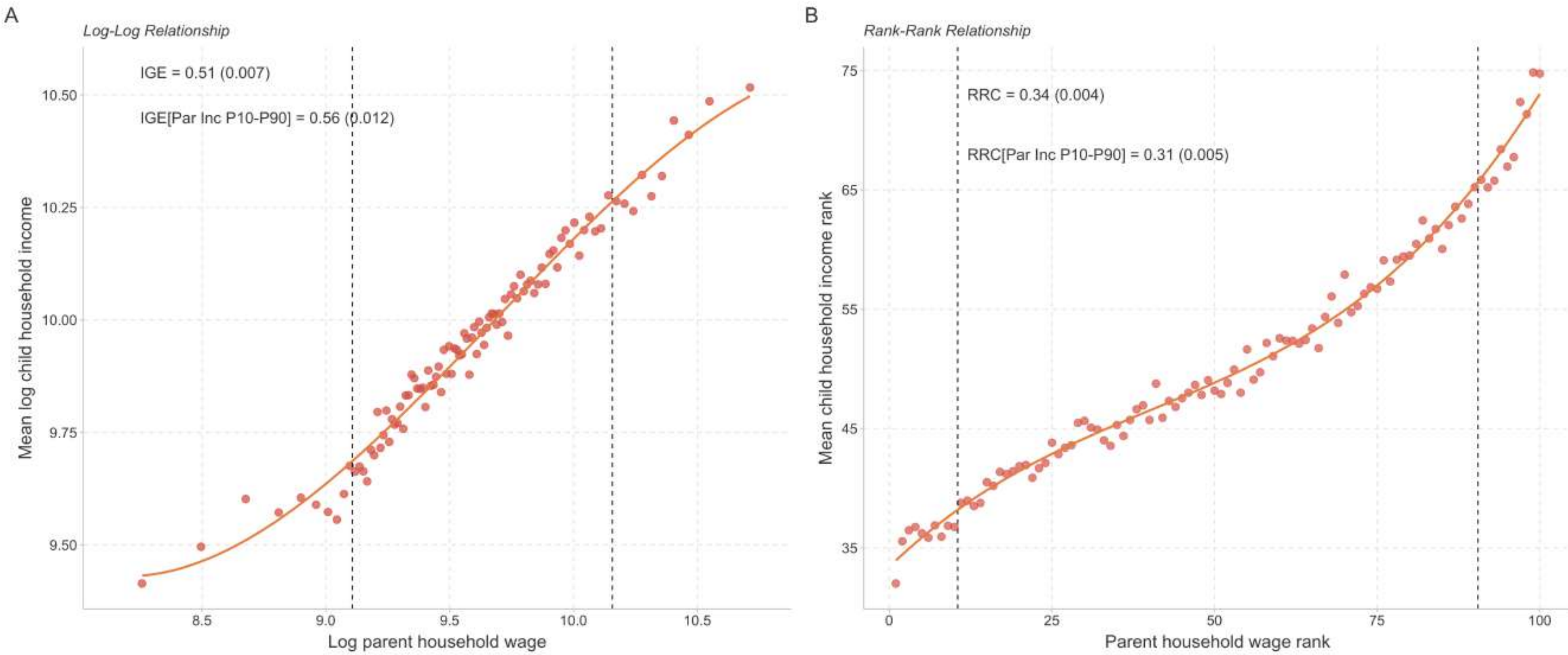
5.1 Intergenerational Income Elasticity (IGE)

Figure 1 panel A displays the conditional expectation of log child income with respect to log parent income. Children with negative or zero incomes are excluded.[28] The log-log CEF is pretty linear throughout parent income in the middle 80%, with some mild non-linearities at the tails. This S-shaped relationship is in line with the evidence from the United States (e.g. Chetty et al. (2014)) or Denmark (e.g. Helsø (2021)) where non-linearities are even more pronounced. These non-linearities imply that (i) the elasticity is not constant throughout the parent income distribution, with smaller magnitudes at the tails, and (ii) it is somewhat sensitive to the inclusion or exclusion of parents at the tails of their income distribution.

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

[28]This is of minor importance when defining child income as household income as such cases are exceedingly rare. We come back to this restriction in Section 6.3.

Figure 1: Conditional Expectation Functions for Log-Log and Rank-Rank Relationships



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). 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), divided by the number of household adults. 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 demographic characteristics in 1990 (age, French nationality dummy, country of birth (6 categories), 2-digit occupation (42 cat.), education (8 cat.), 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), and estimated on a sample of synthetic parents whose average wages 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 (.07% of the sample) and we bin parent incomes into 100 equal-sized (centile) 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. Children with negative or zero incomes are assigned a rank equal to the ceiling of the percentage of such cases in their cohort divided by 2. 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 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 on the full sample and for parents between the 10^{th} and 90^{th} percentiles. The orange line is a 3^{rd} order polynomial fit through the conditional expectations.

Appendix Figure A8 shows our baseline estimates of the intergenerational income elasticity for various child and parent income measures. Our baseline IGE estimate is 0.515, meaning that, on average, a 10% increase in parent income is associated with a 5.15%[29] increase in child income. Intergenerational persistence estimates are larger for household income than for individual income or wage, which may be related to assortative mating. But it is very similar to that obtained when defining parent income as father wage (as has been historically the case in the literature), despite the fact that by construction, estimates based on father wage exclude children only observed with their mother in the 1990 census (6.19% of observations). The IGE is significantly lower for sons (0.467) than for daughters (0.563). 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.

Our estimates are relatively similar to existing ones for France despite significant differences in methodology and data (Appendix Table A7). Our father-son wage IGE (which is the one for which an estimate exists for France) is 0.439, very close to the 0.400-0.438 range found in Lefranc and Trannoy (2005, Table I, Panel A, cols. (1)-(4), p.65) but quite a bit lower than the 0.577 found in Lefranc (2018, Table 2, 1971-75, col. (2), p.823). A likely reason for the difference between our estimate and that of Lefranc (2018) is that he uses only education as a predictor of parent income. As has been shown in the literature and as we show in Section 6.2, using only education as a predictor leads to overinflated estimates of the IGE. In fact, our estimate when using only education is 0.661. Our father-daughter estimates are not in line with those found in Lefranc and Trannoy (2005). Indeed, we estimate that the father-daughter wage IGE is 0.513, while it is 0.298-0.331 in Lefranc and Trannoy (2005, Table I, Panel B, cols. (1)-(4), p.65). Understanding the reason for these differing results is tricky due to important methodological differences as well as differences in the data sources used.

5.2 Rank-Rank Correlation (RRC)

We present estimates of the rank-rank correlation (RRC). 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. Children with negative or zero incomes are assigned a rank equal to the ceiling of the percentage of such cases in their cohort divided by 2. Figure 1 panel B plots the conditional expectation of child income rank with respect to parent income rank. This conditional expectation function appears relatively linear, with slight non-linearities at the tails as observed in many countries (Chetty et al., 2014; Bratberg et al., 2017; Helsø, 2021). As expected, it is upward-sloping implying that, on average, children born to parents with incomes higher up in the

[29] The exact expected change is actually equal to $(1.1^{0.515} - 1) \times 100 \approx 5.03\%$.

income distribution also tend to end up themselves higher up in their own income distribution.

Appendix Figure A9 shows our baseline estimates of the rank-rank correlation for various child and parent income measures. Our baseline estimate of the rank-rank correlation is 0.337, meaning that a 10 percentile increase in parent income rank is associated, on average, with a 3.37 percentile increase in child income rank. The estimates are slightly higher for daughters (0.351) than for sons (0.324), and are also slightly higher when defining parent income as household wage rather than as father wage. The estimates are significantly lower when defining child income as individual income and even lower when using wage, a pattern observed in other countries (Chetty et al., 2014; Deutscher and Mazumder, 2020; Landersø and Heckman, 2017), which could be driven by assortative mating as mentioned for the IGE.

To the best of our knowledge, this is the first time the RRC is estimated for France. In Table 2 we compare RRC estimates for countries for which estimates exist. To enable comparability we only keep studies which pool sons and daughters together and define parent income as the sum/average of father and mother income. For most countries, child income is defined at the household or family level except in Chuard-Keller and Grassi (2021), Heidrich (2017) and Acciari et al. (2021) where it is at the individual level. For both parents and children, all the studies compiled use a comprehensive income definition and not simply wage as was the case for early studies in the intergenerational mobility literature. 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 stability of the RRC to specification variations and common data limitations (e.g. only observing child incomes at relatively early ages) (Nybom and Stuhler, 2017). This international comparison suggests (i) there is less variation across countries in the rank-rank slope than with respect to the intergenerational elasticity and (ii) France exhibits strong persistence across generations in international comparison, given that our RRC estimate is greater than that for all countries with available RRC estimates except the United States.

Table 2: Rank-Rank Correlation in International Comparison

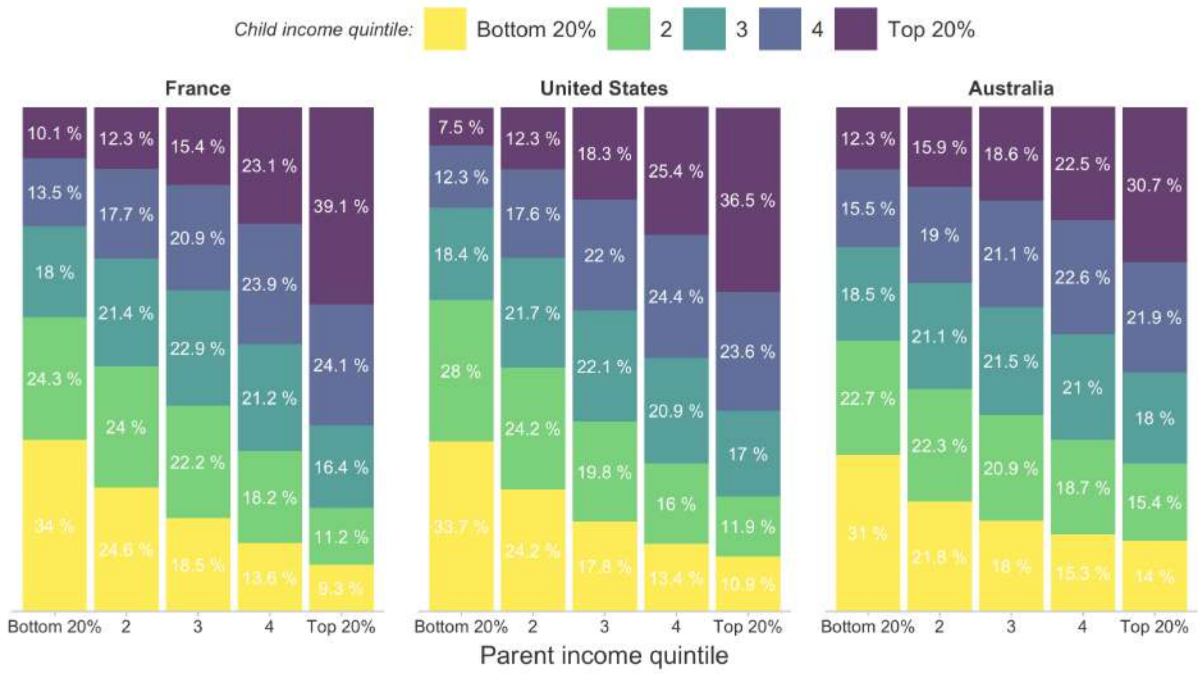
Country	RRC ↓	# obs.	Data	Income Definition	Child Cohort	Child Age or Year at Income Measurement	Parent Age or Year at Income Measurement	Source
Switzerland	0.14	923,262	Social Security Earnings Records	Average total pretax *individual* income	1967-1984	30-33	when child between 15-20	Chuard-Keller and Grassi (2021, Figure 1)
Sweden	0.197	778,484	SIMSAM database[1]	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[2]	1978-80	Bratberg et al. (2017, Table 3)
Norway	0.223	324,870	Full population admin. data	Average pretax *family* earnings	1957-1964	1996-2006	1978-80	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	≈ 410,000	Danish register data	Average total pretax *family* income	1973-1979	2010-2012	when child between 7-15	Landersø and Heckman (2017, Table A17)
Denmark	0.257	205,625	Full populations admin. data	Average total pretax *family* income	1973-1975	2010-2012	when child between 7-15	Eriksen (2018, Table 3.2)
Italy	0.30[3]	1,719,483	Electronic database of Personal Income returns	Average total pretax *individual* income	1979-1983	2016-18	1998-2000	Acciari et al. (2021, p.28)
France	0.337	64,572	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	1980-82	1996-2000	2011-2012	Chetty et al. (2014, Table 1)
United States	0.395	6,414	NLSY79	Average total pretax *family* income (self-reported)	1957-1964	1996-2008[2]	1978-1980	Bratberg et al. (2017, Table 3)

Notes:
[1] Swedish Initiative for Research on Microdata in the Social and Medical Sciences.
[2] Only even years.
[3] This estimate corresponds to the one when adjusting for lifecycle bias, omission of taxpayers and tax evasion as reported on p.28. The baseline RRC estimate reported in Table 3 is 0.22.

5.3 Transition Matrices

The last measure of intergenerational income persistence we estimate is a quintile-by-quintile transition matrix. Using the same parent and child ranking method as for the rank-rank correlation, we can obtain estimates of the conditional probability of being in a quintile as an adult given a parental income quintile. Figure 2 presents our baseline estimates for the transition matrix for France, comparing it 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.

Figure 2: Baseline Quintile Transition Matrix for Different Child Income Definitions



Notes: The first panel of this figure presents our baseline intergenerational transition matrix estimates for our baseline parent and child income definitions. See Figure 1'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 estimations 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).

We find that 10.1% of children born to parents in the bottom 20% reach the top 20% when they are adults. This share is 7.5% in the United States and 12.3% in Australia. Conversely, 34% remain in the bottom 20% of the income distribution. Regarding children born to the top 20%, 39.1% remain at the top, while only 9.3% move down to the bottom of the income distribution, much less than in Australia (14%). It is useful to analyze these figures in light of a society where an individual's income is completely independent of parent income. In such a society, the probability of being in any quintile given a parent quintile would by definition be 20%.

Changing the child income definition does not appear to affect upward mobility though it does seem to influence downward mobility markedly (Appendix Figure A10). When defined as individual income, 13% of children in the top 20% are in the bottom 20% as adults. This figure increases to 15.3% when defined as wage. This very likely reflects labor supply choices conditional on spousal income, rather than actual downward mobility. This highlights the importance once again of using household (or family) level measures of income when analyzing intergenerational income mobility.

In Table 3 we compare some of our conditional probabilities with those found for other developed countries. The same picture as for the IGE and the RRC emerges: in France the persistence in incomes across generations is particularly strong. While it does better than the United States when it comes to upward mobility (10.1% vs. 7.5%), a point we discuss in Section 5.4, it fares significantly worse than countries such as Canada (11.4%), Australia (12.3%) or Sweden (15.7%). It also displays the strongest persistence at the bottom and at the top of the income distribution of all the comparison countries.

Table 3: Transition Matrix in International Comparison

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. (2021, see footnotes 1 and 2)
France	10.1%	34%	39.1%	
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)
Australia	12.3%	31%	30.7%	Deutscher and Mazumder (2020, Table 3)
Sweden[3]	15.7%	26.3%	34.5%	Heidrich (2017, Figure 10, Appendix B)

Notes: See Table 2 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 once controlling 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 row 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.

We analyze persistence at the top of the parent income distribution in more detail in Appendix Figure A11. Specifically 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 using detailed percentile-by-percentile estimates provided in the online appendix of Chetty et al. (2014). We estimate the likelihood of remaining in the top 10% to be about 29% 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.

5.4 Discussion of Baseline Results

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 of the cross-country comparison literature (e.g. see Corak (2016)), but also for the RRC and in terms of transition matrices.

From a more methodological point of view, to the best of our knowledge this paper is only the second one to estimate the RRC using the two-stage procedure, the other being for Italy (Barbieri et al., 2020), and the first to do so for transition matrices. As we discuss in detail throughout Section 6, it appears that the RRC is very stable to numerous potential sample selection choices and other biases that might affect the two-step estimation, especially in comparison to the IGE. Estimating transition matrices with imputed parent incomes requires having a sufficiently large number of common characteristics between parents and synthetic parents in order to estimate income positions as accurately as possible. We believe therefore that RRCs could most likely be relatively precisely estimated for many more countries than is currently the case, and in particular in those where no linked child-parent dataset is available.

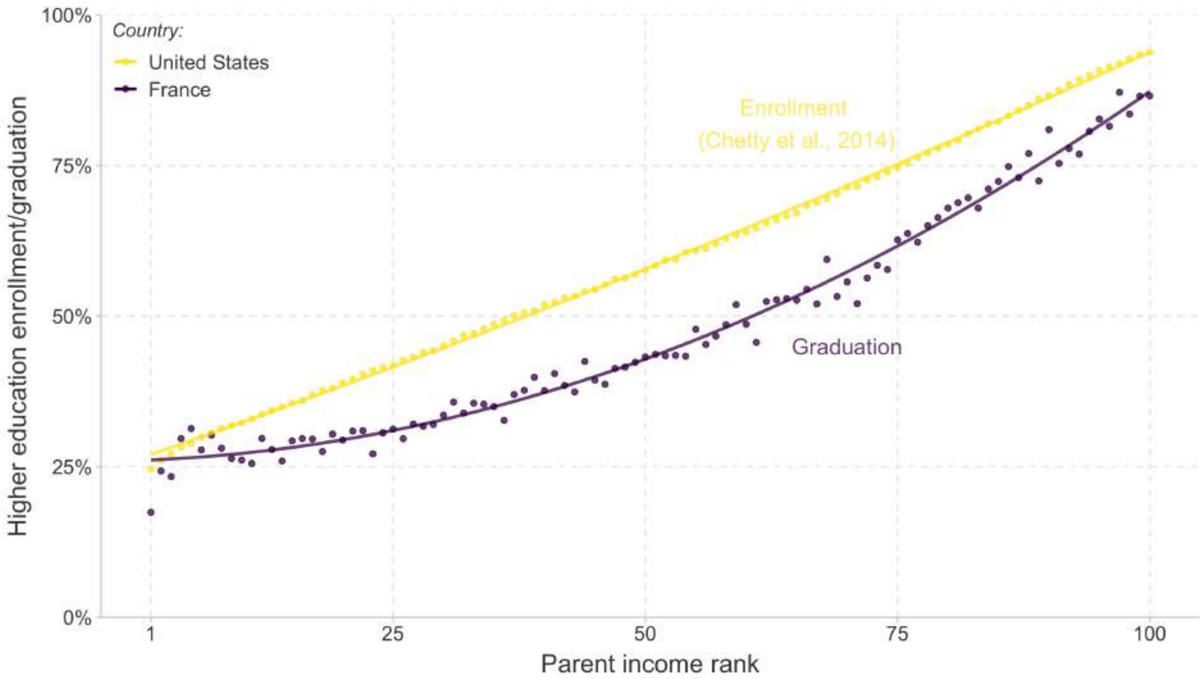
Lastly, though the IGE and RRC estimates in France are very similar to those obtained for the United States, the two countries differ in the probability of reaching the top 20% conditional on having parents in the bottom 20%. A potential explanation could be differences in higher education trajectories along the parent income distribution in France and in the United States. Indeed, the large differences in tuition fees between the two countries make access to higher education a theoretically plausible source of divergence in upward mobility at the bottom of the parent income distribution.

Using the yearly census surveys available since 2004 in the EDP, we observe children's last obtained diploma when they are between 23 and 45.[30] Figure 3 compares

[30]We observe this information for 86.29% of the sample; see Appendix C for details. The share of

higher education graduation rates in France with enrollment rates in the United States[31] (computed by Chetty et al. (2014)) by parent percentile income rank. Graduation rates in France are lower than enrollment rates in the United States, very likely due to college dropouts. 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.[32] 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 a higher education degree), which would lead to a concave relationship.

Figure 3: Graduation From Higher Education by Parent Income



Notes: This figure presents higher education graduation in France vs. enrollment rates in the United States (taken from (Chetty et al., 2014)) by parent income rank. See Figure 1's notes for details on data, sample and income definitions, and Appendix C for additional details on the specific construction of the higher education graduation variable.

The higher access to the top income quintile for children born to parents in the bottom income quintile in France than in the United States, and differences in intergenerational mobility patterns between countries in general, are certainly due to a myriad of intertwined factors. The evidence put forward in Figure 3 is suggestive of the fact that facilitated access to higher education in France for children with the most modest

missing values is pretty well uniformly distributed along the parent income rank distribution.

[31]Specifically, enrollment is defined as attending college at least at some point between ages 18-21.

[32]Appendix Figure A12 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.

family backgrounds may be part of the equation, be it or not in a causal or direct way.

6 Robustness of Baseline Results

6.1 Lifecycle and Attenuation Bias

As discussed in Section 2.2, the existing empirical literature has highlighted two statistical biases that may affect the baseline estimates above: lifecycle and attenuation bias. The former relates to heterogeneous lifecycle earnings profiles among parents and children. We therefore assess how our estimates vary with the age at which child and parent incomes are measured. The latter refers to classical measurement error in the right-hand side variable, parent income.

6.1.1 Lifecycle Bias

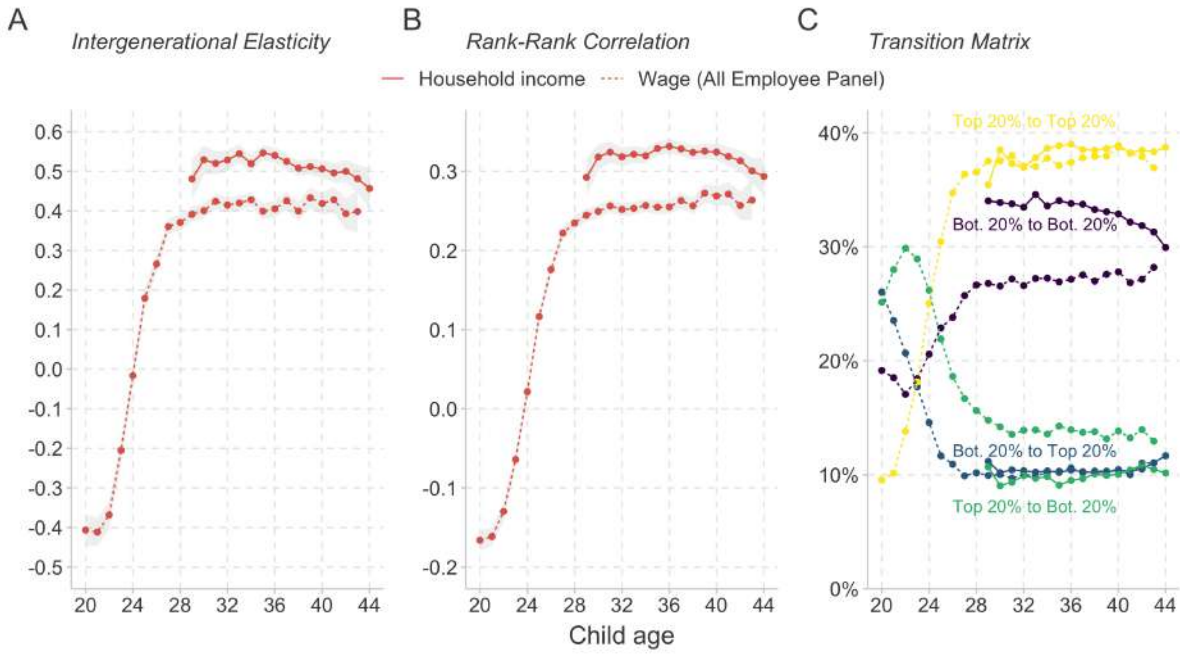
Child lifecycle bias. Figure 4 presents our estimates of intergenerational income mobility when varying the age at which child income is measured. In addition to household income from the tax returns data, we exploit the longer time series wage data provided by the All Employee Panel. Each point represents the estimate of the measure of intergenerational income mobility when measuring child income at a given age.[33] 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 4 panels A and B is that the estimated IGE and RRC increase sharply when child incomes are measured early in the lifecycle and stabilize roughly when child income is measured around 30 years old. The wage IGE (RRC) measured at age 25 is equal to 0.179 (0.116) while it is 0.399 (0.255) at age 35, more than a doubling in magnitude. For household income there appears to be a slight decline in the estimates when children are in their forties. This appears to mostly reflect changes in the underlying cohort sample rather than a real decrease in the estimate.

The results for the transition matrix in Figure 4 panel C suggest our baseline estimates are quite close to the estimates obtained when child income is measured at any age between 29 and 44, except for persistence in the bottom 20% measure which

[33] 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). For the tax returns data, at age 29 the estimate is based only on the 1981 cohort (with incomes measured in 2010), at age 35 the estimate is based on the 1975-1981 cohorts (with incomes measured in 2010 to 2016 respectively), while at age 44 the estimates are based only on the 1972 cohort (with incomes measured in 2016). Therefore the presented estimates cannot be directly compared across ages nor with our baseline results nor between income measures since as we discussed previously different income measures yield different baseline estimates. But the lifecycle bias is unlikely to be overestimated by a secular trend as Lefranc (2018) shows that the IGE has increased in France over the second half of the twentieth century.

Figure 4: Child Lifecycle Bias



Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figures 1 and 2 to changes in the age at which child income is measured. The sample of children varies across ages due to household income being observed only between 2010 and 2016, and wage between 1967 and 2015 (only for individuals born in even years before 2001). For example, child household income measured at age 29 corresponds to the 2010 income of children born in 1981. Shaded areas represent the 95% confidence intervals. See Figure 1's notes for details on data, sample and income definitions.

declines after age 36. The estimates using the All Employee Panel confirm that when measuring child incomes too early in the lifetime, the secondary diagonal elements of the transition matrix (remaining in the same income quintile as one's parents) would be severely underestimated while of the "big transitions" (from bottom to top and from top to bottom) would be severely overestimated.

As a further check, and to overcome the issue related to changes in underlying sample of children, we reproduce the All Employee Panel estimates keeping the sample of children constant. To do so we restrict our sample to children born in 1972 and 1974[34] for whom wages are observed every year between 25 and 43 years old and 25 and 41 years old respectively. Appendix Figure A13 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. Again, we find that measuring child income prior to age 30 or perhaps even slightly later risks seriously underestimating the IGE (panel

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

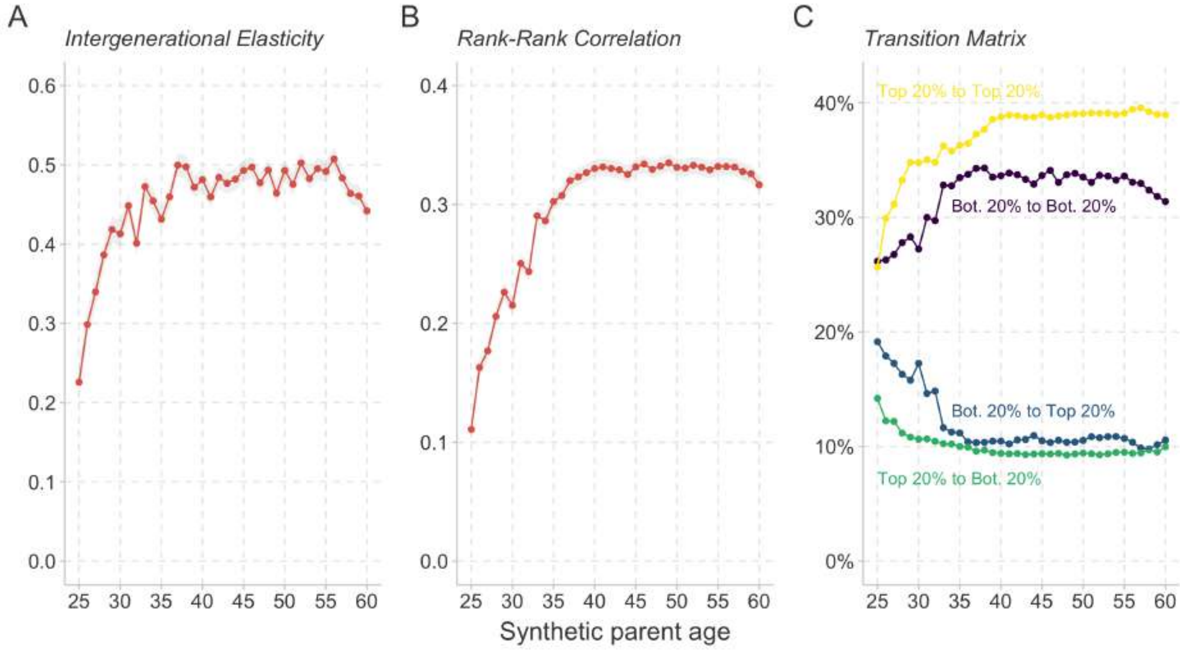
A). For the 1972 cohort, the IGE at age 25 is slightly negative (as in Nybom and Stuhler (2017)) and only 0.153 at age 26 while it is 0.274 at 35. The IGE appears to be more volatile, suggesting that averaging child income over several years is likely to result in more stable estimates. Measuring the RRC prior to age 28 markedly underestimates its magnitude (panel B). For the 1972 cohort, the RRC estimate when child income is measured at age 25 is around 0.034 while it is around 0.260 at age 35. For the transition matrices, the various cells appear to stabilize in the early/mid thirties (panel C).

Overall, we do not find persuasive evidence that the IGE or the RRC varies importantly with the age at which child income is measured so long as it is measured at least in their early thirties. This does not imply that there is no remaining lifecycle bias, as highlighted in Nybom and Stuhler (2016), it is suggestive that our baseline results do not appear to measure child incomes too early in their lifecycle. Our results are roughly in line with those found by Nybom and Stuhler (2017), though they differ regarding the IGE. Relative to Mazumder (2016) we find a significantly smaller impact of child age on the IGE after age 30.

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. Specifically, we run the first stage regressions defining synthetic parent income at age a for a between 25 and 60 years old. Figure 5 shows how our estimates of intergenerational mobility vary with the age at which parent income is predicted. The relationship between the IGE (panel A) and RRC (panel B), and age at which parent income is measured is concave, strongly increasing between 25 and the late thirties and then stabilizing until the mid to late fifties. Indeed when predicted parent income is based on synthetic parent income at age 25, the IGE for the entire sample is 0.226 (RCC = 0.111) while it is 0.481 (RRC = 0.330) at age 40. 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 (with at least 2 income observations) is either too early or too late in the lifecycle. Moreover, it is expected that the estimates based on single income years might be smaller than those based on incomes averaged over several years due to attenuation bias.

This mismeasurement of parent income also affects estimates of transition probabilities (Figure 5 panel C). Relative to our baseline results, measuring parent income at age 25 underestimates the likelihood of remaining at the bottom 20% or top 20% and overestimates the probability of moving upwards (bottom 20% to top 20%) or downwards (top 20% to bottom 20%). The estimates stabilize once again when parent income is measured after age 35 and does not appear that our baseline estimates are affected by lifecycle bias in parent income.

Figure 5: Parent Lifecycle Bias



Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figures 1 and 2 to changes in the age at which parent income is predicted. Shaded areas represent the 95% confidence intervals. See Figure 1's notes for details on data, sample and income definitions.

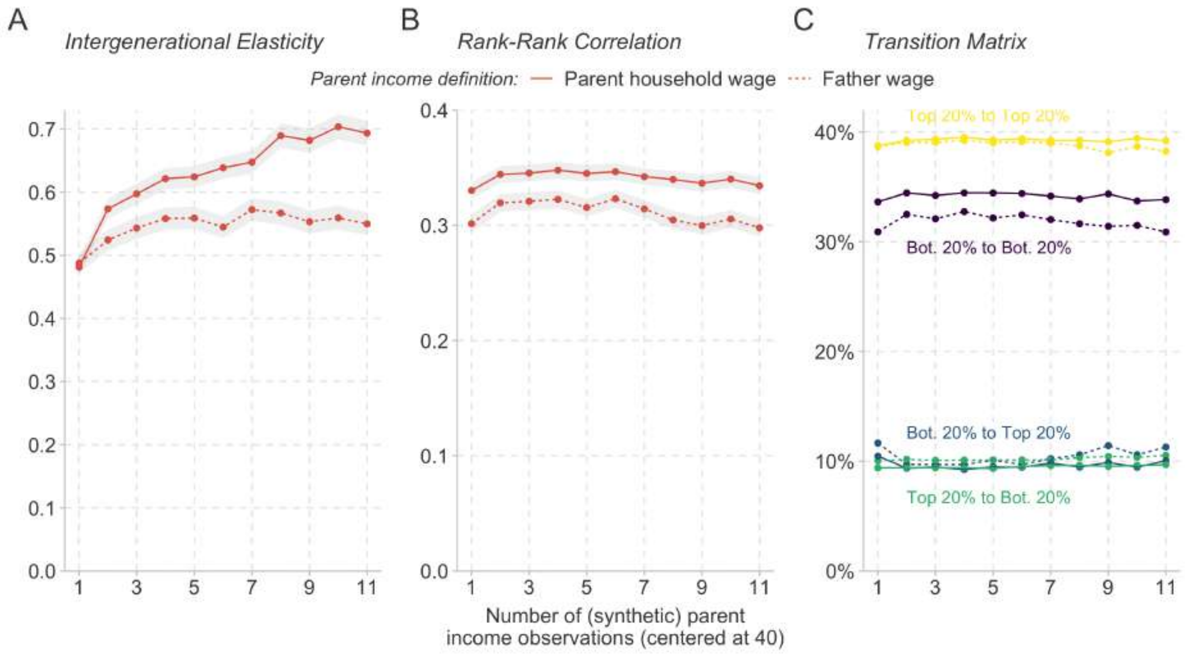
6.1.2 Attenuation Bias

We assess 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.[35]

Figure 6 plots estimates of our persistence measures varying the number of synthetic parent income observations used in the first-stage prediction regression from 1 to up 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 the first-stage prediction is computed varies for each estimate depending on how many synthetic parents had incomes observed in the required age range. We report results both for parent household wage and father wage.

[35]We also check in Appendix Figure A15 the sensitivity of intergenerational mobility to the number of child income observations and confirm that it plays only a very minor role.

Figure 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. That is, 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. Shaded areas represent the 95% confidence interval. See Figure 1's notes for details on data, sample and income definitions.

Figure 6 suggests that attenuation bias might affect our parent household wage IGE (panel A) but not our other estimates of intergenerational mobility. Indeed when defining parent income at the household level, the IGE increases from just below 0.5 when using only one income observation to around 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.[36] 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 parental income is actually observed (Mazumder, 2005). Indeed, since we are already predicting parental income based on observable characteristics, and thus in a sense reducing year-on-year income volatility, averaging

[36]How one interprets the results based on parent household wage depends one 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 might bias 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.

over more years does not affect the estimate much.

The rank-based measures, whether the RRC (Figure 6 panel B) or the transition matrix cells (Figure 6 panel C), are remarkably unaltered by increasing the number of income observations over which synthetic parent income is averaged. 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 be sufficient to obtain reliable estimates.

Additionally, we check whether the lack of change in intergenerational mobility measures with the number of (synthetic) parent income observations could be due to the fact that the sample of synthetic parents varies throughout. We replicate Figure 6 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 prediction (centered around 40 years old again). To do so, we impute wages in 1981, 1983 and 1990, for which the data is not available,[37] 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 A16 displays the results from this sensitivity analysis. 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. The increase in the parent household wage IGE is much less marked, increasing from 0.618 when using one income observation to 0.693 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 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).

So far we have only assessed the robustness of our baseline estimates to the two

[37] As explained in Section 4, the 1982 and 1990 population censuses generated an extra workload which prevented INSEE from compiling the All Employee Panel data for these years.

main biases highlighted by the literature: lifecycle and attenuation bias.[38] However, our estimates may be sensitive to other aspects: (i) the choice of predictors and estimation method used in the first stage, and (ii) the inclusion/exclusion of tails of the child and parent income distribution.

6.2 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 valid. We make sure that this first stage does not spuriously drive the results in one way or another by evaluating its sensitivity to relaxing parametric assumptions and varying both the set of instruments and sample restrictions.

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 parental 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.[39]

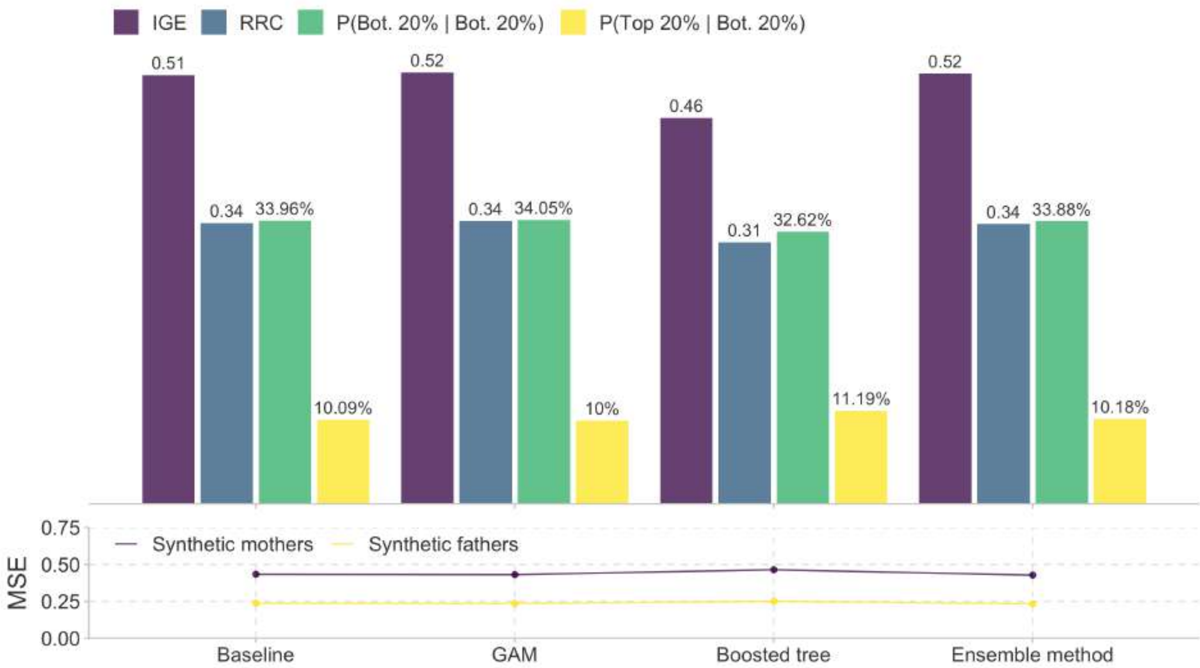
Figure 7 compares the intergenerational mobility estimates and out-of-sample MSE resulting from these three methods using our baseline child and parent income definitions. We do not observe significant differences in MSE between the different prediction methods. The resulting mobility estimates are virtually 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 addi-

[38]Though in the main body of the article we assess child and parent lifecycle bias separately, in Appendix Figure A14 we study how our measures of intergenerational persistence vary with the age at which child and (synthetic) parent income is measured jointly.

[39]See Charpentier et al. (2019) for illustrative applications and additional details on these methods.

tive OLS specification.

Figure 7: 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 1's notes for details on data, sample and income definitions.

The other 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 A17 documents the sensitivity of IGE and RRC estimates to the set of predictors used in the first-stage estimation. We do not assess the sensitivity of the transition matrices because for those measures, the accuracy of the prediction matters more and therefore simple prediction models will necessarily be inadequate. We estimate the IGE and RRC for adding each of the following predictors sequentially (all measured in 1990): education (8 cat.), 2-digit occupation (42 categories), 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). 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 R^2 and root mean squared error

(RMSE), computed as the average from 5-fold cross-validation.

We find that the IGE is 0.66 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 city-level characteristics has no effect on the estimates. Indeed, as can be seen from the adjusted 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 0.32 with only education and 0.33 with all variables. This appears once more to be a strength of the RRC in the TSTSLS context.

6.3 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 reporting 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.

The first issue is particularly salient for the estimation of the intergenerational income elasticity due to the impossibility of taking the log of zero.[40] Many authors simply disregard 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 A18 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 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 0.79 when they are recoded to 1 and 0.54 when recoded to 1,000. The RRC is entirely insensitive to such recoding as ranks are not altered by it.

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%.

[40] 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.

Figure 8 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 0.515 to 0.411 while the RRC only decreases from 0.337 to 0.322. 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.

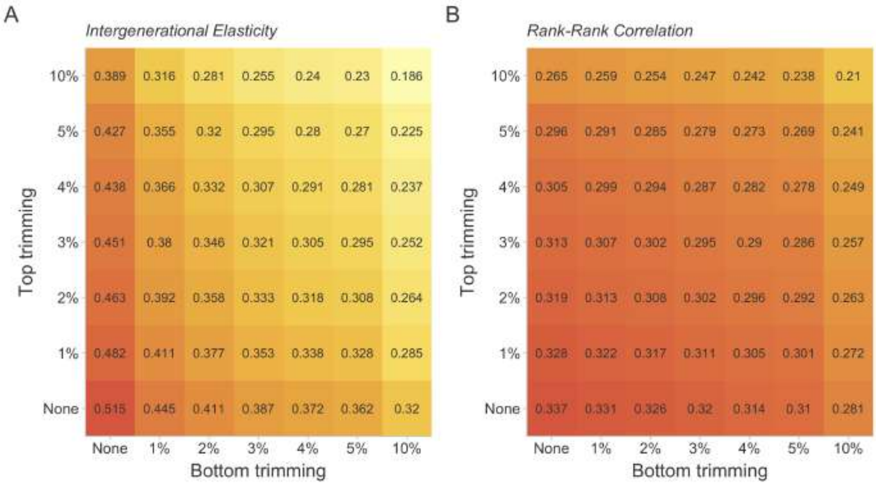
Second, the IGE is quite strongly influenced by minor trimming in the first-stage prediction sample. For example, excluding the bottom and top 2% of synthetic parent incomes leads to an IGE of 0.584. Such exclusions are not uncommon in the literature though their relevance is unclear.[41] Meanwhile the 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 0.333 (from 0.337). This appears to be an additional benefit of estimating the RRC when using with the TSTSLS method.

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.

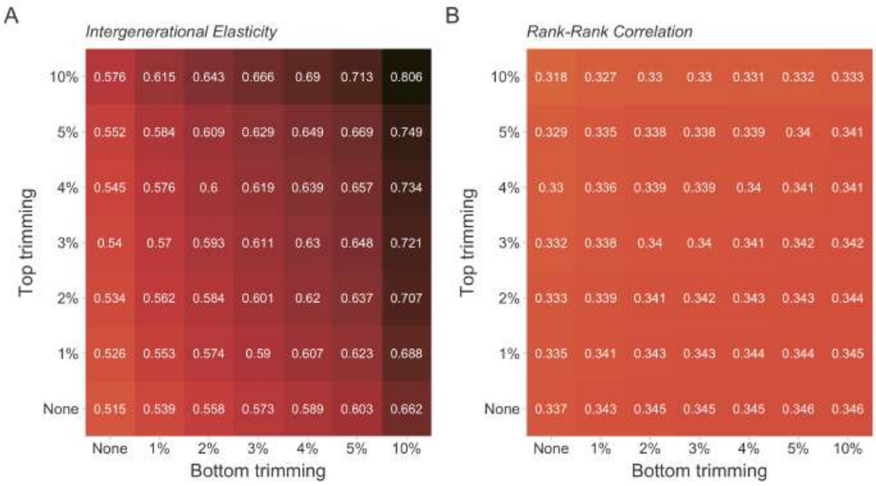
[41] For example, Barbieri et al. (2020) exclude the top and bottom 1% of their sons and synthetic fathers' incomes.

Figure 8: Sensitivity to Child and Parent Income Distributions Trimming

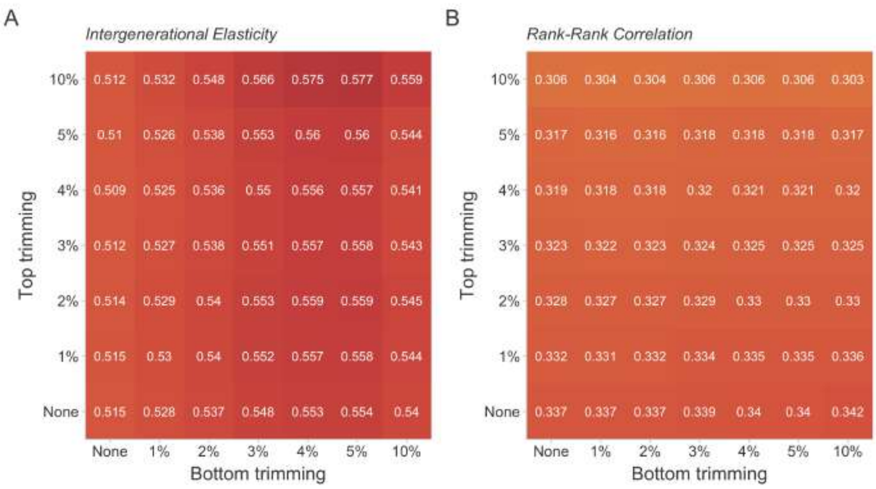
(a) Child Income Trimming



(b) First-Stage Synthetic Parent Income Trimming



(c) Second-Stage Parent Income 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 1's notes for details on data, sample and income definitions.

7 Spatial Analysis

7.1 Heterogeneity across departments

A first step in the understanding of the sources of intergenerational mobility in France is the investigation of where persistence is the highest. We study the geographic variations of intergenerational mobility at the department level. These geographical units divide metropolitan France into 95 territories.[42] Departments have the advantage of covering the whole of metropolitan France, and their borders have not change over the study period. In addition, considering a finer geographic decomposition such as commuting zones would imply the omission of a sizable amount of the population due to insufficient data, as in our sample only 25% of commuting zones have more than 200 observations.

Individuals are assigned to the department they lived in in 1990, when they were between 9 and 18 years old. To ensure our estimates are sufficiently reliable, we focus on the 84 departments with over 200 observations.[43] For each department, we estimate the IGE and the RRC in the same way as the baseline results. Importantly, individuals are still ranked within the national income distribution. This implies that, in terms of percentile ranks, moving up the national income ladder by geographically moving to a higher income department is considered as intergenerational mobility even if relative to others in her new department the individual ends up in the same percentile rank their parents fell into within her childhood department.

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). 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 from the following OLS regression:

$$p_{c,d} = \alpha_d + RRC_d \times p_{p,d} + \varepsilon_d \tag{5}$$

The expected income rank for individuals whose parents locate at the 25$^{\text{th}}$, $\mathbb{E}[p_{c,d} \mid p_{p,d} = 25]$ then writes:

$$\mathbb{E}[p_{c,d} \mid p_{p,d} = 25] = \hat{\alpha_d} + R\hat{R}C_d \times 25 \tag{6}$$

Appendix Figure A20 graphically illustrates how this intergenerational mobility measure is computed for the *Nord* department, the most populated one in 1990. Rank and log income conditional expectation functions for the most populous departments

[42] For practical reasons, we treat Northern and Southern Corsica as a single department.

[43] The number of observations per department is reported in Appendix Table A9.

are available in Appendix Figures A21 and A22. We favor absolute upward mobility over specific cells of the transition matrix because of the size of our sample. 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.

Figure 9 depicts the spatial variations in intergenerational mobility as captured by the three estimators mentioned above.[44] It reveals substantial variations across departments. The distribution of department-level RRCs ranges from 0.21 to 0.45 and is tighter than that of IGEs, which range from 0.25 to 0.79. Yet it varies across departments just as much as it varies across countries.[45] The range of our estimates of absolute upward mobility, from rank 36 to rank 54, is slightly more narrow than that observed in Italy using a comparable geographic unit: from 37 to 63 (Acciari et al., 2021). The set of commuting-zone level AUM estimates documented by (Chetty et al., 2014) for the United States is also wider than observed across French departments, which is certainly due, in part, to the difference in granularity between the two geographic units.

Some spatial patterns emerge from the geographic representation of the three statistics. First, 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 0.28 to 0.40 in departments in Brittany (West) and from 0.46 to 0.71 in departments in Hauts-de-France (North). This pattern is observed not only in terms of relative mobility (IGE and RRC), but also in terms of absolute upward mobility: while children with modest socio-economic backgrounds have relatively high expected income ranks in Brittany (AUM $\in$ (43; 45)), they tend to remain lower in the income distribution in Hauts-de-France (AUM $\in$ (36; 42)).

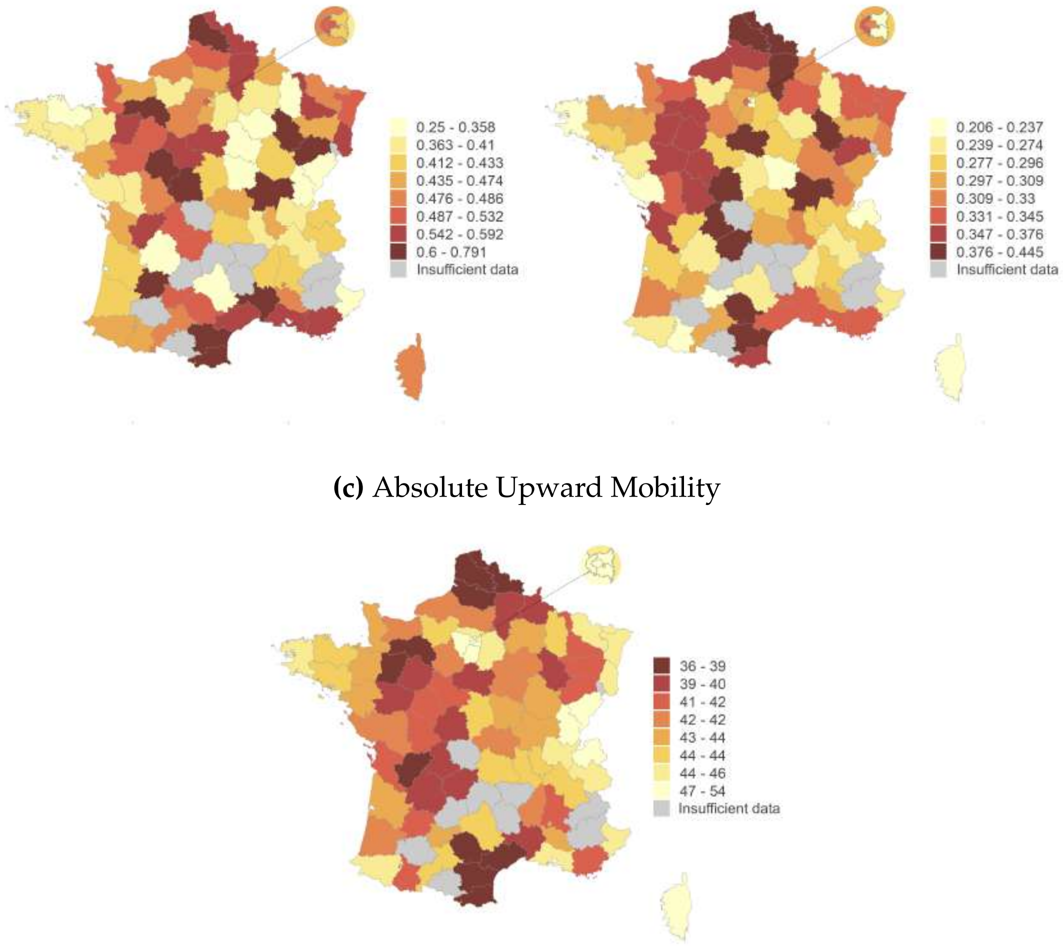
Second, departments that share a border with Switzerland (middle of the eastern border of France) also tend to exhibit high levels of relative and absolute mobility. Yet a high relative mobility is not systematically associated with a high absolute upward mobility. For instance, high relative mobility but low absolute upward mobility are typically observed in the empty diagonal.[46] Another instance of such a discrepancy is observed for the city-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 French average. The conditional expectation functions in Appendix Figure A22 provide an explanation to this idiosyncrasy. They reveal that the Parisian CEF is both shifted upwards relative to other large departments, and flatter at the left end of the

[44] Department-level estimates are reported in Appendix Table A9. Department-level intergenerational elasticities and rank-rank correlations are represented graphically with their confidence intervals in Appendix Figure A23. Spatial variations of intergenerational mobility at the more aggregate region level are documented in Appendix Figures A24 and A25.

[45] See Table 2 for cross-country variations in the rank-rank correlation.

[46] The empty diagonal - *diagonale du vide*, is a band of low-density population that stretches from the southwest to the northeast of France.

Figure 9: Spatial Variations in Intergenerational Mobility



(a) Intergenerational Elasticity

(b) Rank-Rank Correlation

(c) Absolute Upward Mobility

Notes: This figure presents the spatial variations of our baseline intergenerational mobility estimates at the department level. To compute local estimates, individuals are assigned to the department they lived in 1990, when they were between 9 and 18 years old. Departments with less than 200 observations are considered as having insufficient data. For practical reasons, we treat North Corsica and South Corsica as a single department. See Figure 1's notes for details on data, sample and income definitions.

income distribution. The combination of these two features results in relatively good prospects for children whose parents locate at the 25th percentile in Paris relative to other French departments without implying particularly high relative mobility.

Even though the three maps display similar spatial variations, they do not perfectly coincide. Table 4 shows the correlation between each intergenerational mobility measure for the three income definitions we use. The fact that correlation coefficients do not exceed 0.75 suggests the three statistics capture different processes. The most correlated statistics are the IGE and the RRC. The correlation coefficients of these two measures of relative mobility lie around 0.7 depending on the income definition used. The second strongest relationship is observed for the two rank-based estimates, the RRC and absolute upward mobility. Unsurprisingly, the least correlated statistics are the most distant ones in what they capture: the IGE and AUM.

Table 4: Correlation Between Department-Level Intergenerational Mobility Measures

Child income definition	IGE-RRC	RRC-AUM	IGE-AUM
Household income	0.65	-0.60	-0.47
Individual income	0.75	-0.50	-0.51
Wage	0.67	-0.37	-0.30

7.2 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 an initial set of 14 variables, described in Appendix Table A10, classified into 5 groups: demographic, economic, inequality, education, and social capital variables. We measure these variables as close to 1990 as possible so as to reflect the environment individuals grew up in. 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 A11 to A13 and summarized in Appendix Figure A26. 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.

There are three main take-aways. First, the IGE appears to only be significantly related to the unemployment rate, with a correlation of 0.30. This strong association is indeed striking visually when comparing the spatial distributions of the two variables (Figure 9a and Appendix Figure A27d). Second, we find no evidence of a within France "Great Gatsby Curve". The latter 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) and within some countries (Acciari et al., 2021). 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 the United States (Chetty et al., 2014) and Italy (Acciari et al., 2021). Third, absolute upward mobility tends to exhibit much stronger relationships with department characteristics in general, than either the IGE or the RRC. This could suggest that factors that affect absolute mobility might differ from those that affect relative mobility. While the signs of some coefficients are coherent with reasonable priors, like the positive relationship between AUM and the share of high-school graduates or mean wages, others are less clear (e.g. positive correlation with share of single mothers).

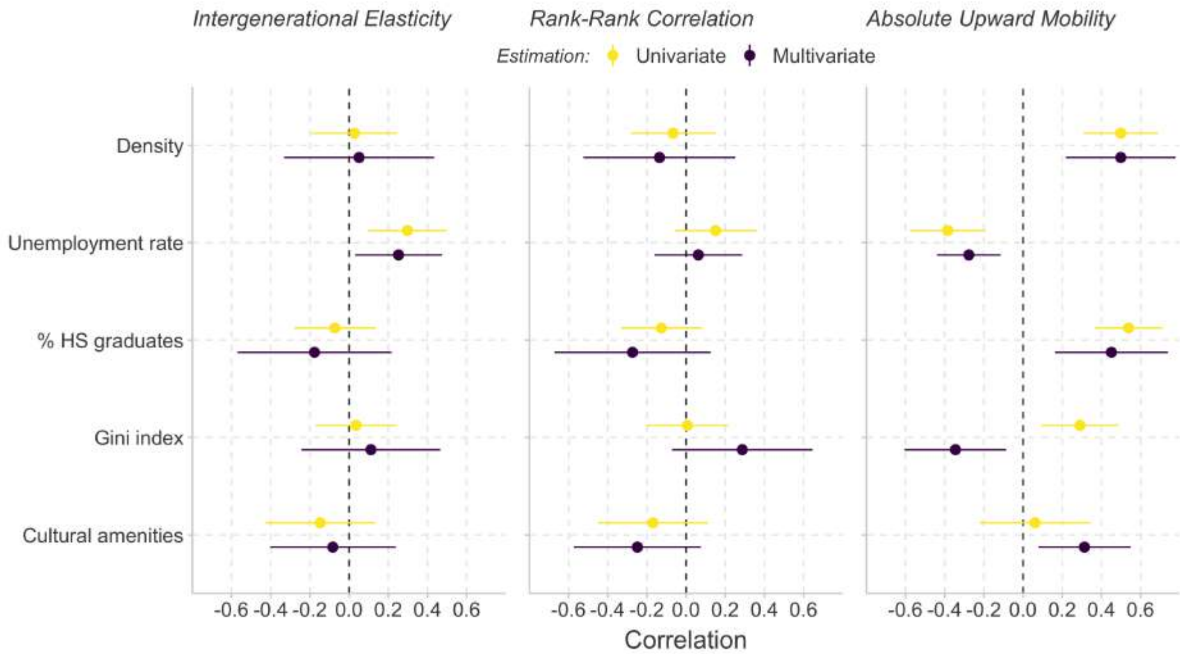
Appendix Figure A28 provides a potential explanation to these inconsistencies by documenting the correlation between all department characteristics. The 14 variables considered are for the most part quite strongly correlated with each other, both within and between variable groups. For instance, the Gini inequality 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. Therefore we focus on a subset of one variable per category, namely population density, the unemployment rate, the Gini index of inequality, the share of high-school graduates, and cultural amenities (number of museums and cinemas per inhabitant), and estimate their relationship with each intergenerational mobility measure jointly in a single regression.[47] The choice of these variables is based either on their intuitive relevance or their occurrence in the literature. Results of the joint regressions are presented in Appendix Table A14 and summarized in Figure 10, which also reports the univariate correlations for comparison purposes.

Multivariate correlations do not change the picture much. The IGE remains significantly correlated only to the unemployment rate while the RRC is not significantly correlated to any of the chosen characteristics. The unexpected positive relationship (though not significant) correlation with the Gini index remains for the IGE and the RRC. However, the sign of this correlation becomes negative for absolute upward mobility once other characteristics are controlled for. This within-country negative relationship between AUM and inequality is also observed in Italy (Acciari et al., 2021) and in North America (Chetty et al. (2014) for the United States and Corak (2020) for Canada). None of the other characteristics change signs and they retain the same level of significance as in the univariate case. Specifically, population density, the share of high school graduates and cultural amenities are positively and significantly correlated to absolute upward mobility, while the unemployment rate and the Gini index are negatively correlated. These results highlight the potentially different roles played by these factors on relative versus absolute mobility.

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 let the causal assessment of the effect of sorting and place characteristics on intergenerational mobility to future studies. The evidence we put forward on the potential sources of intergenerational persistence, which is particularly high in France, strengthens the plausibility of

[47] The spatial distribution of these variables at the department level are presented in Appendix Figure A27.

Figure 10: Intergenerational Mobility and Department Characteristics



Notes: This figure presents the regression coefficients between department-level intergenerational mobility and department characteristics. Coefficients of univariate and multivariate regressions are shown. Both the department intergenerational mobility estimates and the characteristics are standardized, implying that the coefficients can be interpreted as correlations or partial correlations in the multivariate case. See Figures 1 and 9's notes for details on data, sample and income definitions, and Appendix Table A10 for definitions and sources of the department characteristics.

the implication of local characteristics such as income inequality and access to cultural amenities in shaping individuals' intergenerational mobility prospects.

8 Conclusion

We use rich administrative data to provide an overview of intergenerational income mobility in France for individuals born in 1972-1981. The estimations put forward all along this analysis reveal a relatively strong intergenerational income persistence at the national level, among the highest in OECD countries and far from the particularly mobile Scandinavian countries. Our preferred estimate of the intergenerational elasticity in household income is 0.515. It suggests that had a child's parental lifetime income been higher by a given amount, this would have translated by an increase of about half of this amount in her lifetime income as an adult.

Following the recent standards of the literature, we provide the first estimates of intergenerational mobility in terms of percentile rank in the income distribution for France. Our baseline rank-rank correlation suggests that a child would on average rank 3.37 percentiles higher in the income distribution of her birth cohort if her parents had located 10 percentile ranks higher in their respective income distribution. Our sensitivity tests show the rank-rank correlation to be more robust than the intergenerational elasticity to a variety of specification variations. Rank-rank correlations thus appear particularly convenient in settings where the two-sample two-stage least squares method is necessary to palliate data limitations.

We find that this intergenerational persistence is particularly strong at the tails of the parent income distribution. Children born to parents in the bottom 20% of their income distribution have a 10.1% probability of reaching the top 20% as adults. This probability is of 39.1% for children born to parents in the top 20%.

Moreover, we uncover considerable heterogeneity in intergenerational mobility across French departments. Intergenerational persistence appears to be particularly high in the North and South, and relatively low in the Western part of the country. We find that population density, the share of high-school graduates and cultural amenities are positively correlated with absolute upward mobility while income inequality and unemployment rate are negatively correlated.

References

Acciari, Paolo, Alberto Polo, and Giovanni Violante, “"And Yet It Moves": Intergenerational Mobility in Italy,” 2021.

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 2021.

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

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

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.

Charpentier, Arthur, Emmanuel Flachaire, and Antoine Ly, “Econometrics and Machine Learning,” *Economie et Statistique / Economics and Statistics*, 2019, (505d), 147–169.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *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,” 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,” IZA Discussion Paper 9929, IZA, Bonn, Germany 2016.

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

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,” 2008, p. 44.

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

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

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.

INSEE, *Courrier Des Statistiques N6* 2021.

— **, ed.**, *Tableaux de l'économie française. Edition 2020*, Paris: Insee, 2020.

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.

Jugnot, Stéphane, "La Constitution de l'échantillon Démographique Permanent de 1968 à 2012," Documents de Travail de La Direction Des Statistiques Démographiques et Sociales F1406, Institut National de la Statistique et des Études Économiques, Paris 2014.

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.

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.

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.

— **and Nathan Deutscher**, "Measuring Intergenerational Income Mobility: A Synthesis of Approaches," 2021.

Mogstad, Magne and Gaute Torsvik, "Family Background, Neighborhoods and Intergenerational Mobility," 2021.

Nilsen, Øivind Anti, Kjell Vaage, Arild Aakvik, and Karl Jacobsen, "Intergenerational Earnings Mobility Revisited: Estimates Based on Lifetime Earnings," *The Scandinavian Journal of Economics*, 2012, *114* (1), 1–23.

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.

Pekkarinen, Tuomas, Kjell G. Salvanes, and Matti Sarvimäki, "The Evolution of Social Mobility: Norway during the Twentieth Century," *The Scandinavian Journal of Economics*, 2017, *119* (1), 5–33.

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.

The Anatomy of Intergenerational Income Mobility in France and its Spatial Variations

Appendix

Gustave Kenedi Louis Sirugue

A Data

The Permanent Demographic Sample is a socio-demographic panel combining several administrative data sources. It contains information on individuals born on the first four days of October.[48] Individuals born on one of these days are called EDP individuals. The EDP gathers data from 5 administrative sources: (i) civil registers (births, adoptions, marriages, and deaths) 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, which combines data from the annual declaration of social data (DADS) and data on central government employees; and (v) tax returns data since 2010.

Each time an individual born on the first four days of October appears in one of these administrative datasets with the exception of the electoral register, the information contained in it is added to his or her individual ID in the EDP. Therefore all these datasets can be matched together using the common individual identifier. The EDP is an unusual panel, but its singular data collection principle has allowed to progressively gather a very rich set of socio-demographic characteristics throughout an individual's lifecycle, making it increasingly suited to study mobility-related issues.

For our analysis we use data from birth certificates, the 1990 census, the All Employee Panel and the tax returns data. We describe each data source in detail below.

Birth certificates. Variables in the birth certificates dataset are gathered since 1968 from civil status forms, be it birth certificates, transcriptions of a declaratory judgment of birth, or transcriptions of a judgment of plenary adoption. They contains information on the birth of the EDP individual such as gender, date and place of birth, and information on each parent including date and place of birth as well, but also nationality and occupation.

1990 Census. The EDP subset of the 1990 census contains variables from the population census forms filled by EDP individuals. Census information document the socio-demographic status of EDP individuals, as well as, even though to a lesser extent, those of their family and household members. It includes information on date and place of birth, nationality, education, occupation, marital status, household structure, but also on individuals' dwelling, building when relevant, and municipality.

All Employee Panel. The All Employee Panel combines two sources of data: the annual declarations of social data (*déclarations annuelles des données sociales* - DADS) and

[48]The EDP selection criterion has progressively widened to include individuals born on the first days of January, April, and July.

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.[49] 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 1990, the All Employee Panel data were not compiled by INSEE in 1981, 1983 and 1990.

Tax returns data. The tax returns data is 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 1st 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 are married or in a civil union, but also for couples who live together, an increasingly common arrangement.[50] 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.

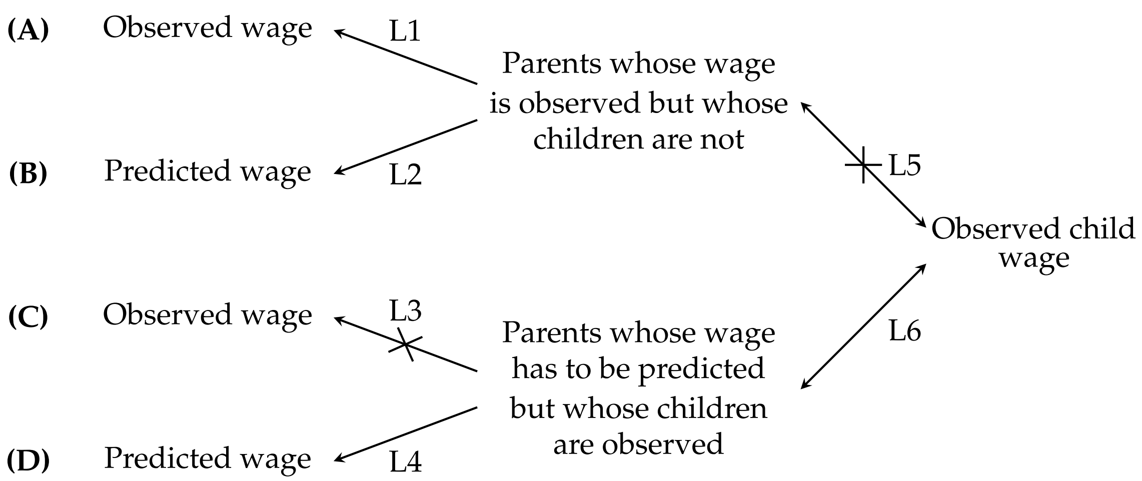
[49] France Télécom and La Poste employees appear only in 1988 as well. We ensure our results are not affected by the fact that civil servants are only observed from 1988 onwards by estimating the first-stage regression computing synthetic parents' on post-1988 wages only, still restricting to when they are between 35 ad 45 years old. Appendix Figure A19 displays the results from this check. The results are largely unaffected.

[50] In France, families headed by non married couples are as frequent as single-parent families (INSEE, ed, 2020).

B TSTSLS Bias-Correction Methodology

This section outlines the methodology we propose to estimate the bias of the IGE induced by predicting parent income. Figure A1 depicts our two-sample two-stage least squares (TSTSLS) setting. The first stage consists in estimating the prediction model using parents whose wage is observed (link L1) but whose children are not. The prediction model is then applied to parents whose wage is unobserved (link L4) to perform the second stage, i.e., regressing child log income on parent predicted log income. This corresponds to case (D), and the validity of this procedure relies on the assumption that there is no systematic bias in parent income predictions (i.e., that L4 correctly reproduces L3). The relationship we wish to estimate is the one corresponding to case (C), and the difference between the coefficient obtained from (C) and the one we obtain from (D) constitutes the bias induced by predicting parent income.

Figure A1: Schematic depiction of the TSTSLS setting



With our data and setting, we propose to estimate the prediction bias of the IGE by using parents whose wage is observed, i.e. our sample of synthetic parents. Specifically, we use both their observed wage and their wage predicted out of sample (using the same first-stage predictors as in our main analysis) to replicate the prediction bias. This corresponds to comparing coefficients from cases (A) and (B) instead of cases (C) and (D), which requires matching our synthetic parents to children (missing link L5). We proceed in three steps:

Step 1: We match children to our sample of synthetic parents following the rules laid out below.

Step 2: For a given match, we compare the resulting intergenerational elasticity estimated with predicted parent wage to our baseline TSTSLS estimate, and infer whether the match is plausible or not: if the estimate obtained with predicted parent wage falls within the confidence interval of our baseline TSTSLS estimate, the match can be considered plausible, otherwise it is considered implausible.

Step 3: We estimate the intergenerational persistence on observed wage rather than on predicted wage for every plausible match. The resulting estimates provide a bounded range for the coefficient we aim to estimate, net of the prediction bias.

We apply this methodology to fathers and sons' wages. Because there are 15,583 fathers and 36,014 sons, it is computationally not feasible to test every possible match. To obtain a reasonable amount of plausible matches, we impose two parameters. These two parameters are:

1. The income interval from which a son can be drawn from, ranging from the initial position of the father (i.e. the closest son from their own relative continuous position on a scale from 1/N to 1) to ± 100% around the initial position of the father (i.e. any son at random within the whole distribution) *(mobility restriction)*

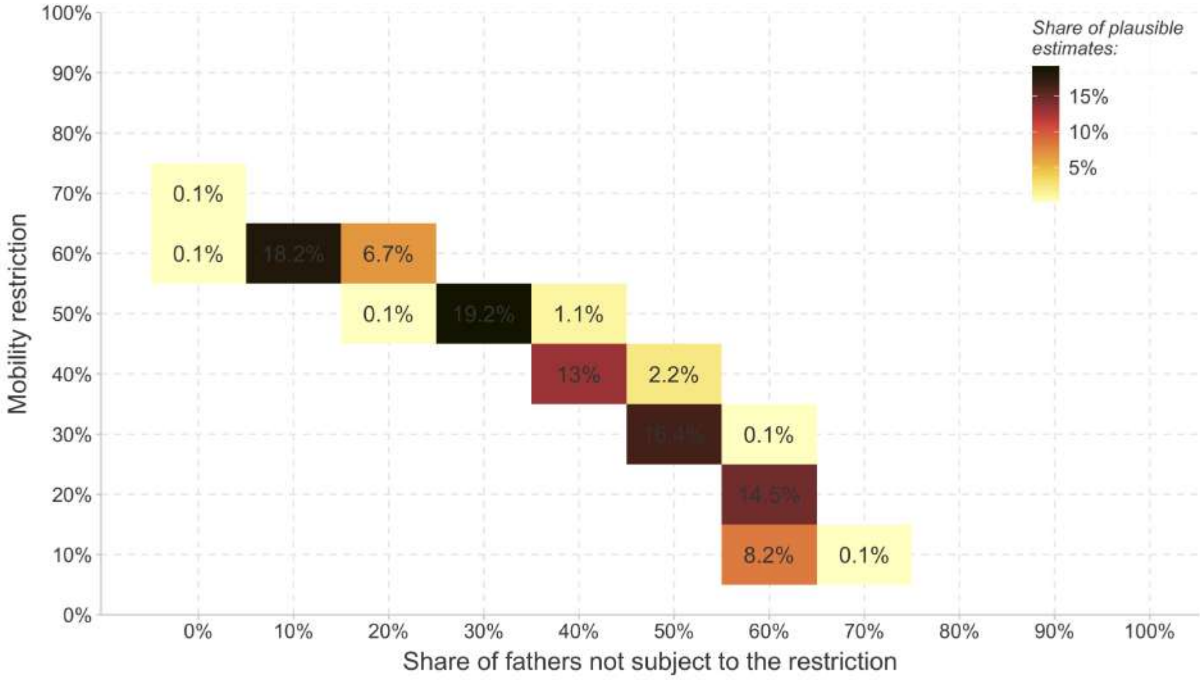
2. The share of fathers not subject to the restriction above, ranging from 0% to 100%.

When either of these two parameters equals 1, no potential match is discarded, and at the limit the confidence interval we ultimately compute contains the true coefficient net of the prediction bias. When both of these parameters equal 0, every match is discarded except the situation in which each son is as close as possible to the position of the father. Every intermediate combination of parameters allows to converge more rapidly to a confidence interval containing a reasonable amount of plausible estimates.

We vary the mobility restriction from 10% to 100% and the share of fathers subject to this restriction from 0% to 100%, both in increments of 10 percentage points. This yields 110 possible parameter combinations. We simulate matches with these 110 combinations of parameters 200 times, yielding 22,000 simulated matches. For every simulated match, we compute the IGE using predicted father wage. Whenever this yields a coefficient that falls within the confidence interval of the baseline estimate computed on parents whose children are observed but whose wage has to be predicted (i.e. $[0.41, 0.47]$), the corresponding match is considered as plausible. Figure A2 shows the couples of parameters yielding plausible estimates and their proportion. Each cell represents the share of plausible estimates obtained from the corresponding couple of parameters among all estimates. This figure shows that plausible estimates are not induced by a specific parametrization, but come from a range of intermediate combinations of the two parameters.

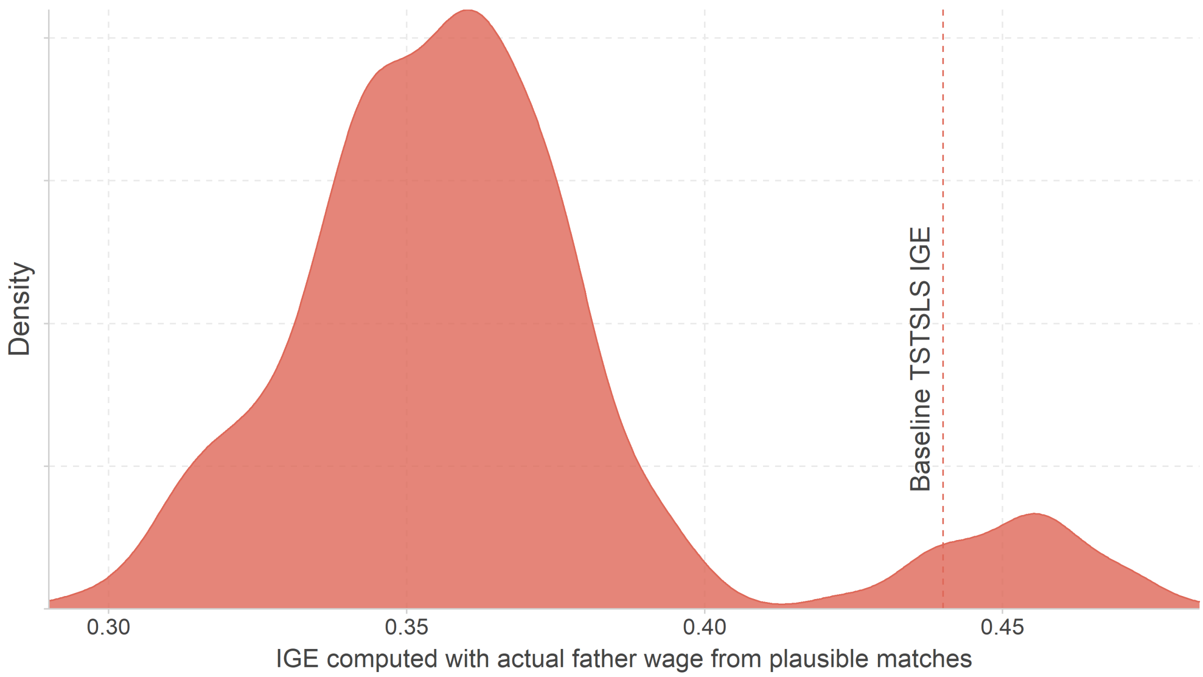
Figure A3 represents the density of IGEs resulting from plausible matches and computed using the actual wage of fathers. It suggests that in our setting, the TSTSLS strategy tends to overestimate the IGE. The mean and the median of this distribution amount to 0.36, which is relatively low relative to our baseline estimate of 0.44, even though it does not change the position of France relative the other OECD countries according to Table A8. As more work is needed to confirm the results from these simulations, we do not discuss them in detail in the paper.

Figure A2: Share of Plausible Estimates per Couple of Parameters



Notes: This figure plots for each combination of parameters the share of plausible estimates obtained from Step 2. See Appendix above for details.

Figure A3: Density of IGEs Computed with Actual (Synthetic) Father Wage from Plausible Matches



Notes: This figure plots the density of estimates obtained from Step 3. See Appendix above for details.

C Details on Analysis of Higher Education Graduation by Parent Income

In the EDP, information on the education level of our sample of children can be retrieved from the population censuses. Specifically, these ask for individuals' last obtained diploma. We code as "graduated from higher education" any last obtained diploma greater than a high school degree. Since 2004, France conducts annual census surveys on a 20% rotating random sample of the population.[51] Thus, a full census is completed every five years. Because individuals are (theoretically) surveyed once every five years, we cannot compute directly the fraction of children ever attending college between age 18-21 within each parent-income percentile bin, the statistic provided by Chetty et al. (2014) for the United States. Around 80% of the sample is observed at least once during this age range, with the remaining 20% not being observed at all. Since individuals may not enroll into higher education straight away and may drop out, having only one observation of enrollment between 18 and 21 is insufficient to infer whether the individual has been enrolled in higher education at any time between 18 and 21. This is why we focus on whether individuals graduated or not from higher education. We are able to retrieve this information for 86.29% of our sample.

[51] Technically, it is a 20% random sample of cities not individuals.

Appendix Figures

Figure A4: Schematic Depiction of the Two-Sample Two-Stage Methodology

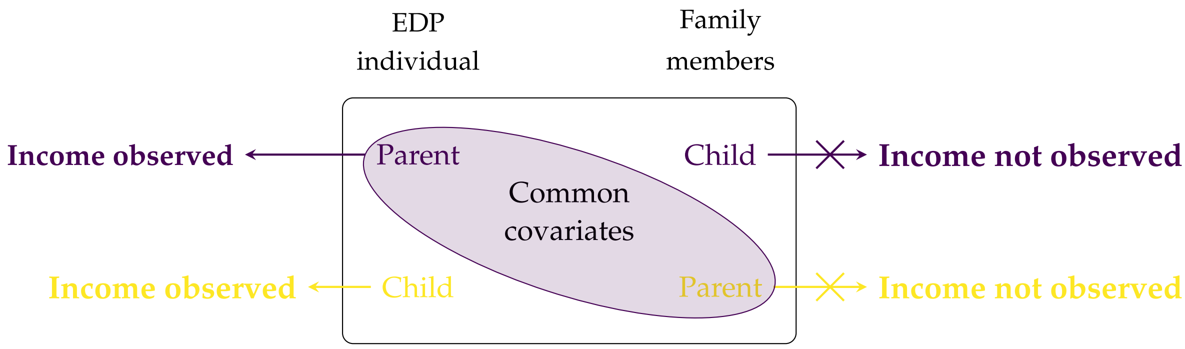
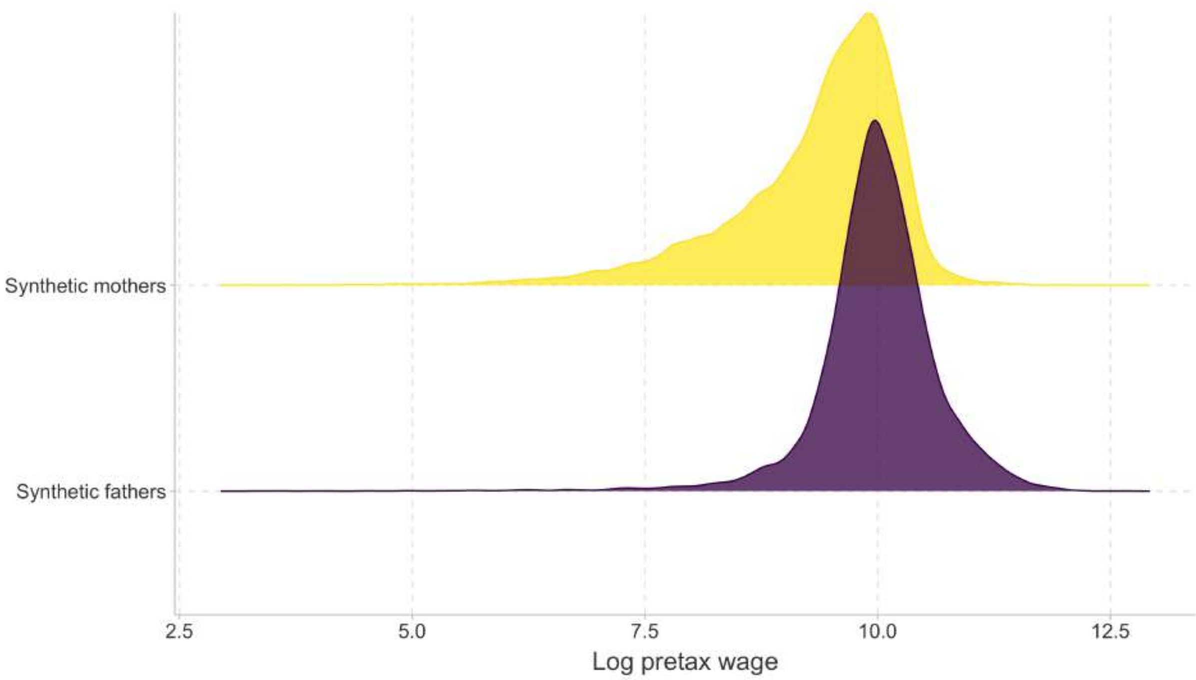
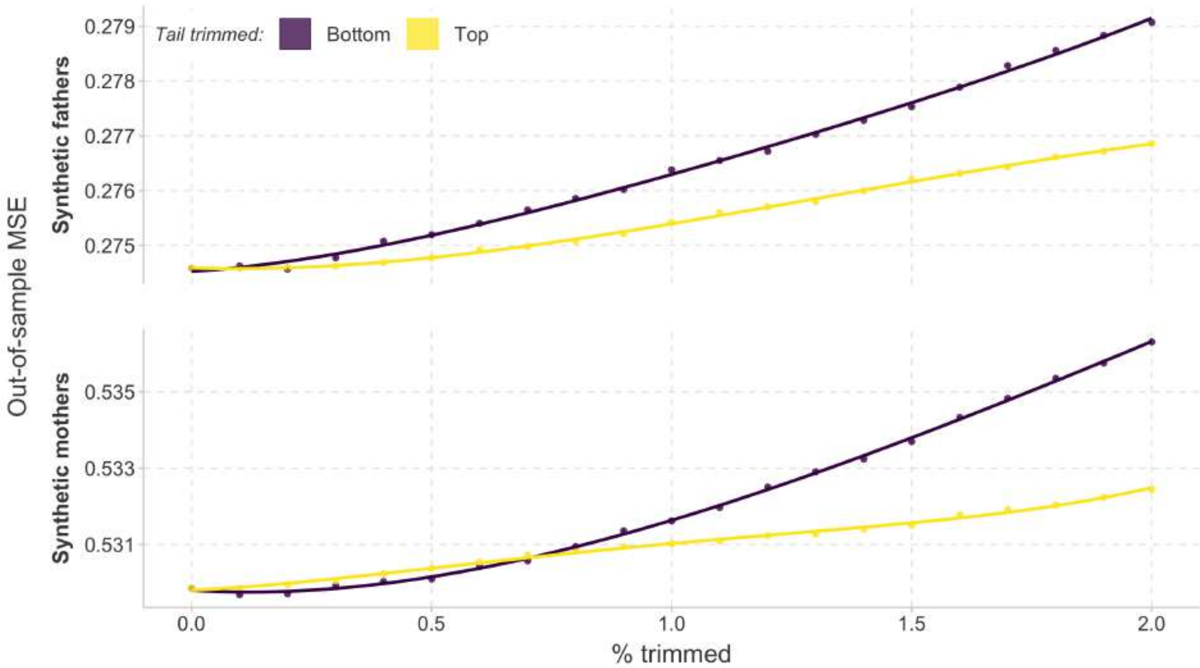


Figure A5: Distribution of Synthetic Parents' Average Pretax Wage Between 35 and 45



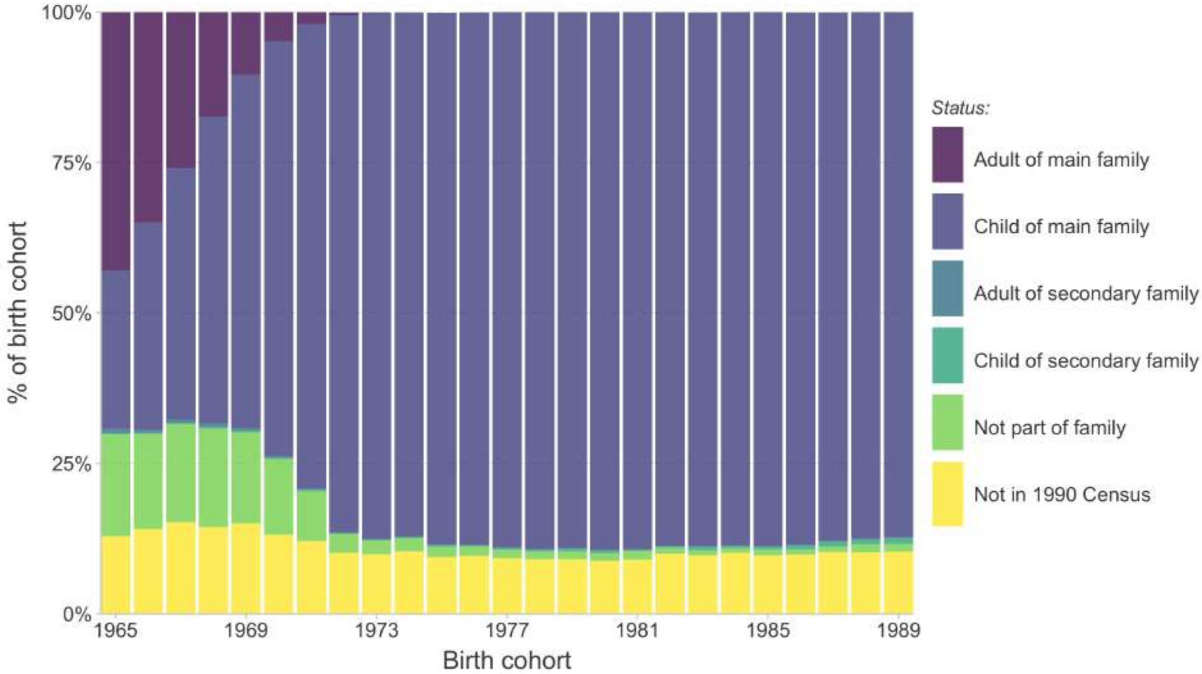
Notes: This figure represents the density distribution of synthetic parents' average pretax wage between 35 and 45. See Section 4.1 for details on sample construction.

Figure A6: Out-of-sample MSE as a Function of Top and Bottom Trimming



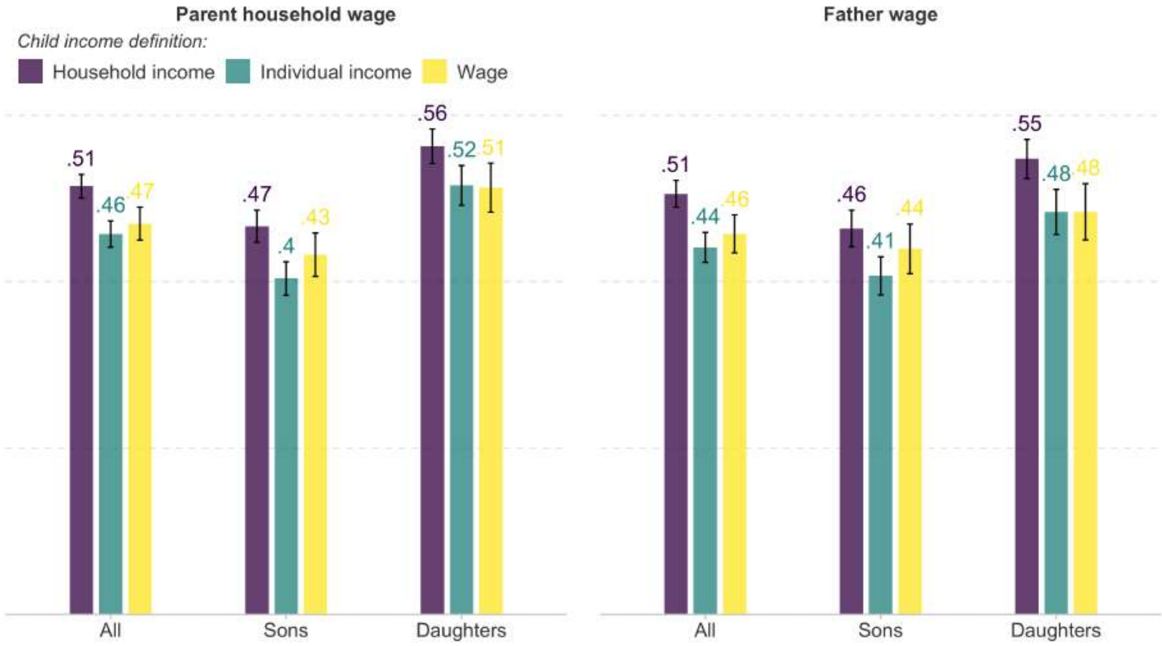
Notes: This figure plots the out-of-sample MSE as a function of trimming various shares of the tails of synthetic parents' income distribution. Our-of-sample MSEs correspond to the average MSE obtained from 5-fold cross-validation. See Sections 3 and 4.1 for details on the exact model being estimates and sample construction.

Figure A7: Family Position in 1990 Census by Child Birth Cohort



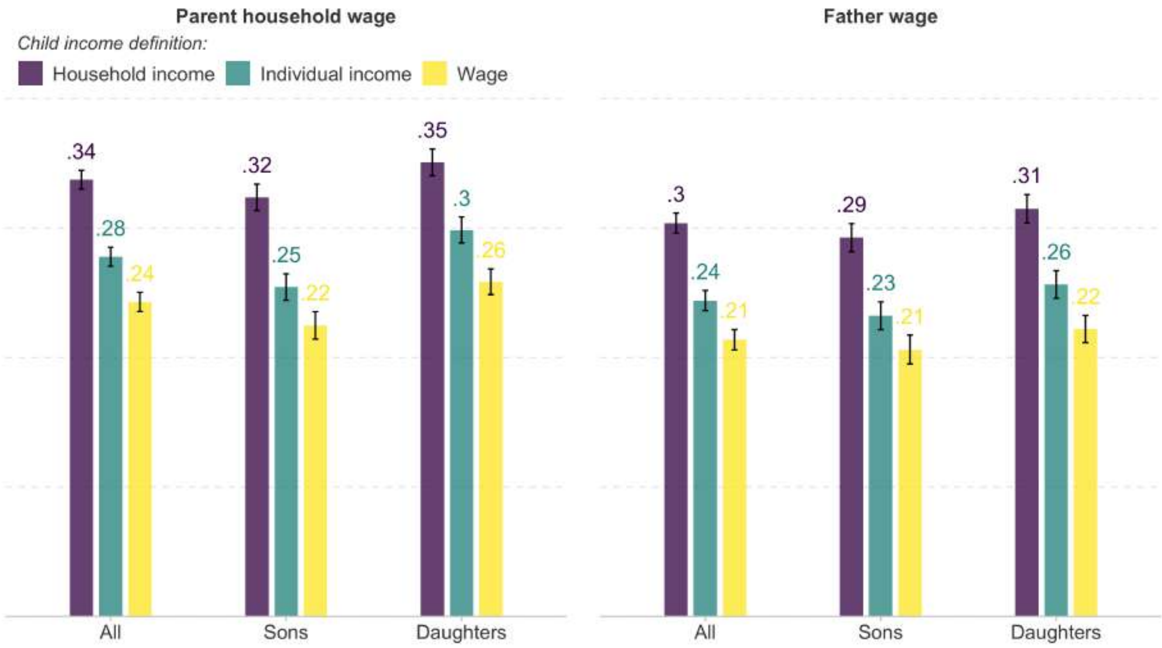
Notes: This figure presents the family position of EDP individuals in the 1990 census by birth cohort. The sample is restricted to EDP individuals born in metropolitan France.

Figure A8: Baseline IGE Estimates for Different Child and Parent Income Definitions



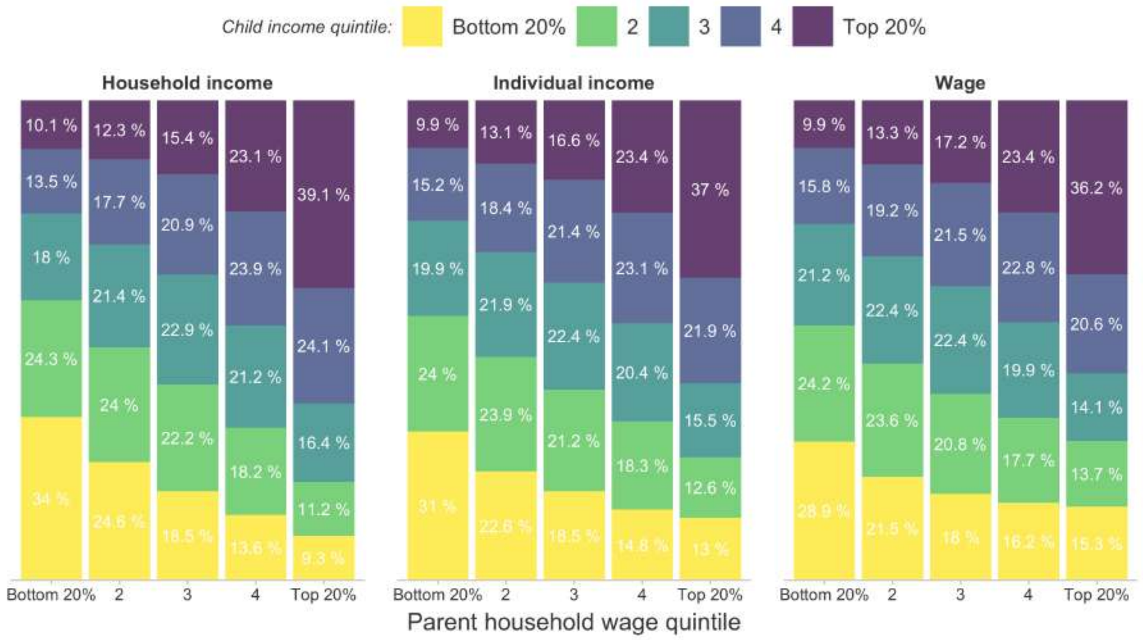
Notes: This figure presents our baseline intergenerational income elasticity estimates for various parent and child income definitions. Each bar represents the coefficient of an OLS regression of child income on parent income, for the entire sample ("All") and for sons and daughters separately. See Section 4 for for details on data, sample and income definitions.

Figure A9: Baseline RRC Estimates for Different Child and Parent Income Definitions



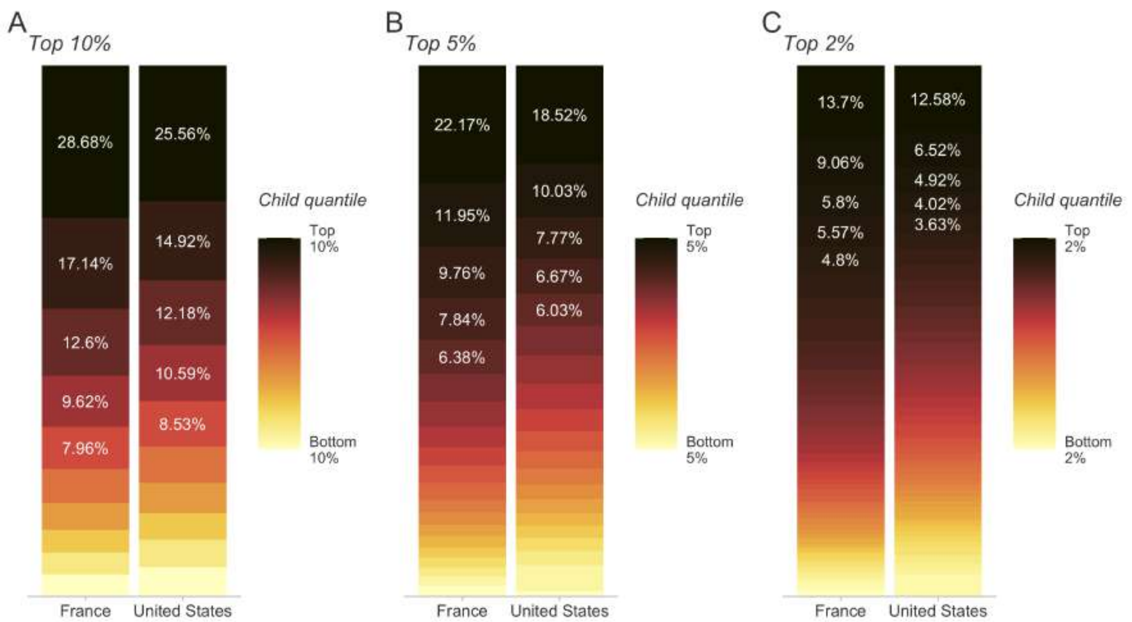
Notes: This figure presents our baseline intergenerational rank-rank correlation estimates for various parent and child income definitions. Each bar represents the coefficient of an OLS regression of child income rank on parent income rank, for the entire sample ("All") and for sons and daughters separately. See Section 4 for for details on data, sample and income definitions.

Figure A10: Baseline Quintile Transition Matrix for Different Child Income Definitions



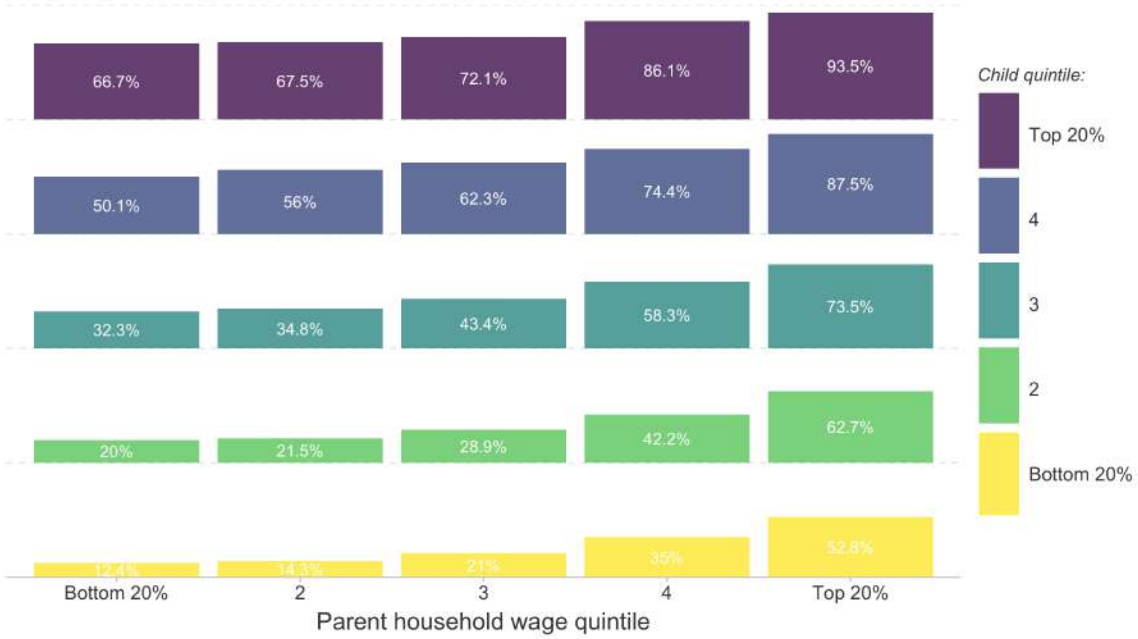
Notes: This figure presents our baseline intergenerational transition matrix estimates for various child income definitions. Each cell corresponds to the percentage of children in a given income quintile who have parents in a given parent income quintile. See Section 4 for for details on data, sample and income definitions.

Figure A11: Top Parent Income Quantiles Transition Matrices in France and US



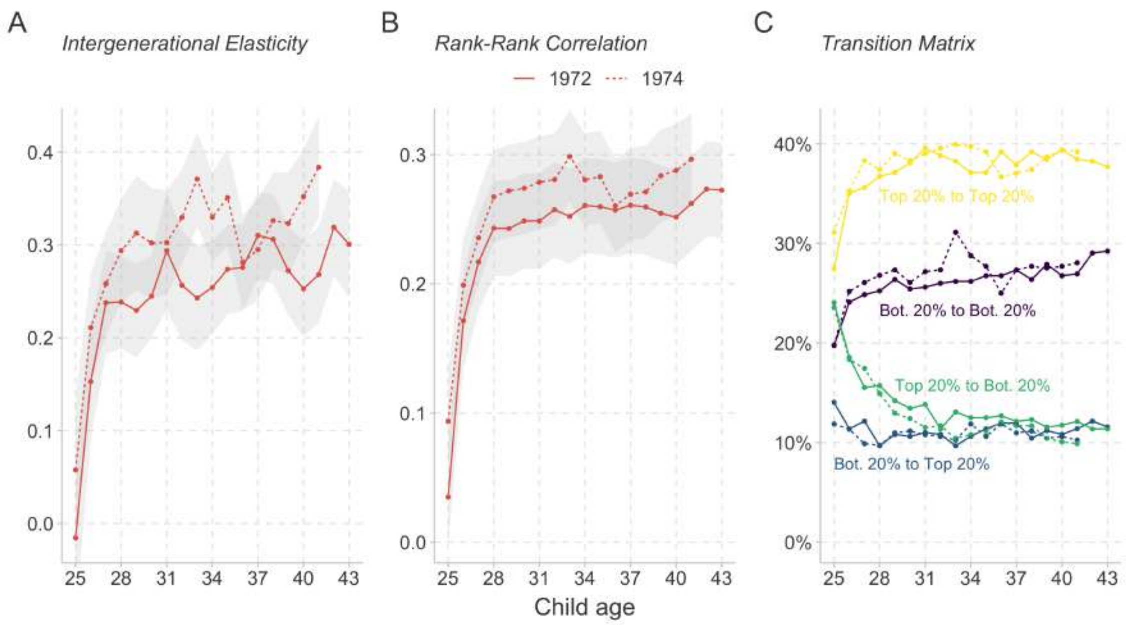
Notes: This figure presents intergenerational transition matrix estimates for children coming from families in the top 10% (panel A), top 5% (panel B) and top 2% (panel C) of the parent income distribution. We compare the transition probabilities we obtain for France with those computed by Chetty et 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 4 for for details on data, sample and income definitions.

Figure A12: 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 4 and 5.4 for for details on data, sample and income definitions.

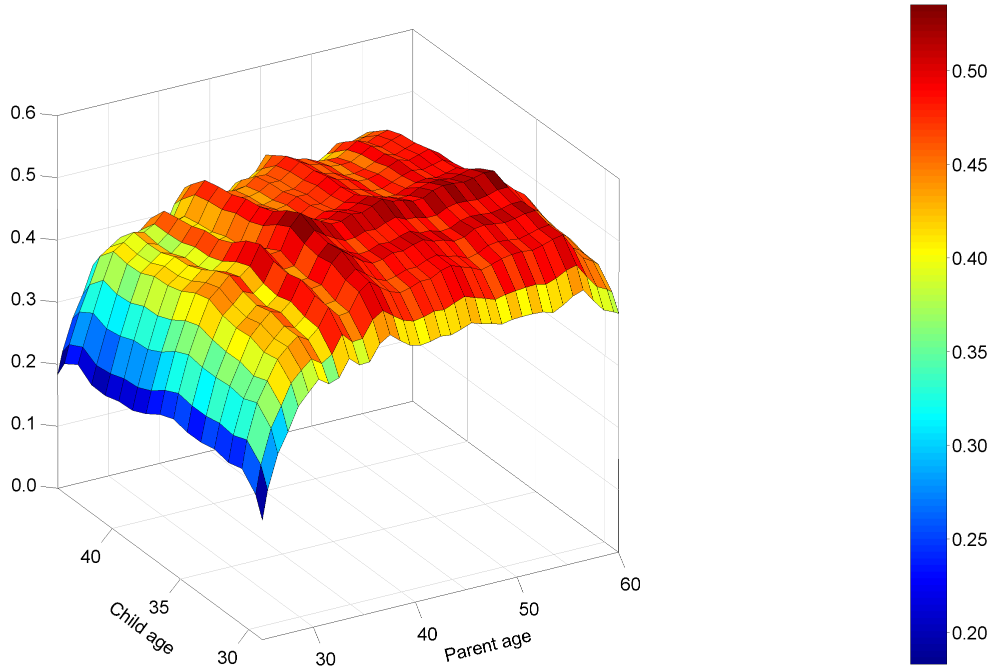
Figure A13: Child Lifecycle Bias - 1972 and 1974 Cohorts (Constant Sample)



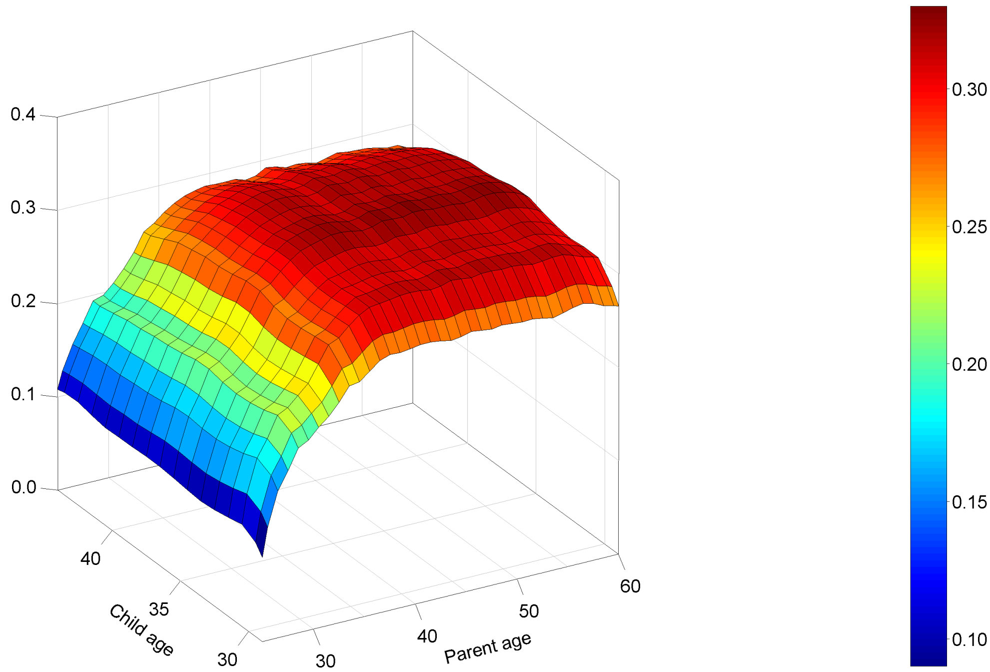
Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figures 1 and 2 to changes in the age at which child income is measured, for children born in 1972 (solid line) and 1974 (dashed line). For both birth cohorts the sample is kept constant, 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% confidence interval. See Sections 4 and 5.4 for for details on data, sample and income definitions.

Figure A14: Child and Parent Lifecycle Bias

(a) Intergenerational Elasticity

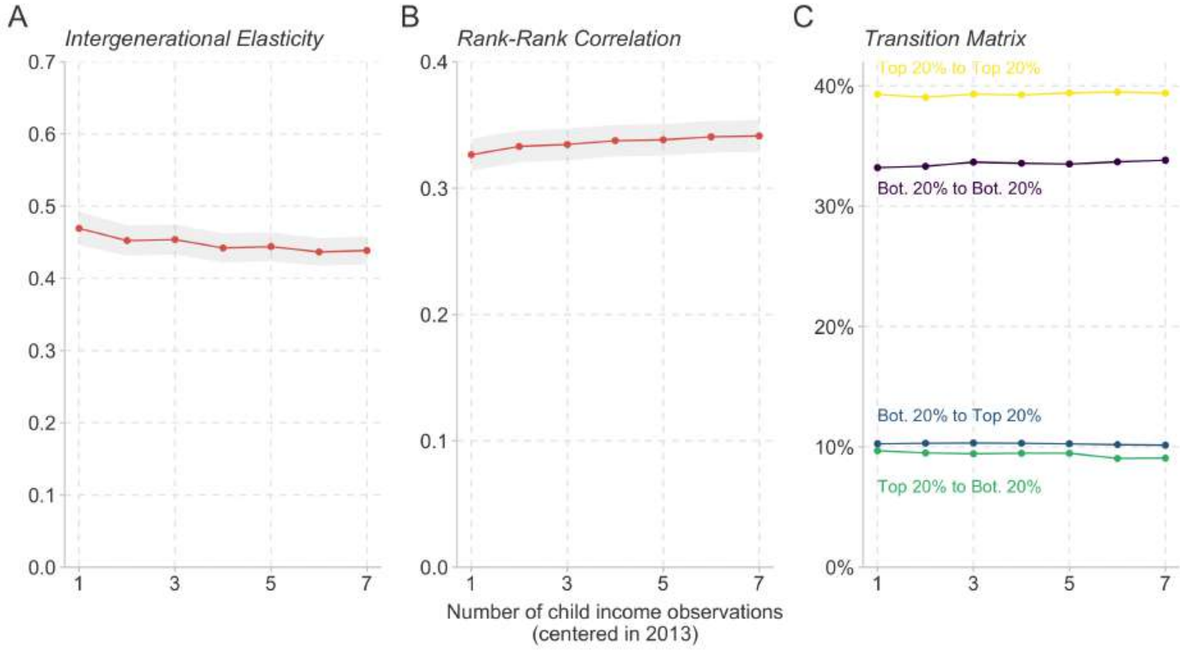


(b) Rank-Rank Correlation



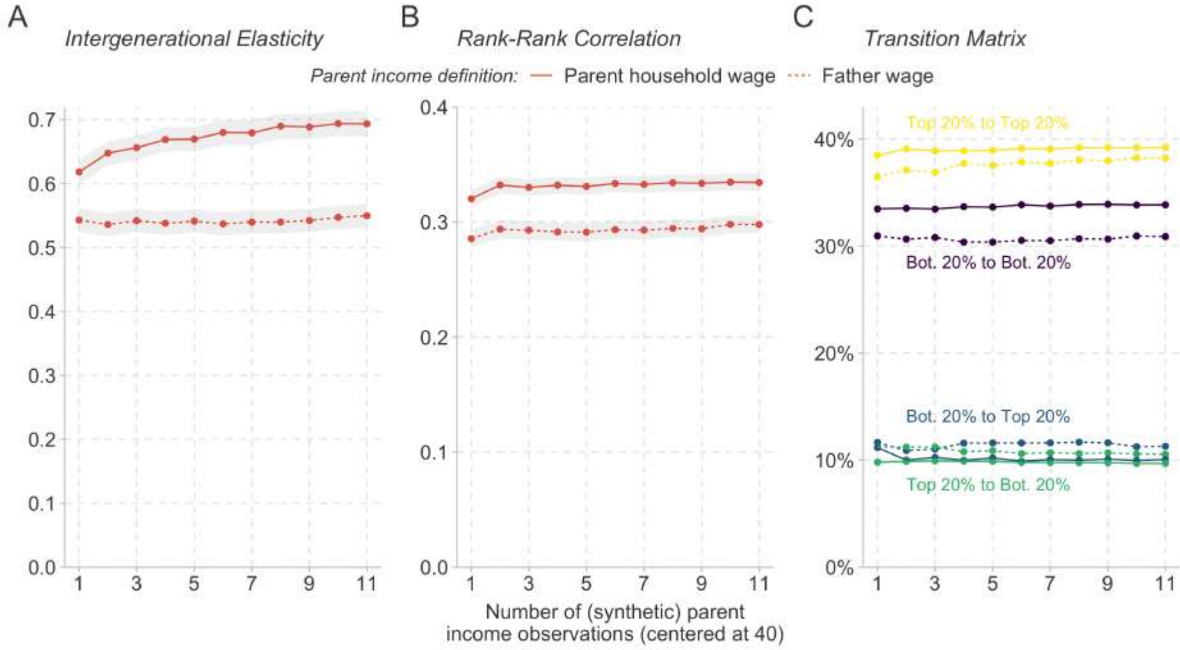
Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates presented in Figure 1 to changes in the age at which child and (synthetic) parent incomes are measured. The sample of children and synthetic parents varies across ages. See Sections Figure 1's notes for for details on data, sample and income definitions.

Figure A15: Sensitivity to Number of Child Income Observations (Constant Sample)



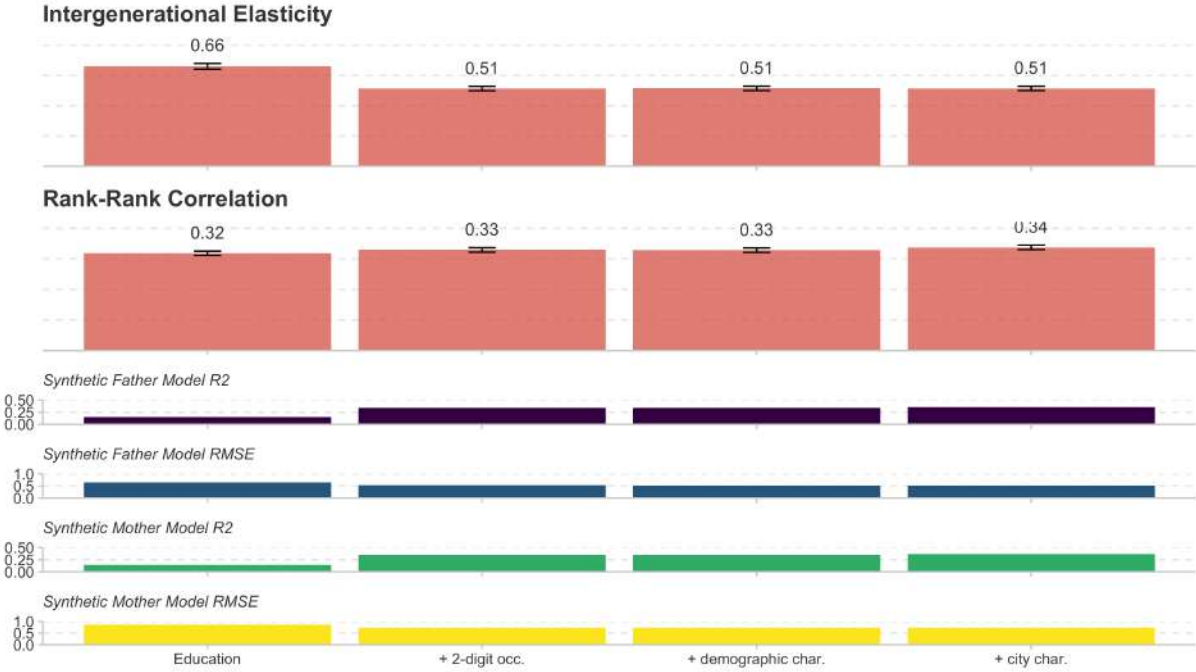
Notes: This figure presents estimates of our persistence measures varying the number of child income observations from 1 to 7 between 35 and 45 years old, keeping the sample of children constant, i.e. keeping only children with 7 household income observations. (The sample varies ever so slightly for 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. 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% confidence interval. See Figure 1's notes for for details on data, sample and income definitions.

Figure A16: Parent Attenuation Bias (Constant Sample of Synthetic Parents)



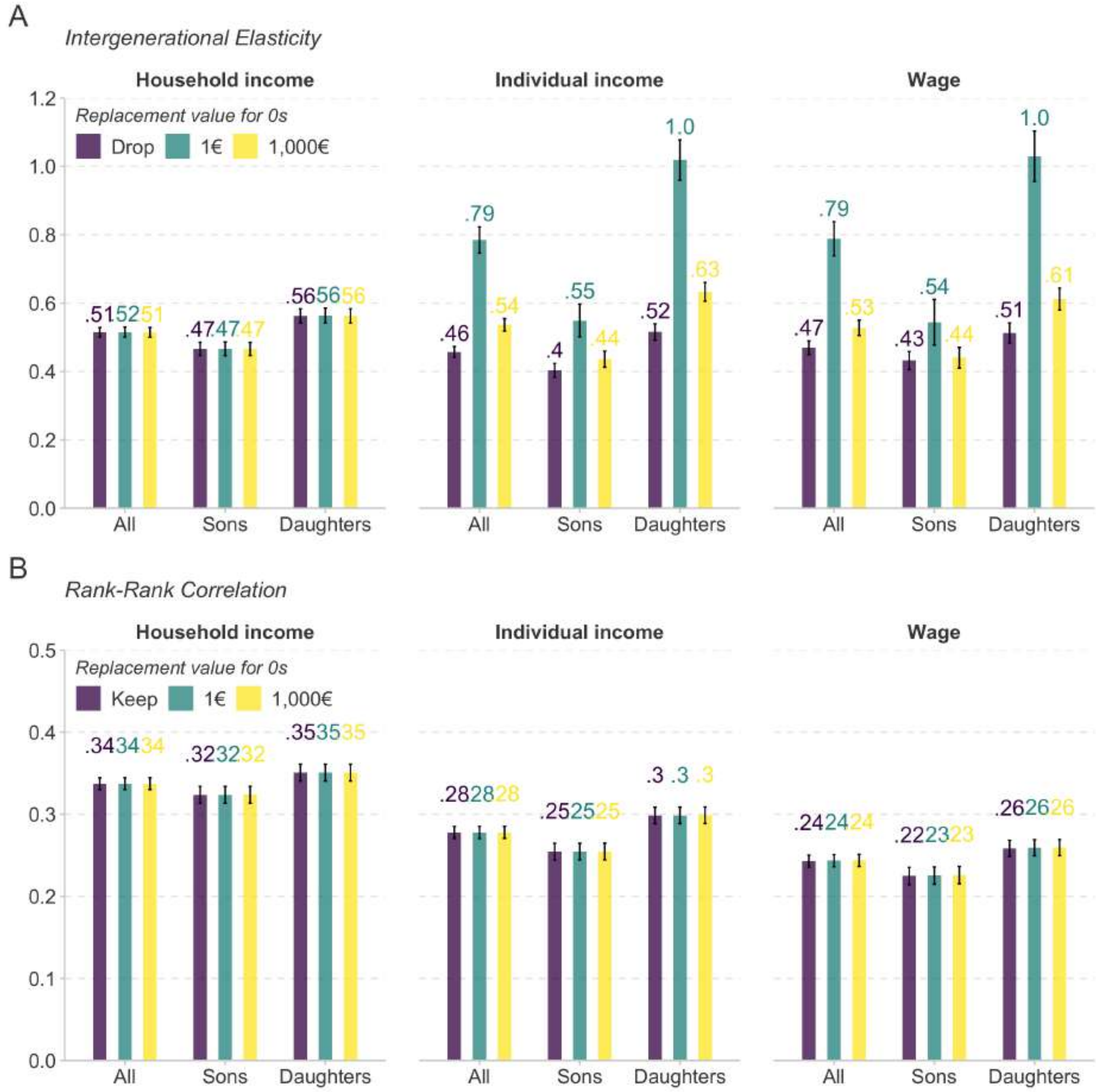
Notes: This figure assesses the robustness of our baseline intergenerational income mobility estimates to the number of income observations used to predict parent income, keeping the sample of synthetic parents constant. The sample of synthetic parents is thus restricted to those with all 11 income observations 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% confidence interval. See Figure 1's notes for details on data, sample and income definitions.

Figure A17: Robustness of Baseline Estimates to Different First-Stage Predictors



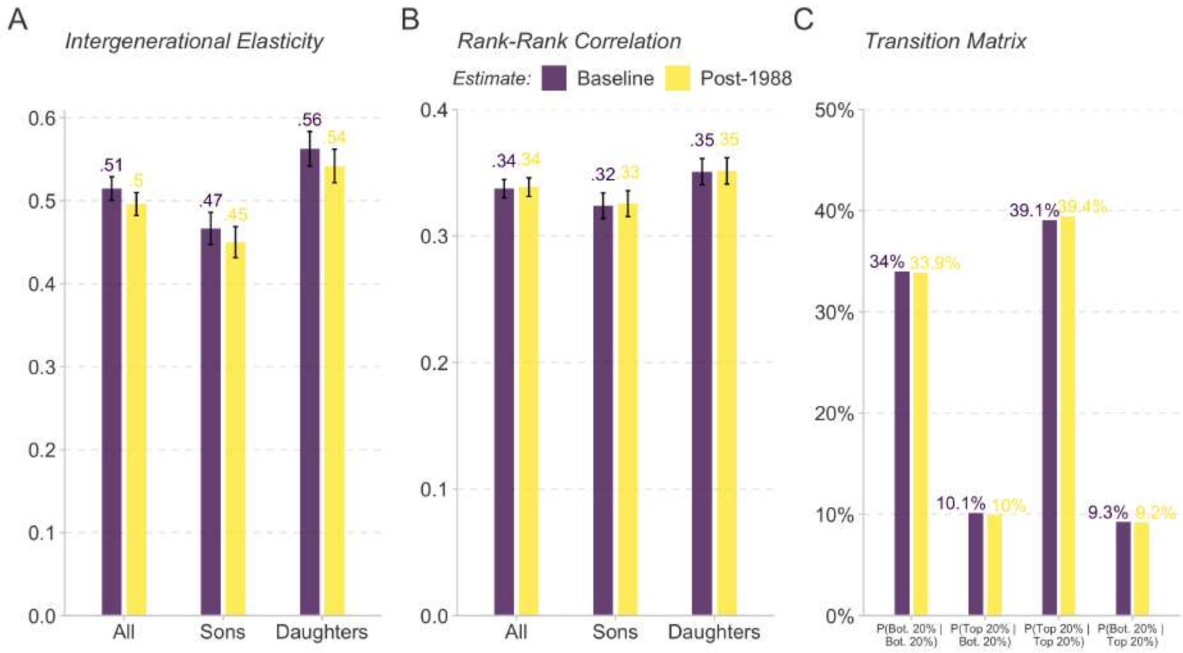
Notes: This figure assesses the robustness of our baseline IGE and RRC estimates to variations in the set of first-stage predictors. Parent income is predicted separately for fathers and mothers using a set a of instruments that vary along the x-axis. The bottom four panels of the figure reports separately for synthetic fathers and mothers the R^2 and root mean squared error (RMSE) associated with each first stage. See Figure 1's notes for details on data, sample and income definitions.

Figure A18: Sensitivity to Different Zero Child Income Replacement Values



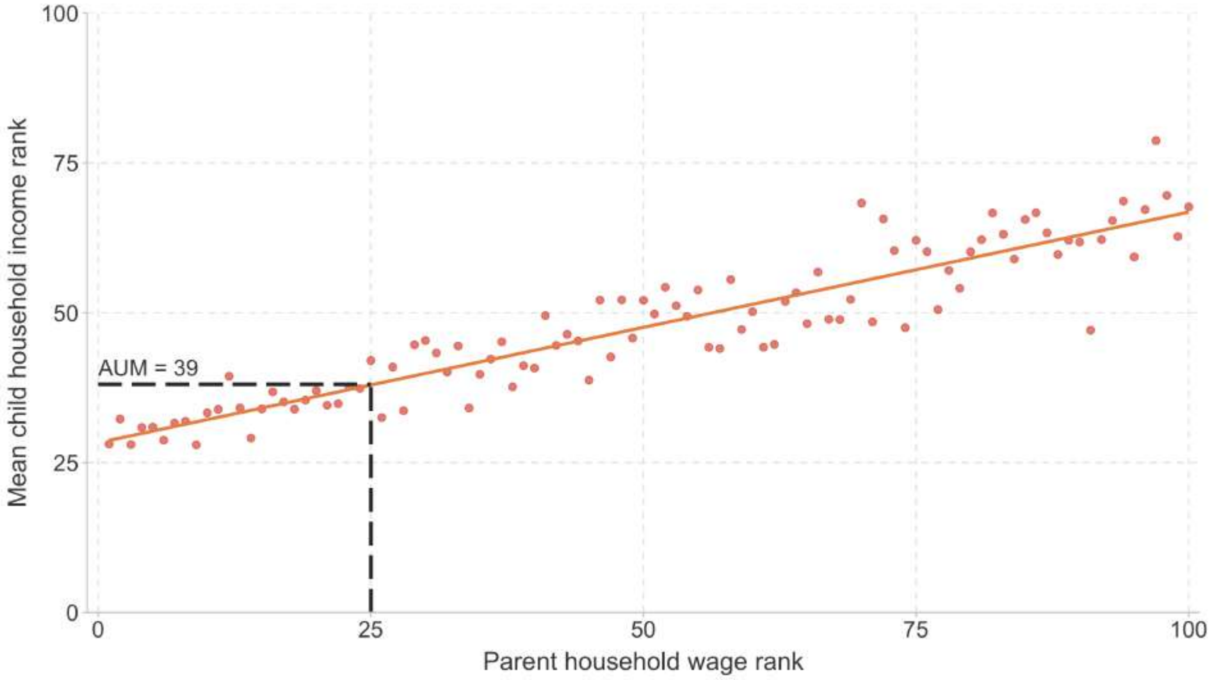
Notes: This figure assesses the robustness of our baseline IGE and RRC estimates to replacing incomes of children reporting negative or zero incomes by 1 euro or 1,000 euros, for different child income definitions. See Section 4 for for details on data, sample and income definitions.

Figure A19: 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). 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 4 for details on data, sample and income definitions.

Figure A20: 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 39, for children whose parents locate at the 25^{th} percentile. The orange line is a linear regression fit through the conditional expectation. See Figure 1's notes for details on data, sample and income definitions.

Figure A21: Department-Level Log-Log Relationships

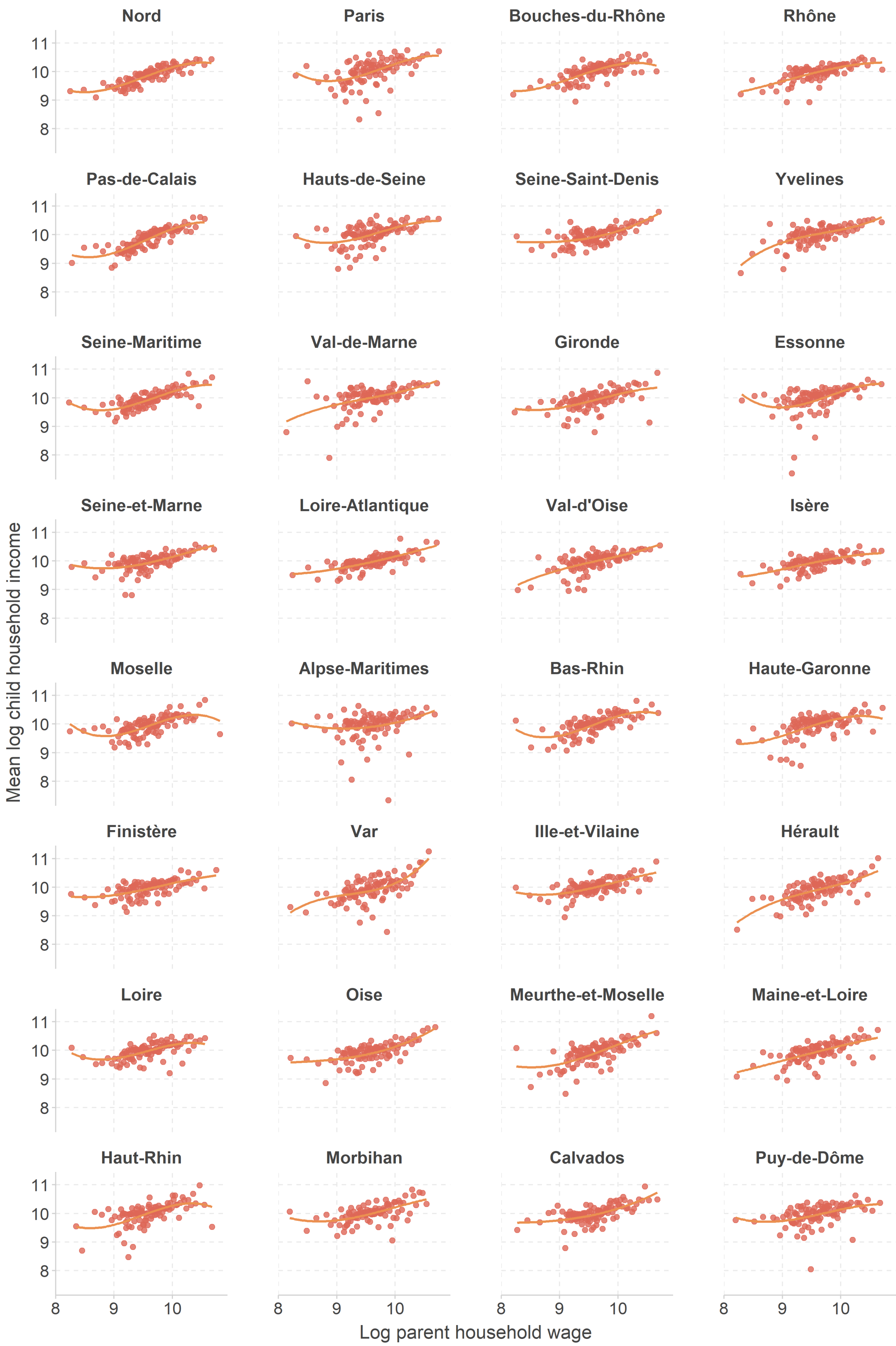
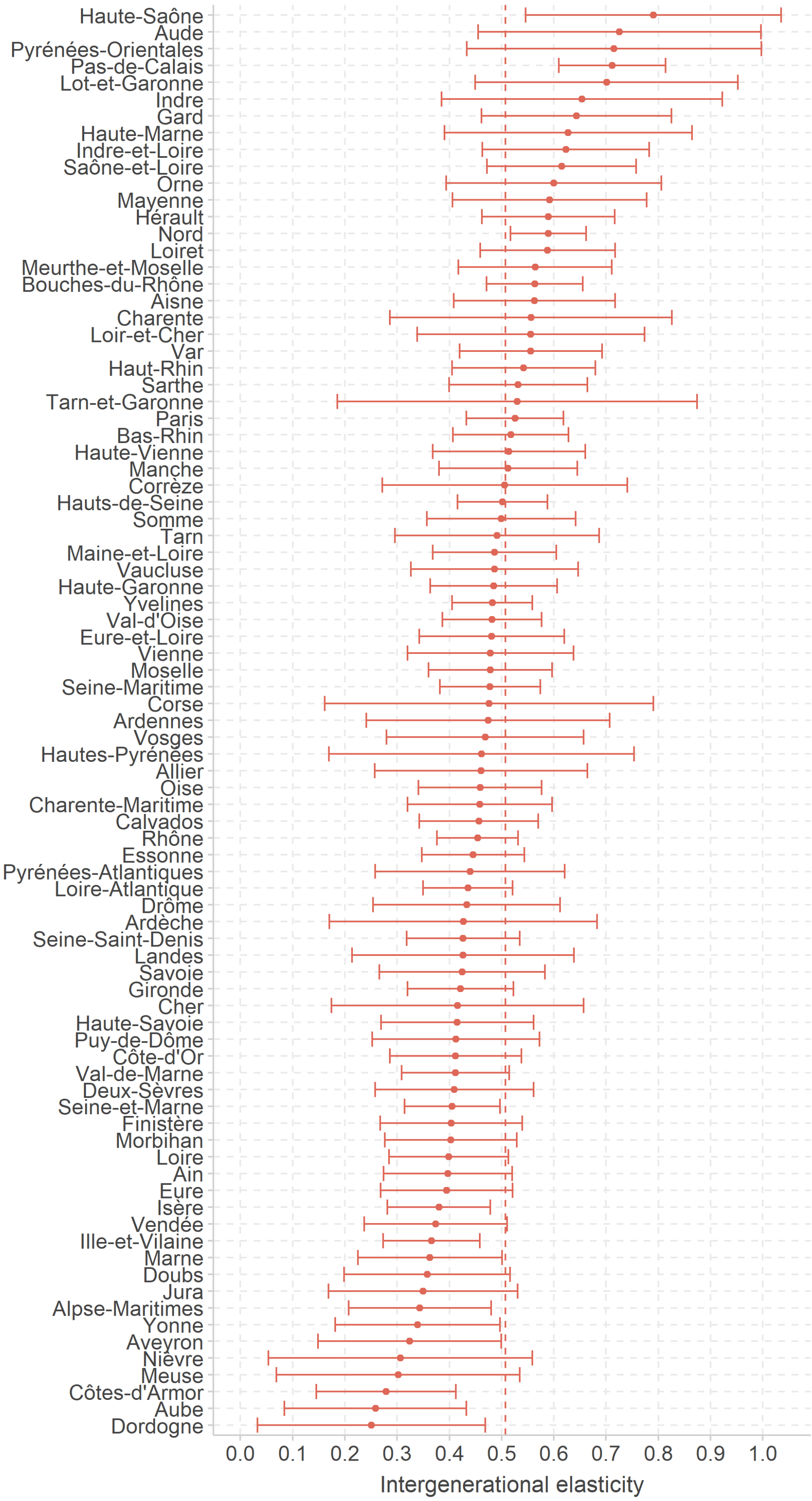


Figure A22: Department-Level Rank-Rank Relationships

Figure A23: Department-Level Intergenerational Persistence Estimates

(a) Department-Level Intergenerational Elasticities



(b) Department-Level Rank-Rank Correlations

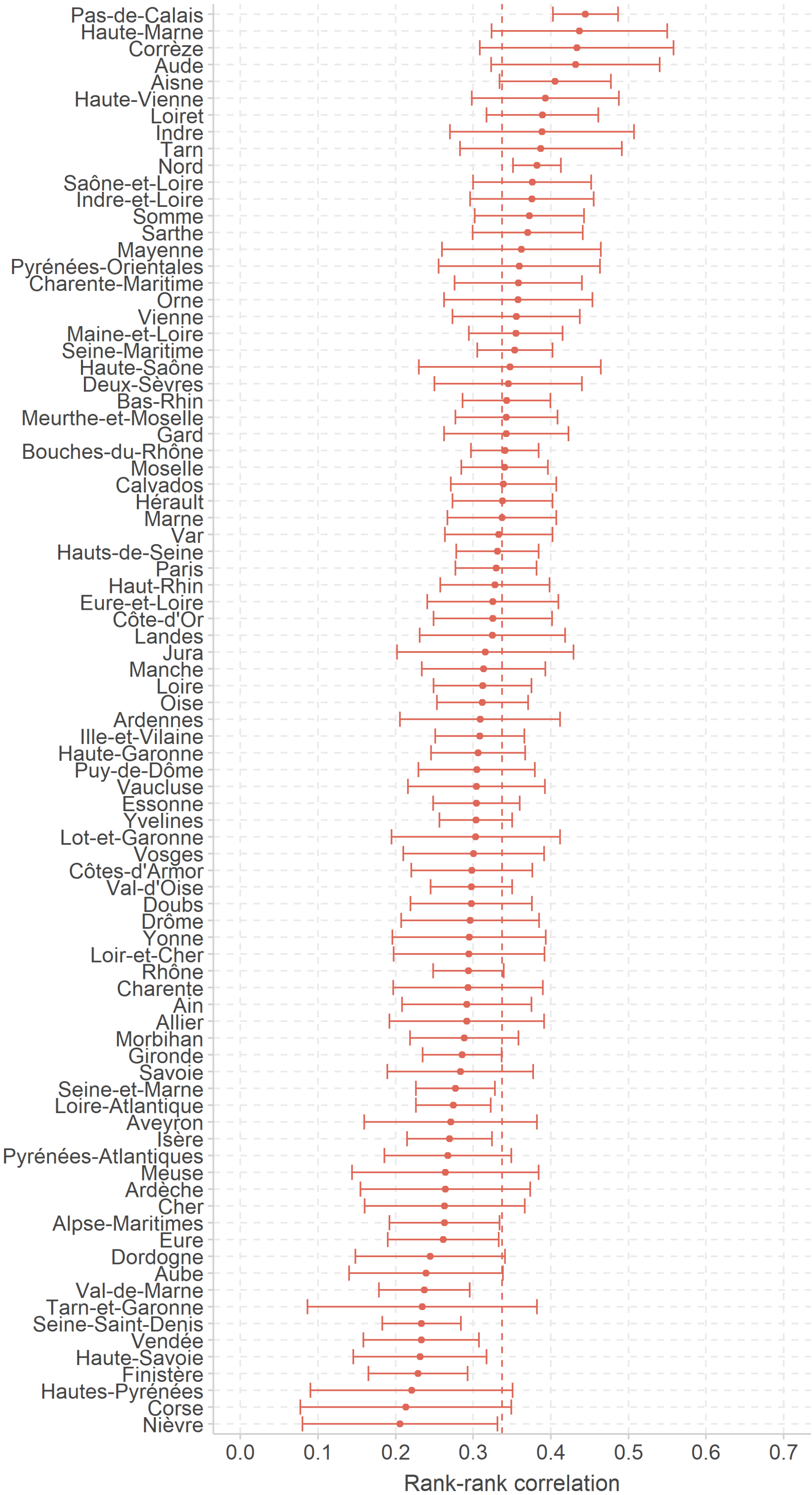
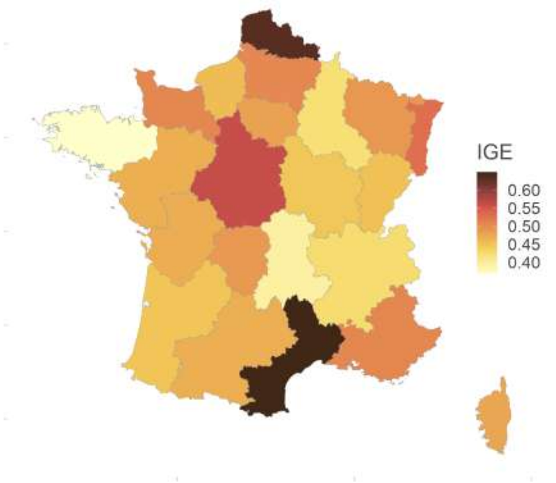
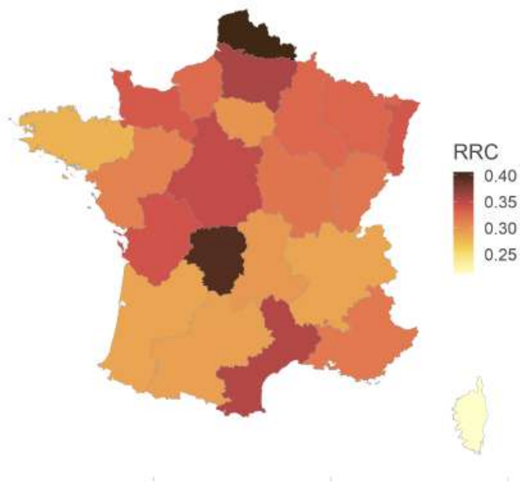


Figure A24: Variations in Intergenerational Mobility at the Regional Level

(a) Intergenerational Elasticity



(b) Rank-Rank Correlation



(c) Absolute Upward Mobility

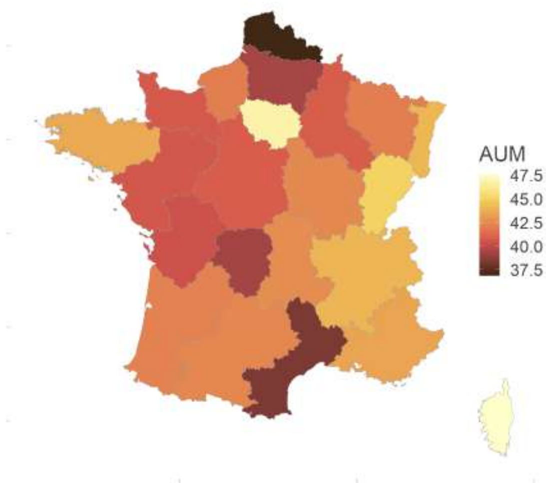
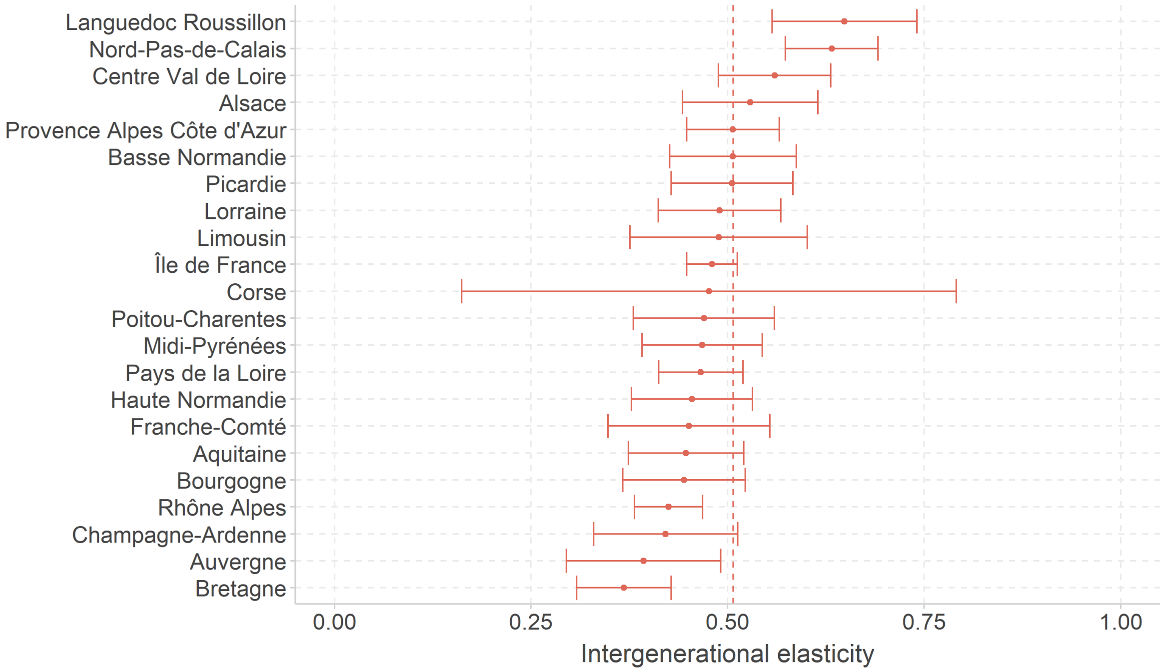


Figure A25: Region-Level Intergenerational Persistence Estimates

(a) Region-Level Intergenerational Elasticities



(b) Region-Level Rank-Rank Correlations

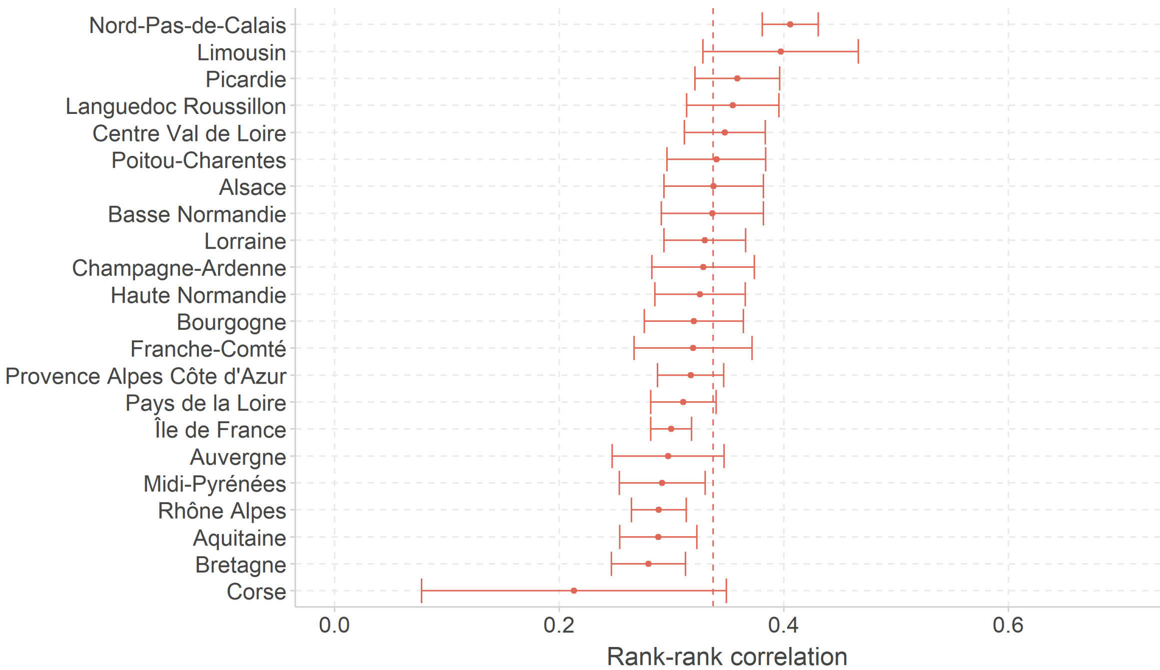
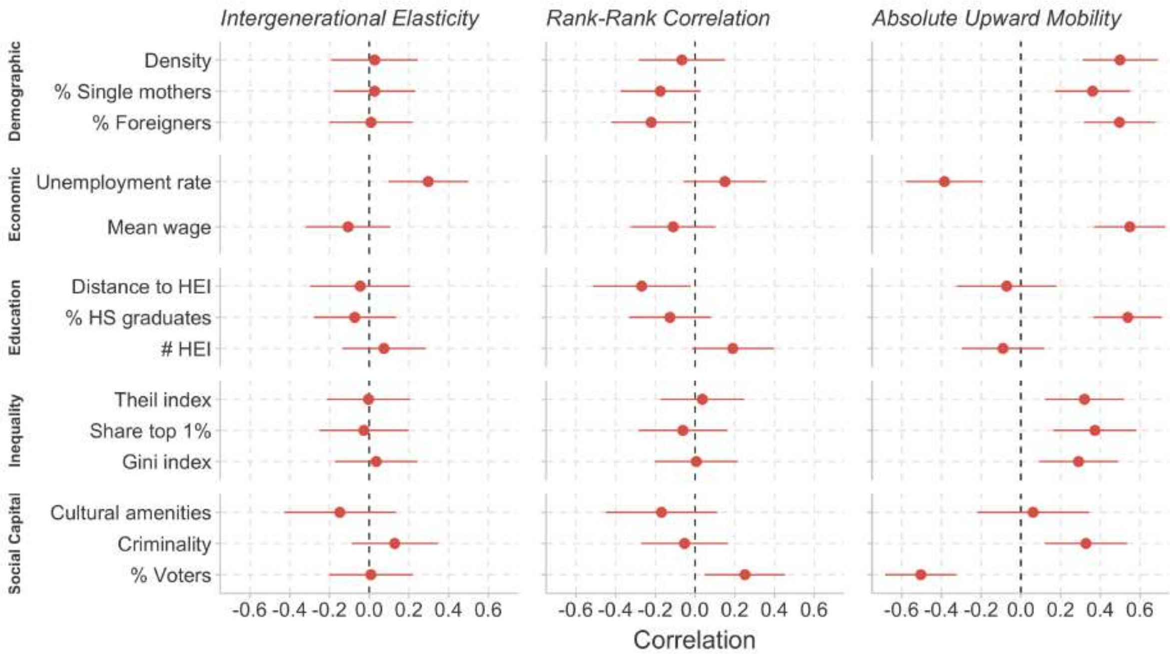


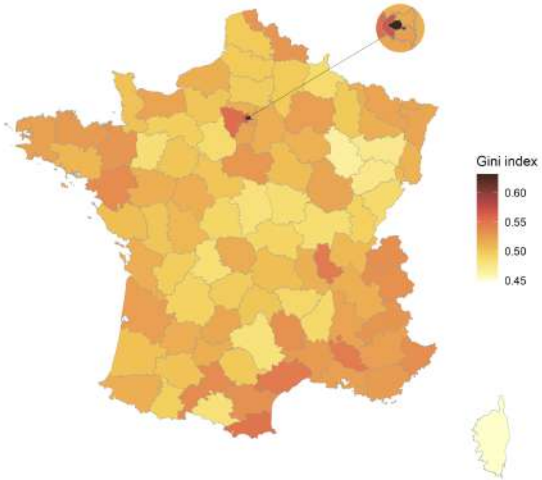
Figure A26: Intergenerational Mobility and Department Characteristics - Separate Estimation



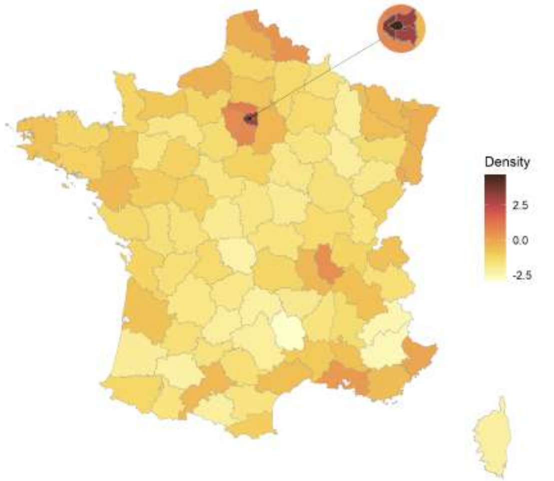
Notes: This figure presents the regression coefficient between department-level intergenerational mobility and department characteristics. Each coefficient is obtained from a separate regression. Both the department intergenerational mobility estimates and the characteristics are standardized, implying that the coefficients can be interpreted as correlations. See Figures 1 and 9's notes for details on data, sample and income definitions, and Appendix Table A10 for definitions and sources of the department characteristics.

Figure A27: Spatial Distribution of Average Department Characteristics

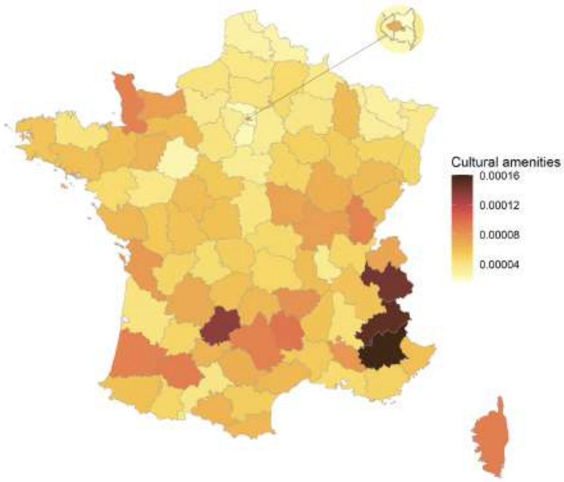
(a) Gini index of inequality



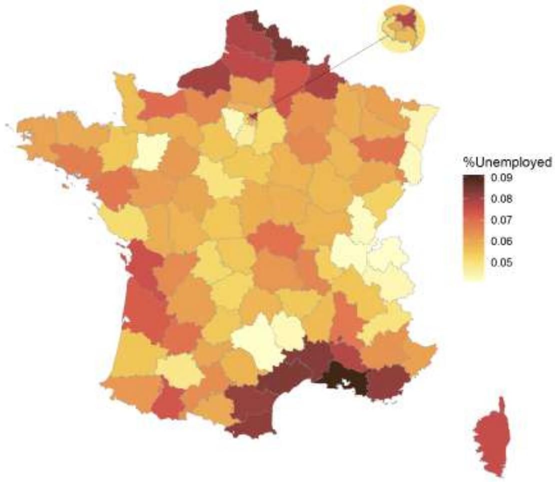
(b) Log population density



(c) Museums and cinemas per capita



(d) Unemployment rate



(e) Share of high school graduates

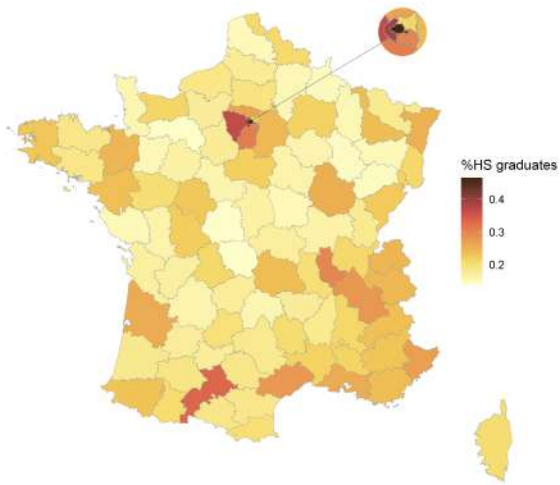
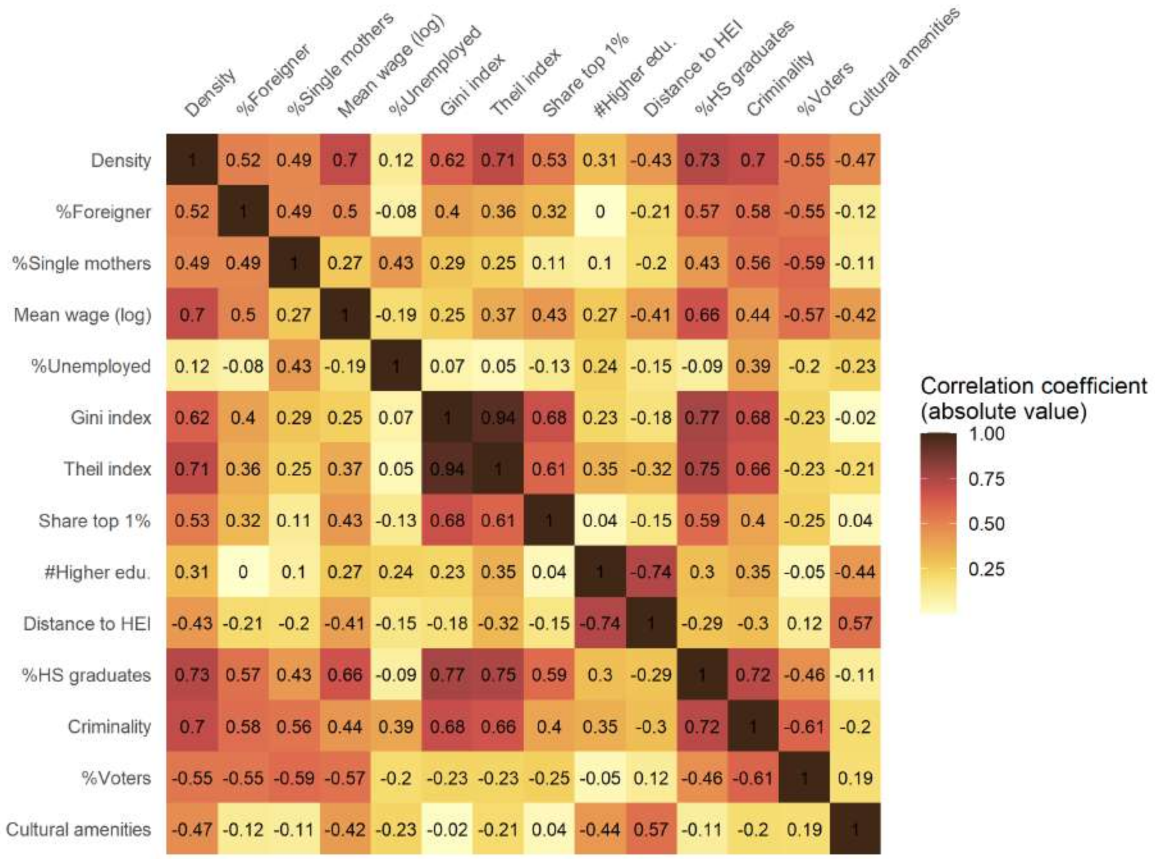


Figure A28: Correlation Between Department Characteristics



Notes: This figure presents the correlation coefficient between all department characteristics considered, defined as follows. See Appendix Table A10 for definitions and sources of the department characteristics.

Appendix Tables

Table A1: Prediction Model Estimated Separately for Synthetic Fathers and Mothers

	Synthetic Fathers	Synthetic Mothers
Age in 1990	-0.002^{**}	-0.010^{***}
	(0.001)	(0.001)
French nationality	0.097^{***}	0.096^{**}
	(0.027)	(0.043)
%Unemployment	-2.162^{***}	-2.409^{***}
	(0.224)	(0.321)
%Foreigners	1.014^{***}	0.511
	(0.227)	(0.324)
Density	0.041^{***}	0.065^{***}
	(0.005)	(0.007)
Population	-0.023^{***}	-0.019^{***}
	(0.005)	(0.007)
%Single mothers	-0.344^{*}	-0.530^{*}
	(0.192)	(0.276)
Maghreb	-0.054^{**}	0.008
	(0.024)	(0.036)
Other Africa	-0.190^{***}	-0.005
	(0.067)	(0.094)
Other Europe	0.047	0.086^{**}
	(0.035)	(0.057)
Rest of the world	-0.148^{***}	0.065
	(0.053)	(0.081)
South Europe	0.125^{***}	0.090
	(0.037)	(0.060)
Primary education	0.077^{***}	0.038^{*}
	(0.014)	(0.020)
BEPC	0.156^{***}	0.270^{***}
	(0.021)	(0.024)
CAP	0.126^{***}	0.204^{***}
	(0.013)	(0.021)
BEP	0.136^{***}	0.289^{***}
	(0.024)	(0.028)
High school diploma	0.245^{***}	0.364^{***}
	(0.018)	(0.024)
Bachelor or technical degree	0.253^{***}	0.460^{***}
	(0.023)	(0.031)
Masters or PhD	0.502^{***}	0.465^{***}
	(0.024)	(0.041)
Single mother	-0.327^{**}	-0.109
	(0.150)	(0.275)
Both spouses active	0.084^{***}	-0.080
	(0.029)	(0.274)
Mother inactive	0.122^{***}	-0.157
	(0.030)	(0.284)
Father inactive	-0.276^{**}	-0.171
	(0.107)	(0.279)
Both spouses inactive	-0.199^{*}	-0.024
	(0.102)	(0.293)
Tradesman	0.196^{***}	0.064
	(0.037)	(0.079)
Head of small firm	1.134^{***}	1.148^{***}
	(0.041)	(0.138)
Public service executive	0.565^{***}	1.155^{***}
	(0.045)	(0.104)

Professors and scientific professions	0.444***	1.077***
	(0.040)	(0.079)
Information, arts, and entertainment professions	−0.022	0.543***
	(0.067)	(0.125)
Administrative executives and sales representatives	0.891***	1.159***
	(0.031)	(0.078)
Engineers, technical company executives	0.838***	1.160***
	(0.031)	(0.129)
Teachers and related	0.332***	0.911***
	(0.041)	(0.069)
Intermediary health and social professions	0.315***	0.800***
	(0.045)	(0.070)
Clerk, religious	−0.532	.982
	(0.524)	(0.728)
Intermediary administrative professions of the public sector	0.428***	0.911***
	(0.042)	(0.076)
Intermediary administrative professions and salesmen	0.487***	0.786***
	(0.029)	(0.068)
Technicians	0.458***	0.859***
	(0.030)	(0.100)
Foremen, supervisors	0.521***	0.838***
	(0.029)	(0.111)
Civil servants	0.295***	0.686***
	(0.034)	(0.064)
Police and military officers	0.345***	0.428**
	(0.036)	(0.178)
Company administrative employee	0.385***	0.612***
	(0.036)	(0.064)
Trade employee	0.211***	0.268***
	(0.052)	(0.068)
Service workers	−0.030	−0.151**
	(0.057)	(0.067)
Skilled industrial workers	0.290***	0.643***
	(0.026)	(0.074)
Skilled crafts workers	0.070**	0.385***
	(0.027)	(0.092)
Drivers	0.138***	0.005
	(0.030)	(0.192)
Skilled handling, storing and transports workers	0.241***	0.514***
	(0.033)	(0.120)
Unskilled industrial workers	0.115***	0.434***
	(0.029)	(0.066)
Unskilled crafts workers	0.001	0.004
	(0.035)	(0.074)
Agricultural workers	−0.522***	0.132
	(0.079)	(0.141)
Former farmer	0.047	
	(0.256)	
Former craftsmen, trade workers, and heads of company	0.082	0.320
	(0.184)	(0.270)
Former executives	0.774***	
	(0.168)	
Former intermediary professions	0.683***	0.677**
	(0.141)	(0.259)
Former employees	0.648***	−0.292*
	(0.136)	(0.153)
Former workers	0.257**	0.430
	(0.110)	(0.427)

Unemployed who never had a job	-0.028	-0.251^{**}
	(0.370)	(0.114)
Military contingent	0.592^{***}	0.470^{***}
	(0.123)	(0.153)
Other inactive younger than 60	-0.016	-0.339^{***}
	(0.105)	(0.094)
Other inactive older than 59	0.318^{**}	-0.342
	(0.151)	(0.259)
Constant	9.684^{***}	9.347^{***}
	(0.068)	(0.292)
Observations	15,583	14,735
Adjusted R^2	0.356	0.370
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table A2: Child Sample Construction

Birth Cohort	Born in Metropolitan France	+ Observed as child in 1990 Census	+ At least one obs. in tax returns data	+ At least one obs. in tax returns data at 35-45	+ No parent in occupation 1 or 31
1972	9,083	7,946	7,582	7,582	7,077
1973	8,647	7,670	7,330	7,330	6,788
1974	8,704	7,713	7,372	7,372	6,830
1975	7,334	6,565	6,290	6,290	5,873
1976	7,762	6,963	6,662	6,650	6,199
1977	7,972	7,175	6,886	6,848	6,395
1978	7,755	7,000	6,748	6,677	6,224
1979	8,473	7,620	7,351	7,233	6,770
1980	8,822	7,965	7,642	7,426	6,945
1981	8,457	7,631	7,344	6,958	6,531
1972-1981	83,009	74,248	71,207	70,366	65,632

Table A3: Average Characteristics of Actual and Synthetic Parents

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%	23.80%
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	782.64	785.50
Log density	46.42	49.12
% foreigners	2.31%	2.33%
Unemployment rate	6.22%	6.26%
% single mothers	6.36%	6.40%
N	134,572	140,136

Notes: See Section 4.1 for details on construction of samples. These statistics are computed before applying any income observation restrictions.

Table A4: Share of Actual and Synthetic Parents by 2-Digit Occupation

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 A3's notes for sample construction.

Table A5: Synthetic Parents Sample Construction

Gender	At least one child born in Metrop. France 1972-1981	+ Observed in 1990 Census	+ Born even year	+ At least two obs. in All Employee Panel at 35-45	+ 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 A6: Number of Observations by Child and Parent Income Definitions

Child income definition	Parent income definition	Number of observations	Negative or 0 child incomes
Household income	Family income	64,572	42
Household income	Father income	58,435	37
Individual income	Family income	65,609	2,870
Individual income	Father income	59,355	2,525
Labor income	Family income	65,609	5,385
Labor income	Father income	59,355	4,792

Notes: The very slight discrepancy in the number of child income observations compared to those reported in Section 4.1 comes from the fact we code to missing 23 father ages in 1990 which were below 14 and above 70.

Table A7: Comparison with Existing Father-Son IGE Estimates for France

	Intergenerational Elasticity	First-Stage Instruments	Data	Income Definitions	Parent Age	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	**0.439**					

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 A8: Father-Son IGE in International Comparison

Country	Intergenerational Elasticity **Father-Son Labor Income ↓**	Income Definitions	Parent age	Child age	Source
Denmark	0.15				Corak (2016, Figure 1)
Norway	0.17				-
Finland	0.18				-
Canada	0.19				-
Sweden	0.27				-
Germany	0.32				-
France	**0.44**				
United States	0.47				Corak (2016, Figure 1)
Italy	0.5				-
United Kingdom	0.5				-

Table A9: Department-Level Intergenerational Mobility Estimates

	Department	Observations	IGE	RRC	AUM
01	Ain	535	.40	.29	46
02	Aisne	735	.56	.41	39
03	Allier	365	.46	.29	42
04	Alpes-de-Haute-Provence	141	*	*	*
05	Hautes-Alpes	112	*	*	*
06	Alpes-Maritimes	773	.34	.26	45
07	Ardèche	313	.43	.26	42
08	Ardennes	376	.47	.31	40
09	Ariège	121	*	*	*
10	Aube	361	.26	.24	42
11	Aude	274	.73	.43	37
12	Aveyron	243	.32	.27	44
13	Bouches-du-Rhône	1,795	.56	.34	45
14	Calvados	781	.46	.34	42
15	Cantal	164	*	*	*
16	Charente	374	.56	.29	39
17	Charente-Maritime	559	.46	.36	41
18	Cher	370	.42	.26	44
19	Corrèze	219	.51	.43	40
20	Corse	236	.48	.21	48
21	Côte-d'Or	549	.41	.33	43
22	Côtes d'Armor	590	.28	.30	44
23	Creuse	102	*	*	*
24	Dordogne	337	.25	.24	40
25	Doubs	635	.36	.30	48
26	Drôme	435	.43	.30	41
27	Eure	738	.39	.26	44
28	Eure-et-Loir	506	.48	.33	42
29	Finistère	979	.40	.23	45
30	Gard	577	.64	.34	40
31	Haute-Garonne	949	.49	.31	44
32	Gers	136	*	*	*
33	Gironde	1,304	.42	.29	43
34	Hérault	788	.59	.34	39
35	Ille-et-Vilaine	1,036	.37	.31	43
36	Indre	235	.65	.39	39
37	Indre-et-Loire	597	.62	.38	41
38	Isère	1,217	.38	.27	44
39	Jura	269	.35	.32	47
40	Landes	326	.43	.32	42
41	Loir-et-Cher	357	.56	.29	41
42	Loire	901	.40	.31	44
43	Haute-Loire	194	*	*	*
44	Loire-Atlantique	1,467	.44	.27	43
45	Loiret	706	.59	.39	40
46	Lot	137	*	*	*
47	Lot-et-Garonne	319	.70	.30	41

	Department	Observations	IGE	RRC	AUM
48	Lozère	63	*	*	*
49	Maine-et-Loire	931	.49	.35	40
50	Manche	566	.51	.31	43
51	Marne	676	.36	.34	43
52	Haute-Marne	263	.63	.44	40
53	Mayenne	329	.59	.36	39
54	Meurthe-et-Moselle	862	.56	.34	41
55	Meuse	238	.30	.26	44
56	Morbihan	778	.40	.29	44
57	Moselle	1,274	.48	.34	44
58	Nièvre	251	.31	.21	44
59	Nord	3,668	.59	.38	39
60	Oise	1,008	.46	.31	42
61	Orne	357	.60	.36	38
62	Pas-de-Calais	2,145	.71	.44	36
63	Puy-de-Dôme	664	.41	.30	44
64	Pyrénées-Atlantiques	571	.44	.27	46
65	Hautes-Pyrénées	209	.46	.22	42
66	Pyrénées-Orientales	356	.72	.36	38
67	Bas-Rhin	1,033	.52	.34	44
68	Haut-Rhin	792	.54	.33	46
69	Rhône	1,583	.45	.29	44
70	Haute-Saône	273	.79	.35	41
71	Saône-et-Loire	661	.61	.38	43
72	Sarthe	635	.53	.37	40
73	Savoie	430	.42	.28	45
74	Haute-Savoie	629	.41	.23	54
75	Paris	1,352	.53	.33	50
76	Seine-Maritime	1,547	.48	.35	42
77	Seine-et-Marne	1,529	.41	.28	45
78	Yvelines	1,645	.48	.30	48
79	Deux-Sèvres	376	.41	.35	42
80	Somme	737	.50	.37	39
81	Tarn	354	.49	.39	37
82	Tarn-et-Garonne	202	.53	.23	43
83	Var	773	.56	.33	41
84	Vaucluse	468	.49	.30	43
85	Vendée	627	.37	.23	42
86	Vienne	464	.48	.36	41
87	Haute-Vienne	357	.51	.39	39
88	Vosges	504	.47	.30	41
89	Yonne	388	.34	.29	42
90	Territoire de Belfort	172	*	*	*
91	Essonne	1,302	.45	.30	47
92	Hauts-de-Seine	1,248	.50	.33	47
93	Seine-St-Denis	1,495	.43	.23	47
94	Val-de-Marne	1,188	.41	.24	51
95	Val-D'Oise	1,366	.48	.30	46

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

Table A10: Definitions and Source of Department Characteristics

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 (≥ 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 (≥ 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.
[2] Base permanente des équipements (BPE) - 2007.

Table A11: Correlation Between Intergenerational Elasticity and Department Characteristics

	Dependent variable:													
	Intergenerational Elasticity													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	0.027													
	(0.111)													
% Single mothers		0.027												
		(0.105)												
% Foreigners			0.009											
			(0.107)											
Unemployment rate				0.298***										
				(0.102)										
Mean wage					−0.106									
					(0.109)									
Distance to HEI						−0.045								
						(0.129)								
% HS graduates							−0.073							
							(0.105)							
# HEI								0.074						
								(0.107)						
Theil index									−0.004					
									(0.108)					
Share top 1%										−0.027				
										(0.114)				
Gini index											0.036			
											(0.106)			
Cultural amenities												−0.148		
												(0.143)		
Crime													0.128	
													(0.111)	
% Voters														0.009
														(0.107)
Intercept	−0.003	0.0003	−0.0002	−0.018	0.009	−0.007	0.002	−0.006	0.0002	−0.0004	−0.0002	−0.026	−0.013	0.0005
	(0.110)	(0.109)	(0.109)	(0.104)	(0.109)	(0.111)	(0.109)	(0.109)	(0.109)	(0.109)	(0.109)	(0.111)	(0.109)	(0.109)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.001	0.001	0.0001	0.093	0.011	0.001	0.006	0.006	0.00001	0.001	0.001	0.013	0.016	0.0001

Note: *p<0.1; **p<0.05; ***p<0.01

Table A12: Correlation Between Rank-Rank Correlation and Department Characteristics

	Dependent variable:													
	Rank-Rank Correlation													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	−0.067													
	(0.111)													
% Single mothers		−0.175*												
		(0.103)												
% Foreigners			−0.221**											
			(0.104)											
Unemployment rate				0.150										
				(0.106)										
Mean wage					−0.110									
					(0.109)									
Distance to HEI						−0.269**								
						(0.125)								
% HS graduates							−0.126							
							(0.105)							
# HEI								0.190*						
								(0.105)						
Theil index									0.037					
									(0.107)					
Share top 1%										−0.062				
										(0.114)				
Gini index											0.006			
											(0.106)			
Cultural amenities												−0.169		
												(0.143)		
Crime													−0.053	
													(0.112)	
% Voters														0.251**
														(0.103)
Intercept	0.007	−0.002	0.006	−0.009	0.009	−0.041	0.004	−0.016	−0.002	−0.001	−0.00003	−0.029	0.005	0.013
	(0.110)	(0.107)	(0.106)	(0.108)	(0.109)	(0.108)	(0.108)	(0.107)	(0.109)	(0.109)	(0.109)	(0.111)	(0.110)	(0.106)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.004	0.034	0.052	0.024	0.012	0.052	0.017	0.038	0.001	0.004	0.00004	0.016	0.003	0.067

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A13: Correlation Between Absolute Upward Mobility and Department Characteristics

	Dependent variable:													
	Absolute Upward Mobility													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Density	0.499^{***}													
	(0.097)													
% Single mothers		0.361^{***}												
		(0.097)												
% Foreigners			0.496^{***}											
			(0.092)											
Unemployment rate				-0.385^{***}										
				(0.099)										
Mean wage					0.549^{***}									
					(0.092)									
Distance to HEI						-0.072								
						(0.128)								
% HS graduates							0.538^{***}							
							(0.088)							
# HEI								-0.090						
								(0.107)						
Theil index									0.320^{***}					
									(0.102)					
Share top 1%										0.373^{***}				
										(0.107)				
Gini index											0.291^{***}			
											(0.102)			
Cultural amenities												0.061		
												(0.144)		
Crime													0.327^{***}	
													(0.106)	
% Voters														-0.504^{***}
														(0.092)
Intercept	-0.053	0.004	-0.013	0.024	-0.046	-0.011	-0.017	0.007	-0.015	0.005	-0.001	0.011	-0.033	-0.027
	(0.096)	(0.101)	(0.094)	(0.100)	(0.092)	(0.111)	(0.091)	(0.109)	(0.103)	(0.102)	(0.104)	(0.112)	(0.104)	(0.094)
Observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85
R^2	0.242	0.144	0.261	0.155	0.301	0.004	0.312	0.009	0.107	0.128	0.090	0.002	0.103	0.268

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A14: Multivariate Correlation Between Intergenerational Mobility and Department Characteristics

	Dependent variable:		
	Intergenerational Elasticity	Rank-Rank Correlation	Absolute Upward Mobility
	(1)	(2)	(3)
Density	0.051	−0.136	0.498***
	(0.195)	(0.198)	(0.143)
Unemployment rate	0.253**	0.062	−0.277***
	(0.113)	(0.115)	(0.083)
Gini	0.111	0.287	−0.345**
	(0.181)	(0.183)	(0.132)
% HS graduates	−0.177	−0.274	0.451***
	(0.201)	(0.203)	(0.147)
Cultural amenities	−0.083	−0.249	0.313**
	(0.164)	(0.166)	(0.120)
Constant	−0.030	−0.025	0.006
	(0.109)	(0.110)	(0.080)
Observations	85	85	85
R^2	0.105	0.082	0.519
Note:			*p<0.1; **p<0.05; ***p<0.01