American Economic Journal: Applied Economics 2022, 14(3): 118–163
https://doi.org/10.1257/app.20210151

And Yet It Moves: Intergenerational Mobility in Italy†

By Paolo Acciari, Alberto Polo, and Giovanni L. Violante*

We estimate intergenerational income mobility in Italy using administrative data from tax returns. Our estimates of mobility are higher than prior work using survey data and indirect methods. The rank-rank slope of parent-child income is 0.22, compared to 0.18 in Denmark and 0.34 in the United States. The probability that a child reaches the top quintile of the national income distribution starting from a family in the bottom quintile is 0.11. We uncover substantial geographical variation: upward mobility is much stronger in northern Italy, where provinces have higher measured school quality, more stable families, and more favorable labor market conditions. (*JEL* D31, J31, J62, R23)

Income mobility across generations is a key socioeconomic indicator. It sheds light on the extent to which individuals with unequal initial conditions are offered equal opportunities to succeed, and, as such, it is considered a proxy for a fair and fluid society.

In spite of its centrality to the academic and policy debate, the body of empirical evidence on intergenerational mobility that economists have collected over the years is rather thin. The reason is that the data requirements are considerable. Very few publicly available datasets around the world have information that allows one to link parents and children and, at the same time, to construct reliable measures of permanent income for both cohorts.

Italy is no exception in this regard. So far, no study exists on intergenerational mobility on a national scale that uses high-quality data on incomes. Sociologists have filled this gap by studying intergenerational persistence of occupational classes (Pisati and Schizzerotto 2004). Economists have opted for a variety of other approaches. Checchi, Ichino, and Rustichini (1999) have documented the degree of persistence in educational attainment. A number of papers have used statistical procedures to impute incomes to parents of children who report their income in the Italian Survey of Household Income and Wealth (SHIW) or in the Italian component

*Acciari: Ministry of Economy and Finance of Italy (email: paolo.acciari@mef.gov.it); Polo: Bank of England (email: alberto.polo@bankofengland.co.uk); Violante: Princeton University (email: violante@princeton.edu). Camille Landais was coeditor for this article. The opinions expressed in this article are those of the authors and do not necessarily reflect those of their institutions. We are grateful to Alberto Alesina, Damien Capelle, Lorenzo Cappellari, Miles Corak, Nicola Curci, David Lee, Salvatore Lo Bello, Costas Meghir, Salvatore Morelli, Sevi Rodriguez-Mora, and Danny Yagan for useful comments. We thank Simon Boserup, Davide Malacrino, Francisco Meneses, and Enrico Moretti for sharing their data.

†Go to https://doi.org/10.1257/app.20210151 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

of the European Survey on Income and Living Conditions.[1] Finally, other authors have made creative use of surnames. Barone and Mocetti (2016) have focused on one particular city, Florence, and linked surnames of tax records in 1411 and 2011; Güell, Rodríguez Mora, and Telmer (2015) have documented large variation in intergenerational mobility across geographical areas within Italy by exploiting the informational content of surnames.

Over the last decade or so, the empirical literature on intergenerational income mobility has witnessed a strong revival thanks to the ability to access large administrative data in a handful of countries (e.g., United States, Canada, Australia, Denmark, Sweden, Norway). These very large datasets have led to precise estimates of mobility indices and have opened the possibility to analyze upward mobility patterns, within countries, at a very disaggregated geographical level. This variation has been exploited to learn what socioeconomic factors are strongly correlated with upward mobility across regions while controlling for the common institutional framework.

This paper adds to the recent wave of studies and introduces a new dataset that allows us to develop the first systematic investigation of intergenerational income mobility for the Italian economy. Our starting point is the administrative electronic database on individual tax returns from the Ministry of Economy and Finance.[2] From this data source, we extract a sample of children born between 1979 and 1983 and match them to their parents through their social security numbers. Our final dataset contains nearly 1,720,000 parents-child pairs with detailed income information for 3 years in each cohort, 1998, 1999, 2000 for parents and 2016, 2017, 2018 for children.

We begin from the analysis of intergenerational mobility at the national level. We estimate that, in Italy, a child born from parents with income below the median of the parental income distribution is expected to reach the forty-fifth percentile of her own income distribution as an adult. In other words, she is expected to move upward but to remain below the median. When we examine the full intergenerational income transition matrix across quantiles, we estimate that, for a child born from parents in the top quintile, the probability of keeping her parents' rank as an adult is 33 percent. For a child in the bottom quintile, the probability of rising to the top quintile is 11 percent. We also find that upward mobility is larger for sons, for firstborn children, for children of self-employed parents, and for children who, once adult, migrate to other regions within Italy.

We also estimate the relationship between the average rank of the child and average rank of the parents in their respective national distributions to learn about relative mobility patterns. We find that this relation is markedly linear, except at the very top, where it markedly bends upward. Its slope—the rank-rank slope (RRS)—is a measure of relative mobility for children with different initial conditions in terms of parental income. The estimated RRS is 0.22. To understand the meaning of this value, consider two children, one from parents in the top decile and one from parents in the bottom decile of the national distribution—a gap corresponding to

[1] Some of these studies are Mocetti (2007); Piraino (2007); Barbieri, Bloise, and Raitano (2018); and Cannari and D'Alessio (2018).

[2] The cross-sectional dimension of this database, i.e., without any intergenerational matches, is the source of the statistics on top incomes in Italy documented by Alvaredo and Pisano (2010) and contained in the World Inequality Database (www.wid.world).

a differential in their fathers' earnings of around €36,700. An RRS of 0.22 means that, when adults, these children will be on average still 2 deciles apart, a gap that translates into nearly €6,100 of annual earnings. The median rank-rank slope (i.e., the slope of the median, as opposed to the mean, rank of the child conditional on parental income) is 0.32 and thus much higher than the mean one. The discrepancy between mean and median is due to the fact that the conditional distributions of child ranks are skewed—to the right at the bottom of the income distribution of parents and to the left at the top. Remarkably, at the upper tail of the income distribution, the mean RRS is close to 1, which implies that rank differentials fully perpetuate a generation later.

Even though the expected rank of a child, conditional on parental rank, has a tightly estimated slope, the R^2 of the rank-rank regression is low. Conditional on a particular percentile of the parental income distribution, even controlling for all observable variables in our dataset, economic outcomes of children remain vastly different. For example, if we condition on children with parents in the ninetieth percentile, the bottom quarter of these children will be below the thirty-fifth percentile of their own national distribution. Among children from families at the tenth percentile, the top quarter of them will be above the sixtieth percentile.

For completeness, we also compute a more traditional measure of intergenerational mobility, the intergenerational elasticity of income (IGE). We estimate an IGE of 0.23. We uncover that this elasticity varies sharply across the distribution: at the bottom it is nearly 0, while above the tenth percentile it reaches 0.29. When we replicate statistical income imputation procedures used by previous studies on Italy to remedy the lack of parent-child matches, we obtain point estimates of the IGE around 0.5, hence much above its true value. We argue this approach leads to an upward bias in the IGE because the instruments used to impute father's income are correlated with child income.

The two main shortcomings of our data are (i) life cycle and attenuation biases due to the short within-individual panel dimension and (ii) possible distortions arising from tax evasion for the self-employed. Correcting for these biases increases somewhat our estimates of intergenerational rank persistence; e.g., the RRS and IGE rise to 0.3. Measures of absolute upward mobility remain more stable, though. For example, the mean rank of a child born from parents with income below the median falls only slightly from 0.45 to 0.43, and the probability that a child reaches the top quintile starting from a family in the bottom quintile decreases from 0.11 to 0.09. Overall, even after these corrections, Italy emerges as less immobile than how it was depicted in previous studies that did not have access to the same high-quality data as we do.

When placing our estimates of positional income mobility in a comparative context, upward mobility in Italy appears higher than in the United States but lower than in Scandinavia. We also compute an alternative measure of mobility—namely, the probability that a son earns at least as much as his father in real terms—used recently by Chetty et al. (2017a) and Berman (2020), which allows to isolate the role of differential income growth and differential income inequality when comparing mobility patterns across countries. Italy and the United States have similar shares of sons who overtake their fathers in terms of income (0.53 and 0.55). This similarity, however, is the result of two strong but exactly offsetting forces: lower income growth and less dispersed income distribution in Italy relative to the United States.

Next, we explore the geographical differences in upward mobility across the 110 Italian provinces. We document a staggering amount of variation, with a steep South–North gradient. Relative to the South of Italy, provinces in the North (especially in the Northeast), are both more egalitarian—i.e., they display higher relative mobility—and more upward mobile—i.e., they display higher absolute mobility (as measured, for example, by the expected rank of a child born from parents below the median). The level of upward mobility in northern Italy exceeds that of Scandinavia and that of the most mobile cities in the United States (e.g., Salt Lake City and Pittsburgh), whereas in southern Italy it is comparable to that of the least mobile cities in the United States (e.g., Atlanta or Charlotte).

We uncover a Great Gatsby curve with a negative slope linking upward mobility and several measures of income inequality across Italian provinces. However, surprisingly, the top income share correlates positively. One interpretation is that the top income share is high in areas where self-employment is prevalent and upward mobility is especially strong for this group.

Our dataset also allows us to assess, for the first time, the relationship between Informational Content of Surnames (ICS) indicators (such as those estimated by Güell, Rodríguez Mora, and Telmer (2015) for the Italian provinces) and true measures of mobility. As one would expect, we find a significant, negative correlation between the two. However, we also show that this relationship weakens considerably—becoming almost flat—for provinces that display ICS indices below the mean. Thus, when the ICS is low, it contains little information about the true extent of mobility, suggesting that researchers should be cautious when using such proxy.

We then investigate which socioeconomic indicators correlate, at the provincial level, with upward mobility. We use nearly 50 markers for productivity, labor market conditions, demographic structure, educational attainment, family instability, crime, and economic openness from ISTAT, the National Statistical Institute. In addition, we have several measures of social capital and a unique and very detailed set of indicators of school quality.

Most of these variables correlate with upward mobility with the expected sign. A limitation of this unconditional analysis is that all these socioeconomic variables are also highly correlated among each other. We therefore proceed with a multivariate conditional correlation analysis where we extract a small number of principal components for each broad category to collapse the number of covariates. Overall, the included categories explain nearly 90 percent of the geographic variation in rates of upward mobility. The key explanatory variables are the local labor market conditions, indicators of family instability, and three specific indices of school quality: quality of early childhood education, quality of school organization and services, and students' grades and test scores.

The rest of the paper is organized as follows. Section I defines the measures of intergenerational mobility used in the analysis. Section II describes the dataset, outlines the sample selection procedure, and provides some descriptive statistics. Section III discusses our findings on the degree of intergenerational mobility at the national level. Section IV tackles potential sources of bias in our baseline estimates. Section V compares mobility outcomes in Italy to those in other countries. Section VI documents the patterns of geographical variation in upward mobility across provinces in Italy. Section VII explores the correlation between upward mobility and local characteristics of provinces that could account for geographical variation. Section VIII concludes.

I. Measures of Intergenerational Income Mobility

In this section, we briefly discuss the measures of intergenerational income mobility we use throughout the paper. No single measure is perfect. Each one has advantages and shortcomings over the others, and each one answers a specific question.

A. *Relative Mobility*

Relative mobility is the subject of most prior investigations of intergenerational mobility (see Solon 1999; Black and Devereux 2011; Jäntti and Jenkins 2015 for surveys of the literature). These studies focus on relative outcomes of children with different family backgrounds, and ask "What is the expected income of children of low-income families *relative* to those of high-income families?"

Our main measure of relative mobility is the correlation between child and parental income ranks (Dahl and DeLeire 2008; Chetty et al. 2014), i.e., an index of positional mobility. Let R_i denote child i's percentile rank in the income distribution of children (from 1 to 100) and R_i^P denote the percentile rank of i's parents in the income distribution of parents. A linear regression of child rank on parental rank yields

$$R_i = \alpha + \beta R_i^P + \varepsilon_i, \tag{1}$$

where the constant α measures the expected rank of a child born from parents at the bottom of the income distribution $\left(R_i^P = 0\right)$ and the rank-rank slope (RRS, or rank-rank persistence coefficient) β measures the strength of the correlation between a child's position and her parents' position.

By construction, this regression on national data has only one free parameter since it must be true, by taking averages of both sides of (1), that $50 \cdot (1 - \beta) = \alpha$. Values of β close to zero denote a very mobile society where the expected rank of children is always around the median independently of parental rank. Values close to one depict a society with high persistence in relative positions across generations. Thus, high relative mobility corresponds to a low value for β.[3]

[3] By computing $\Delta \cdot \beta$, we can answer the question "What is the difference in expected rank between two children with parents who are Δ percentiles apart in the national income distribution?" And by simple iteration, we can ask how many generations it would take, on average, for descendants of families originally Δ percentiles apart to belong to the same percentile of the income distribution, i.e., the value N that solves $\beta^N \Delta = 1$. This back-of-the-envelope calculation requires the assumption that permanent income across generations follows an AR(1) process. Existing empirical work on multiple generations finds a correlation between outcomes of children, parents, and grandparents that is higher than what one would expect under the AR(1) assumption (Braun and Stuhler 2017; Lindahl et al. 2015). Thus, this calculation might be a lower bound.

We are also interested in assessing whether mobility at the top differs from mobility in the rest of the distribution. For example, we may think that belonging to the upper class of society yields *disproportionately* better opportunities to perpetuate social status across generations. For this purpose, we construct an index of relative mobility at the top, or top mobility ratio (TMR), as follows. First we compute the RRS by running the rank-rank regression (1) on the top decile of the parental distribution $\left(\beta^{91-100}\right)$. Next, we run it on the bottom 90 percent and obtain β^{1-90}. We then define

$$(2) \qquad TMR = \frac{\beta^{91-100}}{\beta^{1-90}}.$$

The higher this ratio, the stronger the persistence in ranks across generations at the top of the income distribution relative to the rest of the distribution.

The most commonly used index of relative mobility in the literature is the intergenerational income elasticity (IGE), which captures the elasticity of child income with respect to parental income and is estimated as the OLS coefficient of a linear regression of log child income y_i on log parental income y_i^P, i.e.,

$$(3) \qquad IGE = ICC \times \frac{SD\left(\log y_i\right)}{SD\left(\log y_i^P\right)},$$

where *ICC* is the intergenerational correlation coefficient between log income of parent and child and *SD* is the standard deviation. An IGE of 0.5, for example, means that a 20 percent differential in parental income translates into a 10 percent differential in child income. Mazumder (2016) discusses the relation between RRS and IGE.[4]

There are two main advantages of the RRS compared to the IGE. First, the IGE is based on log income. As a result, one has to either drop the zeros in income or use an imputation procedure. Conclusions can be sensitive to selection and imputation assumptions. Second, the RRS can be used to measure mobility differentials among subgroups of the population (e.g., geographical areas) because the RRS for different groups can be estimated based on ranks of the same national distribution.[5]

B. *Absolute Mobility*

Absolute mobility indices measure the outcomes of children from families at a given income or rank in the parental income distribution. They are typically used to study the economic performance of children from poor families.

[4] RRS and ICC are closely related, scale-invariant measures of the extent to which child income depends on parental income. Theoretically, the IGE differs from the RRS only if income inequality changes significantly across generations: if, for example, $SD\left(\log y_i\right) > SD\left(\log y_i^P\right)$, then the effect of parental income on child income is larger, and this is reflected in a higher IGE.

[5] The IGE estimated within groups is, instead, only informative about persistence or mobility with respect to the group-specific mean, not the aggregate mean.

We follow Chetty et al. (2014) and report measures of absolute upward mobility (AUM), defined as the mean rank (in the national child income distribution) of children whose parents are below the median of their own national income distribution, or

$$AUM = \mathbb{E}\left[R_i \,|\, R_i^P \leq 50\right]. \tag{4}$$

When the rank-rank relationship is linear, the average rank of children with below-median parental income equals the average rank of children with parents at the twenty-fifth percentile of the national income distribution, i.e., $\mathbb{E}\left[R_i \,|\, R_i^P = 25\right]$, which can be easily computed from (1) as $\alpha + 25 \cdot \beta = 50 - 25 \cdot \beta$. This is how we compute our AUM indicator at the national level.

Another measure of absolute mobility we analyze is the probability of rising from the lowest to the highest quintile of the income distribution (Corak and Heisz 1999) —we call it Q1Q5. This probability can be interpreted as the fraction of those who make it to the top starting from the bottom:

$$Q1Q5 = \Pr\left\{R_i > 80 \,|\, R_i^P \leq 20\right\}. \tag{5}$$

Finally, we provide nonparametric transition matrices by percentile and marginal distributions that allow readers to construct alternative measures of mobility beyond those we document here.

C. *Indicators for Within-Country Comparisons across Regions*

We are also interested in analyzing the geographical heterogeneity in intergenerational mobility within Italy. Let R_{ig} denote the rank in the national income distribution of children for a child i who grew up in geographical area (or region) g. Let R_i^P denote its parents' rank in the national distribution of parental income. By running regressions of the type

$$R_{ig} = \alpha_g + \beta_g R_i^P + \varepsilon_{ig}, \tag{6}$$

where we continue to rank both children and parents based on their positions in the national income distribution (rather than the distribution within their region), we obtain estimates of the region-specific indicators of relative (β_g) and absolute (AUM_g) mobility.

It is important to note that, while in the linear national rank-rank relation (1) there is only one free parameter, and a one-to-one mapping between AUM and RRS, this is no longer the case for the regional regressions since the average national rank of residents of a specific region need not be the median. As long as linearity also holds at the regional level, the AUM for region g can still be accurately approximated as

$$AUM_g = \alpha_g + 25 \cdot \beta_g. \tag{7}$$

This is the measure of absolute upward mobility that we use throughout our regional analysis. Using directly the twenty-fifth percentile of the national income distribution avoids the problem that in poor provinces the income distribution can be much more concentrated around lower values than in rich ones, thus affecting local estimates of the AUM.

D. *Indicators for Cross-Country Comparisons*

The measure of mobility that we use in cross-country comparisons is the probability that children earn at least as much as their parents in real terms at their same age (e.g., Chetty et al. 2017a; Berman 2020):

$$\Pi = \int\int \mathbf{1}\{y(R) \geq y^P(R^P)\} C(R, R^P)\, dR dR^P. \tag{8}$$

The indicator function in this formula equals 1 if the Rth quantile of the child's real income distribution $y(R)$ is higher than or equal to the R^Pth quantile of the parent's distribution $y^P(R^P)$. $C(R, R^P)$ is the copula giving the joint distribution of the pair of ranks (R, R^P).

Equation (8) illustrates that differences in Π between countries can be due to differences in either the copula $C(R, R^P)$ or the marginal income distributions of children and parents, encoded in $y(R)$ and $y^P(R^P)$. In turn, these marginal distributions could diverge between countries because of differential income growth or differential income inequality.

To disentangle these three effects, when we compare Italy to another country (specifically, the United States), we run three counterfactuals in the spirit of Chetty et al. (2017a) and Berman (2020). In the first counterfactual, we compute (8) using the US copula $C^{US}(R, R^P)$; i.e., we fix the joint distribution of quantiles for parents and children. In the second counterfactual, we compute Π imposing average income growth for the United States. Namely, let

$$\tilde{y}(R) = s(R)\, Y^P G^{US},$$

where $s(R)$ is the share of aggregate real income of Italian children earned by those at rank R of their distribution, Y^P is aggregate income of the Italian parents, and G^{US} is the US income per capita growth factor between the parent and the children generations. Thus, $\tilde{y}(R)$ gives us the counterfactual income of Italian children at each quantile R. Finally, in the third counterfactual, we impose the income distribution of the United States without shifting its mean. Letting $s^{US}(R)$ and $s^{US}(R^P)$ be the shares of aggregate real income earned by children and parents in the United States at each percentile of the respective distributions, we obtain

$$\tilde{y}(R) = s^{US}(R) Y \quad \text{and} \quad \tilde{y}(R^P) = s^{US}(R^P)\, Y^P$$

as the counterfactual income levels at each quantile that match the US income distribution for children and parents, respectively.

II. Data

In this section, we describe our dataset, outline our sample selection procedure, and present some descriptive statistics.

A. *Description of the Dataset*

Our data source is the electronic database of Personal Income Tax returns assembled by the Department of Finance of the Italian Ministry of Economy and Finance.[6] The database is used for the official tax return statistics published annually and for economic analysis supporting policy decisions. It is also the source of cross-sectional statistics on income inequality for Italy in the World Income Database (Alvaredo and Pisano 2010). It is the first time, however, that the dataset is used to link children to their parents in order to construct measures of intergenerational income mobility. As of today, no other existing sources of income data allowed such link in Italy.

The database combines information from all three income tax forms available to Italian taxpayers: (i) form Modello Unico (MU), which is the most common; (ii) form 730, the simplified income tax form available to employees and pensioners whose income consists of only few items; and (iii) form 770, which is compiled by the withholding agent of employees, pensioners, and self-employed workers and is accepted by the Italian tax authorities when the taxpayer has only one source of income, no other dwellings than their primary home, and no itemized deductions (e.g., medical expenses, charitable donations, mortgage interests).[7]

Our extract comprises numerous variables on demographic characteristics and income. Demographics include the province of birth and residence, birth year, marital status, and an occupation/sector identifier (ATECO code).[8] The income variables include total gross (before-tax) income and all its components, i.e., income from dependent labor, self-employment (divided into professional and entrepreneurial income), unemployment benefits, financial assets, housing, land, and farms.[9,10] Finally, we also have information on individual tax liabilities.

[6] In Italy the tax unit is the individual. The accuracy of the information in this database is preliminarily statistically checked and validated by the Department of Finance. This process is mainly performed by SOGEI SpA, an in-house company of the Ministry of Economy and Finance, through a series of algorithms that check the coherence between data reported in different sections of tax returns, correct abnormally high values, etc. In 2012, 100 percent of tax returns were transmitted electronically to the tax authorities.

[7] Standard deductions such as allowances for children and dependent spouses are applied by the withholding agent.

[8] The ATECO code is available only for self-employed individuals who are required to report it in the tax return. The code describes the specific economic activity that generates the income.

[9] Some income sources we do not observe are fellowships/scholarships, child/family benefits, some forms of financial income, and social assistance transfers.

[10] The difference between professional and entrepreneurial income is subtle. The key distinction is the extent to which the business uses physical capital beyond human capital. For example, a lawyer or a freelance journalist is considered, for tax purposes, a professional. An owner of a firm that produces clothes is considered an entrepreneur. Entrepreneurial income also includes income from privately owned businesses, i.e., firms where individuals own a share of total private equity (e.g., partnerships, closely held companies).

B. *Sample Construction*

Linking parents and children in our database is possible starting from 1998. This is the first year in which, in order to claim deductions for dependent children, parents must report the child's Social Security Number (SSN) on their own tax return. Figure A1 in the online Appendix shows the corresponding section of the tax form.[11]

Data for parents are extracted from tax returns of years 1998, 1999, and 2000 by selecting all taxpayers who claim allowances for children born between 1979 and 1983 (age 15–19 in 1998). Then, through children's SSNs, we recover their tax returns when they are adult in years 2016, 2017, and 2018 (the last year available). Each record in our dataset contains information on a child, their father, and their mother.[12] Our initial dataset comprises of 1,726,141 records. We then drop observations where the age differential between mother (father) and child is above 45 (50) or below 14 years to correct for possible reporting errors taxpayers might make (they have to check specific boxes) when they indicate whether the dependent is a child or another relative. Through this step, we only lose 6,658 records, but we are certain that we are identifying parents-child matches. The final sample size is 1,719,483 matched records.

By focusing on the age range 15–19 for dependent children, we are sure to catch the vast majority of individuals in this cohort. According to the main Italian household survey, the Survey on Household Income and Wealth, 93 percent of children between ages 15 and 19 were dependent in 1998. Children who were 15–19 years old in 1998 are 34–38 years old when we observe their income roughly two decades later. As a result, we can also be sure that they already have some labor market experience, even in a country like Italy where entry in the labor market is much delayed relative to the United States.[13]

Sources of Bias.—The matching procedure outlined above misses some parent-child pairs for two reasons that are specific to our dataset. First, starting from 1998, the child's SSN was mandatory for forms MU and 730 but not for form 770. For those who filed form 770 (around 25 percent of taxpayers, typically low-income ones) a match is not possible. Second, certain individuals are not required to submit any tax form. These are individuals (i) with no income whatsoever; (ii) with only tax-exempted income such as social assistance payments, social transfers in kind, or fellowships; (iii) with only rental income from housing and land below

[11] Specifically, parents-children relationships can be identified because a taxpayer must indicate on the form the name and SSN of the spouse and the SSN and relationship for each of the dependents for whom a deduction is claimed.

[12] Even if spouses are separated or divorced and live apart, we can retrace couples when they both claim a positive percentage of deduction for the same individual SSN. If only one adult claims a 100 percent deduction and there is no information about the spouse on the tax return, we conclude that that taxpayer is a single parent.

[13] By comparison, Chetty et al. (2014) select essentially the same birth cohorts (1980–1982). They observe parental earnings in 1996–2000, again approximately at our same point, and children earnings in 2011–2012 when the children are 30–32 years old, hence roughly 5 years earlier than we do. Observing child income at that age would have been extremely problematic for us because in Italy a large share of youths at that age is not yet well integrated in the labor market, which would have led to a selected income sample.

€500; (iv) with financial income exclusively from interest, dividends, and capital gains. Because of (i–iii), the very poor are not covered in our sample.[14]

Our sample construction is also potentially subject to two sources of bias that are prevalent in administrative data and discussed at length in the literature (Solon 1992; Haider and Solon 2006): (i) attenuation bias because our proxy of lifetime income for both parents and children is a 3-year average, and thus it measures permanent income with error; (ii) life cycle bias because income for fathers and children is observed at ages that are 12 years apart, and the elasticities of current income to lifetime income at these two ages are likely to differ.

Finally, Italy is known to have significant tax evasion. Because its incidence is not uniform across income levels, occupations, and geographical areas, observed income ranks may be different from the true ones, and measures of positional mobility may be affected.

In Section IV, we bring in additional data to deal with all these sources of bias.

C. *Descriptive Statistics*

Table 1 contains selected descriptive statistics about parents and children in the baseline sample.

On average, fathers (mothers) are 48 (45) years old in 1998. Over 90 percent of the fathers are married. The data show that in 1998, the father is the top earner in nearly 90 percent of the families in the sample, and around 60 percent of families have two sources of income. The median gross income of fathers is nearly €19,000 in 1998, and that of mothers is approximately half of it.[15] The correlation between the two (excluding zeros) is positive and significant, around 0.25.

Children are on average 36 years old in 2017. Median nominal income of sons is only about 20 percent higher than that of their fathers nearly 20 years earlier. This modest growth partially reflects the age gap between the two groups, but it is mainly due to the dismal aggregate growth of the Italian economy over this period and the fact that growth, as in many other developed countries, mostly accrued to the top earners, so it escaped median income.[16] However, daughters' income in 2017 is 70 percent higher than their mothers', reflecting the steady rise in the female employment rate among more recent cohorts. The composition of income reveals that, for nearly 77 percent of children, compensation for dependent labor is the major component, followed by self-employment income for 17 percent of the child population. Fewer than 6 percent are rentiers—their main source of income being financial and housing capital.

Additional statistics on the income distributions and individual characteristics are presented in Table A1 in the online Appendix. The table shows that, as expected, the

[14] Individuals in group (iv) are very few because it is highly unlikely that someone in that group and in our age classes would not have other types of taxable income.

[15] Italy adopted the euro in 2002; thus, incomes for 1998, 1999, and 2000 are expressed in Italian liras in the database. We transformed them in euros, the currency in which they are reported in 2016, 2017, and 2018.

[16] Over 1998–2017, average annual nominal GDP growth per capita in Italy was 2 percent but 0 in real terms.

Table 1

Statistic	Observations	Value
Parents in 1998		
Mean father's age	1,600,529	48
Mean mother's age	1,196,230	45
Percentage of married fathers	1,577,095	92.8%
Percentage of fathers residing in the North	1,600,529	42.8%
Percentage of fathers residing in the Center	1,600,529	19.6%
Percentage of fathers residing in the South	1,600,529	37.6%
Percentage of fathers residing in same region as they were born	1,600,529	81.0%
Percentage of fathers residing in same province as they were born	1,600,529	73.5%
Percentage of fathers born abroad	1,600,529	2.0%
Percentage of families where top earner = father	1,600,529	87.1%
Percentage of families with two positive incomes	1,719,483	60.9%
Correlation father-mother income (both positive)	1,046,607	0.2524
Median total parental income	1,719,483	23,173
Median total father's income	1,600,529	18,628
Median total mother's income	1,196,230	10,188
Children in 2017		
Mean age	1,719,483	36
Percentage of females	1,719,483	48.4%
Percentage residing in the North	1,635,680	48.5%
Percentage residing in the Center	1,635,680	21.5%
Percentage residing in the South	1,635,680	30.0%
Percentage born abroad	1,716,255	1.6%
Median total income	1,635,680	19,641
Median total income males	850,769	22,215
Median total income females	784,911	16,802
Percentage of individuals whose major income component is:		
Wage	1,635,680	76.5%
Entrepreneurship	1,635,680	8.3%
Other self-employment	1,635,680	9.0%
Capital	1,635,680	5.3%

Notes: Descriptive statistics for the final sample. Income is nominal and expressed in euros. See Table A1 in the online Appendix for more detail on the income distribution and its components.

income distribution is markedly right skewed: the top 1 percent of the distribution accounts for 6.5 percent of total income in 2017.[17]

D. *Comparison with Survey Data*

To gain more confidence in the reliability of our data, we verify that the distribution of labor earnings we obtain in the final dataset is consistent with that from survey data.

The best source for comparison is the Survey on Household Income and Wealth, administered every two years by the Bank of Italy. In SHIW, we keep individuals aged 34–38 in 2016 (there is no survey in 2017) in order to compare them to children in our data, and men and women older than 31 in 2000 to compare them to

[17] In the Italian population, without the age restrictions we impose, this share is close to 9 percent in 2017. For comparison, in the United States, the top 1 percent income share is around 19 percent in the same year.

Table 2

Net wage	Mean	SD	p99	p95	p90	p75	p50	p25	p10	p5	p1
Children											
SHIW (2016)	15,234	6,713	34,000	26,000	23,500	19,000	15,000	11,500	7,000	4,200	0
Tax return (2017)	15,085	10,061	41,204	28,419	24,446	19,838	15,486	9,190	3,225	773	0
Fathers											
SHIW (2000)	14,362	6,643	38,734	24,015	20,658	16,527	13,428	11,817	7,747	4,132	0
Tax return (2000)	16,970	8,191	45,030	29,172	24,635	19,881	16,244	13,302	8,932	5,000	0
Mothers											
SHIW (2000)	11,479	5,440	26,339	18,592	17,043	14,461	11,879	8,263	4,132	1,549	0
Tax return (2000)	10,946	6,312	27,070	19,531	17,501	14,910	11,641	6,756	2,000	896	0

Notes: Comparison between after-tax wage income distributions in SHIW (household survey) and in our administrative tax return data. See text for details. Income is expressed in euros.

fathers and mothers: these are the same cohorts of parents in our tax return data.[18] Since SHIW only reports after-tax income data, we compare the distributions of this measure of income in SHIW and in our administrative data. Table 2 reports a number of statistics of the distributions in the two datasets. Overall, the distributions line up well, with the possible exception of children at the very bottom of the income distribution.

III. Intergenerational Mobility in Italy

We begin our empirical analysis by documenting the relationship between child and parental income at the national level.

Our definition of income is total gross income. Parental income is the sum of income of both parents. Child income is always defined at the individual level. Unless otherwise specified, all incomes are nominal.

We measure the rank of parents R_i^P as their percentile in the national distribution of parental incomes and the rank of children R_i as their percentile in the national distribution of child incomes. Figure 1 presents a binned scatterplot of the mean percentile rank of children as a function of their parents' rank, $\mathbb{E}\left[R_i \,|\, R_i^P = r\right]$.

Our first main finding is that the conditional expectation of a child's rank given her parents' rank is well approximated by a linear relationship throughout the income distribution, except at the very bottom, where it flattens out, and at the very top, where it bends upward.[19]

[18] We choose 2016 for children because 2017 is not a survey year. We use the 2000 wave of SHIW for parents, instead of the 1998 wave, because we observe tax liabilities of parents only for the 2000 fiscal year in our administrative data. As noted in the text, we need this information to draw a meaningful comparison with income in SHIW. We perform additional adjustments to make the two samples more comparable. Following Cannari, Ceriani, and D'Alessio (1997), we drop individuals who (i) report zero or missing income in the survey, (ii) report housing asset income below €500, and (iii) report only interest income. These criteria exclude from SHIW individuals who do not file tax returns.

[19] This linearity in the rank-rank relationship emerges also from other studies based on administrative records for Australia, Canada, Denmark, Norway, Sweden, and the United States (Boserup, Kopczuk, and Kreiner 2017; Bratberg et al. 2017; Chetty et al. 2014; Corak 2017; Deutscher and Mazumder 2019). Also in these studies, the data show significant deviations from linearity only at the extremes of the distribution.

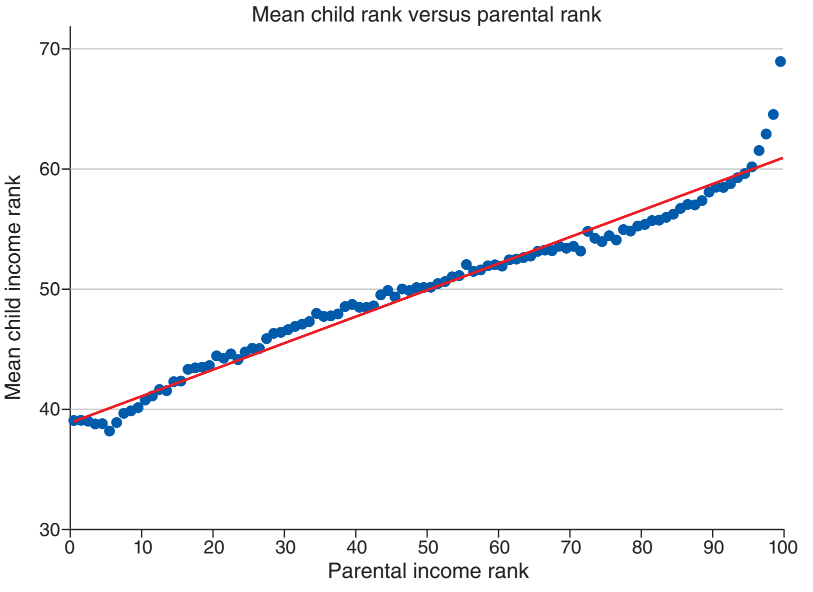

Figure 1. Child Mean Rank Conditional on Parental Income Rank

Notes: Blue dots: data. Red line: linear fit. The constant of the red line is 39 and the slope 0.22.

Running the OLS regression in (1), we estimate that a 1 percentage point (pp) increase in parental rank is associated with a 0.22 pp increase in the child's mean rank, as reported in column 1 of Table 3. This estimate of the RRS implies that, if we take two families, one at the ninetieth percentile of the income distribution and one at the tenth percentile, a generation later the child of the rich family is expected to still be two deciles above the child from the poor family. On average, it takes two generations for these gaps in initial conditions to be mostly offset so that the descendants of the two families (i.e., their grandchildren) would be expected to belong to the same decile.

The first column of Table 3 reports other positional measures of mobility. The top mobility ratio is 4.8, a reflection of the fact that in Figure 1, the rank-rank curve steepens toward the upper end. Thus, ranks persist a lot more at the top of the income distribution.[20] This value of the TMR implies that in the top decile the slope is close to 1: the children of two families, both in the top decile but at the two extremes of the decile, will both reverse toward the median but will still be 10 percentiles apart a generation later.

The AUM index equals 0.45, which means that a child of parents with income below the median is expected to end up in the forty-fifth percentile of her income distribution. The Q1Q5 index implies that, at the national level, children born in the

[20] One reason why this happens is mechanical. The distribution is right skewed, and percentiles are further apart at the top relative to the middle. They are also somewhat closer to each other at the bottom, which explains the flattening of the rank-rank relation for the first few percentiles.

Table 3

	Dependent variable: Child income rank					
	Core	Male	Female	Married father	Div/Sep father	2 earners
Parental income rank	0.220	0.222	0.227	0.225	0.177	0.219
	(0.0007)	(0.0010)	(0.0010)	(0.0008)	(0.0040)	(0.0010)
Constant	0.390	0.449	0.323	0.388	0.394	0.392
	(0.0004)	(0.0006)	(0.0006)	(0.0005)	(0.0024)	(0.0006)
AUM	0.445	0.504	0.379	0.445	0.438	0.446
Q1Q5	0.112	0.156	0.061	0.110	0.118	0.117
TMR	4.824	4.268	5.241	4.730	7.215	5.035
Observations	1,719,483	887,401	832,082	1,464,143	58,048	1,184,767

Notes: National indicators of intergenerational mobility for the core sample and various subgroups. Standard errors are in parentheses. "Male/Female" refers to children. "Married and Div/Sep father" restricts to children whose fathers are married and divorced/separated in 1998. "2 earners" restricts to children whose parents both earned positive income in 1998.

Table 4—National Quintile Transition Matrix (Percent)

	Child quintile				
Parental quintile	1st	2nd	3rd	4th	5th
1st	28.90	24.54	19.46	15.95	11.15
2nd	21.74	22.15	21.69	20.08	14.34
3rd	18.32	19.99	21.27	21.92	18.50
4th	15.96	18.05	20.85	22.63	22.51
5th	15.08	15.27	16.73	19.42	33.49

bottom quintile of the income distribution face an 11 percent probability of belonging to the top quintile as adults.[21]

Tables 4 and 5 contain the full national intergenerational transition matrix across quintiles and deciles. The probability that a child of a family from the top income decile remains in the top decile is over 26 percent, more than 6 times higher than the probability that a child from a family in the bottom decile ascends to the top one as an adult. Interestingly, the persistence rate in the top decile drops from 0.26 to 0.20 when computed on wage income only, whereas the other elements on the diagonal of the transition matrix are barely affected. This observation suggests that financial and business income play a significant role in perpetuating family ranks across generations at the very top of the income distribution.[22]

A. *Robustness to Local Deflating*

A concern underlying our national statistical analysis is that all incomes are nominal, while purchasing power varies substantially across geographical areas in Italy. A given income value in the South, appropriately deflated, becomes higher than its

[21] We also estimated RRS, AUM, and Q1Q5 separately for our five cohorts and did not detect any clear trend—if anything, a minor increase in mobility.

[22] This effect is even stronger as we zoom in on the upper tail. For example, when switching from total income (which includes financial and business income) to wage income only, persistence in the top 1 percent drops by more than half, from 0.13 to 0.06.

Table 5—National Decile Transition Matrix (Percent)

	Child decile									
Parental decile	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
1st	16.69	14.53	13.34	11.60	9.89	8.71	7.55	7.26	6.27	4.16
2nd	13.41	13.17	12.59	11.55	10.60	9.72	8.86	8.23	7.14	4.73
3rd	11.46	11.60	11.69	11.13	11.10	10.50	9.96	9.21	7.94	5.42
4th	10.22	10.21	10.67	10.80	10.91	10.89	10.61	10.38	8.96	6.36
5th	9.34	9.66	10.07	10.46	10.64	10.80	10.82	10.82	10.06	7.34
6th	8.57	9.07	9.59	9.87	10.30	10.79	11.13	11.07	11.05	8.55
7th	7.83	8.35	8.84	9.72	10.39	10.87	11.39	11.33	11.46	9.82
8th	7.59	8.14	8.40	9.14	9.75	10.69	11.18	11.36	12.00	11.75
9th	7.52	7.87	7.91	8.45	8.92	9.58	10.35	11.10	12.70	15.62
10th	7.37	7.41	6.90	7.29	7.52	7.44	8.17	9.24	12.42	26.25

counterpart in the North in real terms. A location-specific deflating procedure might therefore affect the ranks of parents and children in the national distribution.

To examine this issue, we used regional consumer price indexes constructed by the Bank of Italy for the year 2006 (Cannari and Iuzzolino 2009a). We explore four different indexes: the first one is a price index that allows for regional dispersion in prices of food, clothing, and furniture only. The other three indexes include alternative ways of accounting for the location-specific cost of housing services and other services. We also used an alternative deflating procedure based on the province-level price indexes for year 2005 constructed by Boeri et al. (2018) following the methodology proposed by Moretti (2013). The cross-regional correlation between all these cost-of-living indexes is very high, ranging from 0.53 to 0.99.[23]

Overall, these price indexes reveal cost-of-living differentials of up to 30 percent between the least expensive regions (e.g., Calabria) and the most expensive ones (e.g., Lombardia). Table 6 shows that our national results are quite robust to alternative deflating procedures. Our estimates of relative mobility rise somewhat when we use these regional deflators, but indexes of absolute mobility are nearly identical.[24] In light of the robustness, in what follows, we keep using nominal income.

B. *Median Regression and Conditional Distributions of Child Ranks*

The left panel of Figure 2 reports the median child rank conditional on parental income rank in addition to the mean rank already reported in Figure 1. The relationship between median child rank and parental rank remains linear, except at the very top, but it is a lot steeper than for the mean rank: the median rank-rank slope is 0.32. The sharp discrepancy between mean and median arises because the conditional distributions of child ranks are very skewed, as illustrated in the right panel of Figure 2. At the bottom quantiles, most of the mass is in the bottom ranks, whereas at the top quantiles, it is in the top ranks. Therefore, the median is lower than the mean for low parental ranks and higher than the mean for high parental ranks. This pattern induces a steeper positive relationship between parental rank and child median rank.

[23] Table A2 in the online Appendix reports all pairwise correlations between the indexes.

[24] A dampening of relative rank persistence is what one would expect if high-income areas are the high-cost-of-living ones. However, when there is a great deal of within-province income variation, the attenuation is small.

TABLE 6

	Dependent variable: Child income rank				
	Real 1	Real 2	Real 6	Real 9	Real "Moretti"
Parental income rank	0.211	0.195	0.190	0.191	0.199
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Constant	0.394	0.402	0.404	0.404	0.400
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
AUM	0.446	0.450	0.452	0.452	0.450
Q1Q5	0.115	0.121	0.123	0.123	0.120
TMR	5.378	5.909	6.206	6.189	5.571
Observations	1,634,970	1,634,970	1,634,970	1,634,970	1,634,970

Notes: National indicators of intergenerational mobility for the core sample based on PPP-adjusted income. Standard errors are in parentheses. "Real 1" to "Real 9" are estimates on the core sample after incomes of parents and children have been adjusted for regional differences in the price level. The numbers 1, 2, 6, and 9 denote the type of price index used, taken from Cannari and Iuzzolino (2009a). Specifically, Index 1 is constructed under the assumption that food, apparel, and furniture are the only consumption categories that vary in terms of price level across regions. Index 2 also includes house price variation across regions based on data from "Osservatorio sul Mercato Immobiliare" (Housing Market Monitor) within the Italian Revenue Service. Index 6 uses rents from Bank of Italy's Survey on Household Income and Wealth (SHIW), instead of house prices from the Housing Market Monitor, in order to account for variation in the cost of housing services. Moreover, for expenditure categories other than food, apparel, and furniture, regional price differences are imputed based on regression estimates using US data. Index 9 makes different assumptions. First, rents from SHIW used in this index are adjusted for house quality. Second, instead of imputing values based on US estimates, regional price differences in expenditure categories that fall into health care, maintenance, and other services (which account for 16 percent of consumption expenditure) are taken from data published by the Italian Ministry of Economic Development. These price differences are also adjusted for service quality. The remaining 22 percent of consumption expenditure for which there is no direct information is assumed to have no regional variation in price. This is the preferred index by the authors. The last column deflates incomes based on the local indexes constructed by Boeri et al. (2018) following the methodology proposed by Moretti (2013).

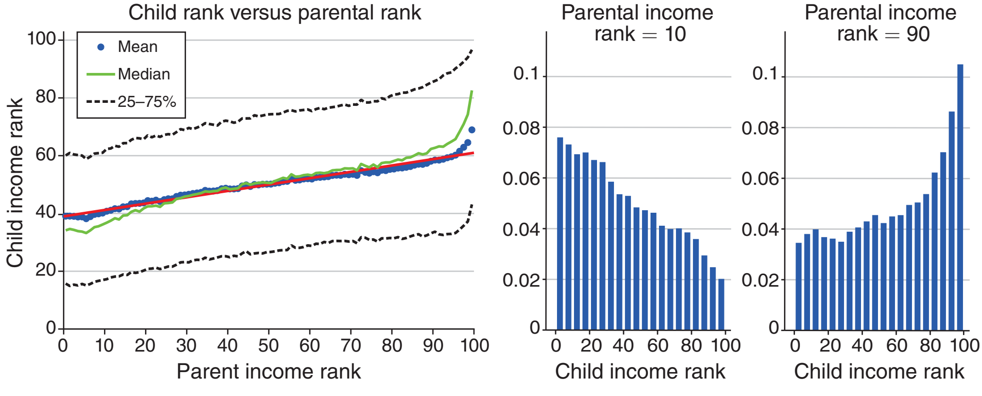

FIGURE 2

Notes: Left panel: mean child rank, median child rank, and interquartile range of child rank conditional on parental income rank. Right panel: conditional distributions of child rank at parental income percentiles 10 and 90.

The left panel of Figure 2 also reports the 25–75 percentile range of child outcomes at each parental rank. The plot reveals a wide dispersion of outcomes around the mean. The interquartile range of the conditional distribution of child ranks, averaged

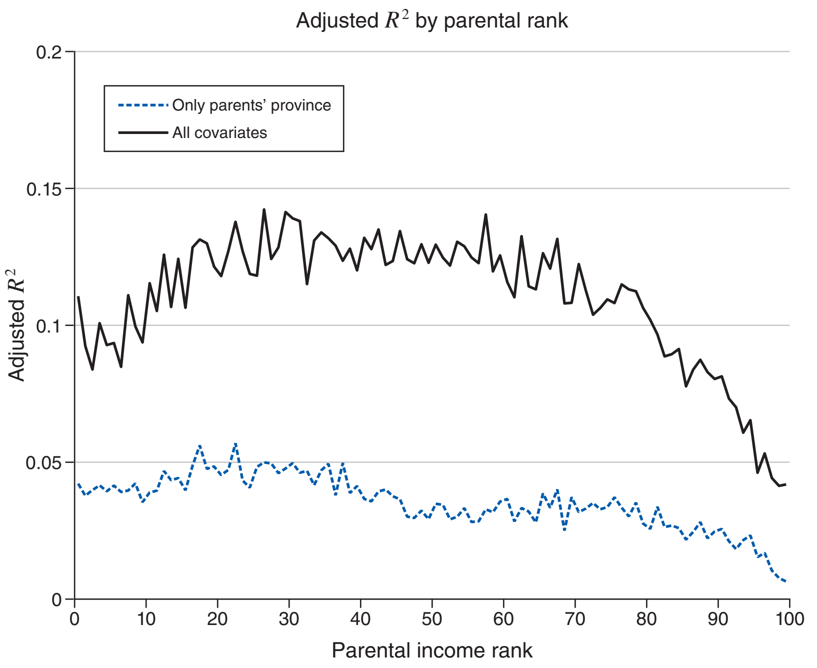

FIGURE 3. ADJUSTED R^2 OF THE 100 WITHIN-PARENTAL-RANK REGRESSIONS OF INDIVIDUAL CHILD RANK IN THE NATIONAL DISTRIBUTION ON VARIOUS PARENTAL AND CHILD COVARIATES

across parents' ranks, is 0.48.[25] For example, at the tenth percentile of parental income, one-quarter of children have incomes above the sixtieth percentile, and at the ninetieth percentile of the parents' distribution, one-quarter of children have incomes below the thirty-fifth percentile.[26] Put differently, even though the slope of the rank-rank relation is strongly statistically significant, the regression has a very low R^2, around 0.05.

This finding is not uncommon in the literature and recently has led to interest in methods that combine multiple parental predictors beyond income (e.g., Blundell and Risa 2019). Here, we build on this approach and ask "What explains, in our data, the vast within-quantile variability?" To answer this question, we regress child rank in each percentile of parental rank (i.e., we run 100 separate regressions) on a large number of individual covariates that include, for parents, marital status, age, province of residence, a self-employment dummy, and a geographical mover dummy; for children, age, gender, a self-employment dummy, and a geographical mover dummy. Figure 3 shows that, jointly, these variables explain around 11 percent of the within-parental-rank variation.[27] The hump-shaped pattern suggests that they account for more variation in outcomes of children from middle-class parents and less for children from poor and rich parents. Province of residence of the parents in

[25] Interestingly, Boserup, Kopczuk, and Kreiner (2017) also report for Denmark an interquartile range of the conditional distribution of child income ranks around 0.4.

[26] The dispersion in the distribution of ranks is not substantially affected by gender. When considering the income distribution of sons only, the interquartile range is 0.47.

[27] This is a relatively low R^2 compared to typical Mincer regressions. The main reason is that our data do not contain information on educational levels.

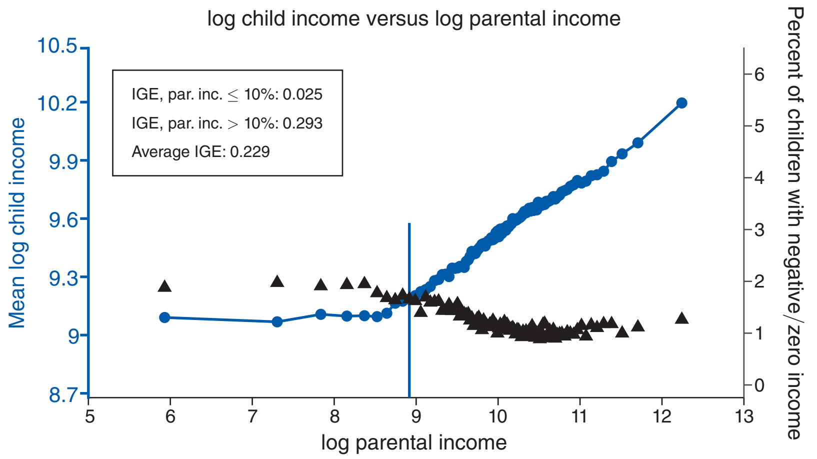

FIGURE 4. RELATION BETWEEN LOG INCOME OF THE CHILD AND LOG INCOME OF THE PARENTS

Notes: The vertical line denotes the tenth percentile. The figure also reports the fraction of children with negative or zero income at each bin (i.e., the percentage of observations dropped).

1998 (the location where the children grew up) accounts for much of the explained within-rank variation: between 1/5 and 1/2. In Section VI, we will document in detail the existence of sharp differences across Italian provinces in the degree of intergenerational mobility.

The main conclusion is that most of the conditional variability in child income ranks remains unexplained, suggesting that there is an enormous amount of unobserved heterogeneity left, even within parental rank and within province.

C. *A More Traditional Indicator: The IGE*

Historically, the most common indicator used in the literature to measure intergenerational mobility is the intergenerational elasticity of income (IGE) that we defined in Section I. Figure 4 plots the relation between log income of the child and of the parents, and the share of observations dropped among children's records because of zero or negative income. Only about 1.6 percent of observations are dropped, so this does not seem a serious problem in our dataset. This finding stands in contrast with Chetty et al. (2014), who report that in their data the IGE is very sensitive to how one handles zero income values.

The figure also reveals that the relationship is linear for much of the distribution, but in the bottom decile, it flattens out dramatically. While the IGE estimated over the entire sample is 0.23, the IGE estimated on the bottom decile is only 0.03, and the IGE estimated excluding it is 0.29.[28] The overall IGE is remarkably close to our estimate of the RRS.[29]

[28] This sharp drop in curvature at the bottom of the income distribution emerges for the United States as well (Chetty et al. 2014).

[29] Berman (2017) proves that under log-normality of the income distribution, the ratio of the RRS to the IGE converges to $\frac{3}{\pi}$, i.e., a number very close to 1. If we take the unconditional estimate of the IGE, our data offer sharp

In conclusion, the IGE and the RRS convey a coherent message about intergenerational mobility in Italy. However, as sufficient statistics to summarize mobility across the entire distributions, they both miss at the tails: the limitations of the IGE are especially evident at the bottom of the income distribution and the limitations of the RRS at the top.

Comparison with Existing Estimates of the IGE for Italy.—Previous authors have estimated the IGE on Italian cross-sectional survey data using imputation procedures based on Two-Samples 2SLS (TS2SLS). The key limitation of these surveys is that one cannot link fathers and sons: the surveys contain only income data for adult sons along with some demographic characteristics of their fathers but not their income. Researchers proceeded in two stages. First, from previous surveys sampling individuals observed during the childhood years of the adult sons, an instrumental variable (usually education) is used to predict income. Next, this instrument, present in the children's dataset as well, is used to impute a *pseudo father* income value to each child record. These studies all obtain estimates of the IGE ranging between 0.35 and 0.55 and hence higher than ours. Point estimates, though, are rather imprecise because of the small sample sizes, roughly 2,000–3,000 observations (Barbieri, Bloise, and Raitano 2018; Cannari and D'Alessio 2018).[30]

We have made an attempt to replicate this methodology as closely as possible on our data, running an exercise "as if" we did not have fathers matched with children but only two separate cross sections. In the absence of education among our observables, we use as instruments father's age, his province of birth, and his share of self-employment income out of total income. An F-statistic over 700 for the first stage rules out a weak instrument problem. The IGE estimated from observed child income and imputed father's income is 0.50.[31]

Our findings establish the presence of an upward bias in this procedure. The most likely reason is that the instrument commonly used in this literature, father's education, has an independent effect on child income beyond parental income. We observe precisely this problem in our full dataset. When we regress child income on actual father's income and the instruments in our dataset with matched fathers and sons, we find that the instruments remain strongly significant in the regression, thereby confirming the source of upward bias.

In sum, it appears that the key challenge for the TS2SLS approach in this literature is finding a valid instrument, one that predicts parental income but remains orthogonal to child income.

support to this approximation. However, if we exclude the bottom decile, the approximation is poor, possibly because of a Pareto tail in the empirical income distribution (estimated to be roughly 2 in our dataset).

[30] This gap between methodologies exists also for the United States. For example, when Björklund and Jäntti (1997) use this same imputation procedure on US data, they arrive at an IGE between 0.4 and 0.5, compared to an estimate around 0.35 obtained by the authors themselves on PSID and by Chetty et al. (2014) on matched father-son tax return data.

[31] It is also precisely estimated (SE $=$ 0.006 based on 200 bootstrapped replications) given that our sample size is over 300 times larger than that of existing studies.

TABLE 7

	Dependent variable: Child income rank				
	Only wage	> 2/3 wage	> 2/3 entr.	> 2/3 prof.	> 2/3 cap.
Parental income rank	0.231	0.231	0.228	0.216	0.211
	(0.0012)	(0.0011)	(0.0017)	(0.0038)	(0.0047)
Constant	0.378	0.379	0.386	0.404	0.396
	(0.0007)	(0.0007)	(0.0008)	(0.0029)	(0.0013)
AUM	0.436	0.437	0.443	0.458	0.449
Q1Q5	0.090	0.093	0.111	0.141	0.127
TMR	2.869	3.105	6.221	7.106	5.558
Observations	860,931	991,014	299,909	57,047	64,637

Notes: National indicators of intergenerational mobility for the core sample and various subgroups. Standard errors are in parentheses. "Only wage" is the restriction to children in a household who earns wage income only in 1998. "> 2/3 X" refers to the restriction to children in a household whose income in 1998 was made up by component X for more than 2/3. "Prof." means other self-employment income than entrepreneurial income (i.e., mostly from professional activity).

D. *Analysis for Different Population Subgroups*

Tables 3, 7, and 8 report results for various subgroups of the population of children. In all these cases, the position of parents and children remains the same: it is the national distribution of the sample of column 1 in Table 3.

Comparing males and females in Table 3 and focusing on the RRS, it appears that relative mobility is essentially the same. However, the AUM index and the Q1Q5 transition rate reveal that women have significantly lower absolute upward mobility. If we take a boy and a girl both from families in the bottom quintile, the probability for the boy of reaching the top quintile is more than twice as large. One of the determinants behind this result is that in Italy, female labor force participation is still quite low (it was 41 percent in 2019).

When we condition on the major source of parental income (labor, self-employment, and capital) in Table 7, we find more pronounced upward mobility for the children of self-employed professionals (e.g., artists, architects, lawyers, doctors, pharmacists): the Q1Q5 is 1.5 times larger than for children of wage earners.

Table 8 shows that children who, in 2017, reside in a different region from the one where their parents lived in 1998 (movers) display much higher income mobility. For example, their Q1Q5 transition rate is 0.23, compared to a baseline value of 0.11, and their AUM index is the fifty-seventh percentile, relative to a baseline value equal to the forty-fifth percentile. Therefore, a geographical move is strongly associated with an upward move in economic conditions. This is also true, but to a lesser extent, for children of fathers who are themselves movers, i.e., for those who in 1998 lived in a different region from the one where they were born, including foreign born (last column).[32]

[32] Tables A3 to A5 in the online Appendix repeat our analysis of national mobility and all these exercises on subgroups of the population for two alternative definitions of parental income: income of the father and income of the top earner of the household (who is the father in 87 percent of the cases). Tables A6 to A9 in the online Appendix report mobility measures and transition matrices for fathers and sons—the most comparable definitions across generations. In all instances, results are similar to those obtained for our baseline definition.

Table 8

	Dependent variable: Child income rank				
	Father top	Mother top	Mover reg., father	Mover reg.	Father born abroad
Parental income rank	0.226	0.193	0.183	0.129	0.195
	(0.0008)	(0.0017)	(0.0020)	(0.0022)	(0.0056)
Constant	0.387	0.400	0.426	0.541	0.410
	(0.0005)	(0.0009)	(0.0013)	(0.0013)	(0.0033)
AUM	0.444	0.448	0.472	0.573	0.459
Q1Q5	0.108	0.122	0.134	0.227	0.134
TMR	4.576	6.155	5.318	4.823	5.344
Observations	1,383,653	335,830	273,142	178,696	31,335

Notes: National indicators of intergenerational mobility for the core sample and various subgroups. Standard errors are in parentheses. "X top" refers to the restriction to children whose parent X was the top earner in 1998. "Mover reg., father" refers to the restriction to children whose fathers were born in a different region from their region of residence in 1998. "Mover reg." refers to the restriction to children whose region of residence in 2017 is different from the region of residence of their fathers in 1998. "Father born abroad" restricts to children whose fathers were born abroad.

E. *Siblings: The Role of Birth Order*

Even though our dataset does not contain explicit identifiers for siblings, we can easily recover siblings by matching children who have the same parental record, i.e., a record for parents that coincides with respect to all the variables in our dataset. Through this procedure, we are able to identify around 171,000 parents with at least 2 children.[33] In order to control for the impact of age on income of siblings, we restrict the sample to cases where, in our dataset, we see the siblings at approximately the same age. Specifically, we restrict the attention to children who are 36 or 37 years old in 2017—around the mean age of children in our dataset—leaving us with around 85,000 records of parents with at least 2 children.

We exploit this sample of siblings by asking whether birth order matters: is there any systematic difference between earlier-born and later-born children in terms of upward mobility?[34] Table 9 shows the results of our analysis. We find that the median rank of earlier-born children is between 0.3 and 2.9 percentiles higher compared to the median rank of the later-born ones. Interestingly, this gap increases steeply with parental income and is statistically different from zero only for families at or above the middle quintile of the income distribution.

IV. Corrections for Potential Sources of Bias

In this section, we verify the robustness of our analysis with respect to (i) attenuation and life cycle biases arising from the fact that we only have a few years of

[33] By requiring a perfect match for all the variables in the dataset, we impose a very restrictive criterion, but we prefer to end up with a smaller sample without false positives to avoid any sort of measurement error.

[34] Some of the children we label as "earlier born" may have older siblings we do not observe because of our age restrictions on the sample. These cases, however, only arise in families with at least 3 children, which are only 6 percent of the total number of Italian families. Thus, the vast majority of earlier born are firstborn.

Table 9

Parental rank	Median child rank: Earlier born	Median child rank: Later born	95% CI for the difference
100	65.2	62.3	[1.6, 4.0]
80	56.6	54.7	[0.7, 3.2]
60	52.0	49.9	[0.7, 3.6]
40	44.5	43.5	[−0.1, 2.2]
20	36.1	35.9	[−0.9, 1.1]

Note: Median rank of firstborn children versus later-born children by parental income ventile.

data available for each cohort and that we observe fathers and children at different ages and (ii) the omission of certain types of taxpayers from our sample and the underreporting of income for those in the sample because of tax evasion.[35]

A. *Attenuation and Life Cycle Bias*

Throughout our analysis, we have used 3-year averages of income measured around age 36 for children and age 48 for fathers as proxies of their permanent income. There are two potential problems with this measure. First, because income is volatile, by averaging over a short span as we do, the transitory component does not fully average out and can lead to an attenuation bias in the estimates of intergenerational persistence (Solon 1992). Second, because child income is measured 12 years earlier than father's income, its elasticity with respect to lifetime income could be lower than that of father's income—a life cycle bias that, again, would lead to an overestimate of intergenerational mobility (Haider and Solon 2006). Figure 5 illustrates why the life cycle bias may be present in our sample. The figure plots the estimated RRS and AUM computed on samples with increasing three-year age ranges for children (age is measured in the year income is observed). Both indicators start leveling off only after age 40, while our sample mean is 36. As a comparison, in both the United States and Sweden, the RRS levels off in the early 30s.[36] This discrepancy is not surprising since Italy is one of the countries with the longest school-to-work transition.

To address these two sources of bias, we need a reliable strategy to extrapolate the relation between current and permanent income at various ages. We implement two alternative approaches. In the first one, we leverage the short panel dimension present in our administrative data to estimate an error component model for income and simulate lifetime income using this model. We start by pooling fathers and sons in order to construct data for the longest possible age range. We then estimate a quartic polynomial in age for log income. Next, we residualize log income with respect to

[35] In the online Appendix, we describe two other corrections that leave our estimates virtually unchanged: the omission of poor individuals who do not file a tax return and the absence of certain types of capital income from the tax returns.

[36] See figure III.A in Chetty et al. (2014) and figure 1c in Nybom and Stuhler (2017). We should note, though, that Mazumder (2016) argues that the figure in Chetty et al. (2014) is somewhat misleading because of the changing age of parents and, even in the United States, life cycle bias does not entirely level off in the early 30s.

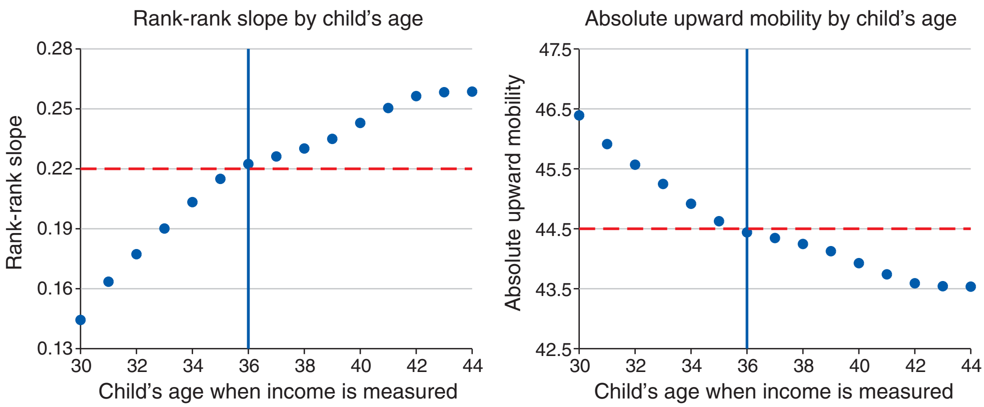

Figure 5. Estimates of RRS (Left Panel) and AUM (Right Panel) Based on Different 3-Year Age Ranges for Children

Notes: The x-axis reports the midpoint of the age range measured in the year in which child income is observed. The vertical blue line indicates the midpoint for our sample, while the dashed red line indicates the estimate of RRS/AUM in the sample.

this age profile and assume a time-stationary persistent-transitory process for log residual income, a common representation in the literature on income dynamics:

$$\log y_{i,a}^{j} = \kappa_{i,a}^{j} + \epsilon_{i,a}^{j}$$

$$\kappa_{i,a}^{j} = \rho \kappa_{i,a-1}^{j} + \eta_{i,a}^{j},$$

where a denotes age, $j \in \{f, s\}$ is an indicator for father or son, $\kappa_{i,a}^{j}$ is a persistent income component following an AR(1) process with autocorrelation ρ and standard deviation of shocks σ_{η}, and $\epsilon_{i,a}^{j}$ is an uncorrelated shock with standard deviation σ_{ϵ}. The two shocks are orthogonal to each other, i.i.d. across all individuals, and normally distributed. Let σ_{κ_0} be the initial standard deviation of the persistent component at age $a = 0$. The initial conditions $\left(\kappa_{i,0}^{s}, \kappa_{i,0}^{f}\right)$ of a son-father pair are drawn from a bivariate normal distribution with correlation ϱ_0. Table 10 reports estimates for the structural model parameters $\left\{\rho, \sigma_{\eta}, \sigma_{\varepsilon}, \sigma_{\kappa_0}\right\}$.[37]

The experiment then proceeds as follows. We run 100 simulations of a panel of 849,921 father-son pairs (the number of pairs in our dataset) for 25 years and compute for each father and son their average lifetime income (the ideal measure of permanent income) and the average income over 3 consecutive years (the noisy measure of permanent income that we have available) at ages 47–49 for fathers and 35–37 for sons. For these simulations, we reconstruct income using both the estimated deterministic age profile and the stochastic process above. Setting $\varrho_0 = 0.395$ reproduces almost exactly our estimated values of the rank-rank slope, AUM, and Q1Q5 for fathers and sons (Table A6 in the online Appendix) when we use the simulated three-year proxy for income. The last step is computing the

[37] In the online Appendix, we provide more details on the identification and estimation strategy.

Table 10—Parameter Estimates of Persistent-Transitory Process for log Residual Earnings

σ_{κ_0}	ρ	σ_η	σ_ϵ
0.588	0.983	0.108	0.132

Table 11—Average Mobility Statistics Using 3-Year Average Income or Lifetime Income Based on Simulations of the Persistent-Transitory Model

	RRS	AUM	Q1Q5	IGE
3-year average	0.226	0.444	0.115	0.236
Lifetime income	0.276	0.431	0.099	0.288

rank-rank slope and the other mobility statistics using the simulated measures of average lifetime income.

Table 11 reports mobility statistics averaged across the 100 simulations. As expected, the short sample induces a downward bias in relative mobility. The size of the bias, however, is modest: around 22 percent for the rank-rank slope (i.e., the true RRS would be 0.28) and slightly larger for the IGE (i.e., the true IGE would be 0.29). The bias is more limited for indicators of upward mobility such as AUM and Q1Q5.

Our second approach leverages a different data source, the Italian Social Security (INPS) dataset, kindly made available to us by Hoffmann and Malacrino (2019).[38] These data do not allow to match family members across generations, but they have the advantage of following individual work histories for over 30 years (1985–2016). In this long longitudinal sample, the rank correlation between a 3-year average of individual earnings at the age of our sons (35–37) and their lifetime income is 0.92, and the rank correlation between a 3-year average of earnings at the age of our fathers (47–49) and their lifetime income is 0.96. These high values give us confidence that we have a strong proxy for permanent income.

In order to further assess the robustness of our results to attenuation and life cycle bias, we use the INPS dataset to estimate a log-linear statistical model relating three-year average earnings $y_{i,a}$ to lifetime average earnings $\bar{y}_i$:

$$\log y_{i,a} = \beta_a \log \bar{y}_i + \epsilon_{i,a}, \tag{9}$$

where a denotes a three-year age interval. For our purposes, a equals 35–37 years of age for sons and 47–49 for fathers. We assume that lifetime earnings of sons and fathers are jointly normally distributed with identical marginal distributions $N(\mu_{\bar{y}}, \sigma_{\bar{y}})$ and correlation λ and that innovations $\epsilon_{i,a}$ are also normally distributed with standard deviation σ_ϵ^a. Table 12 presents the parameter estimates for this model.

[38] INPS collects data on employer-employee relationships in order to compute social contributions and pension benefits. The dataset is based on workers born on 24 randomly selected dates from the universe of all Italian dependent employees in the nonfarm private sector. The data represent a 6.6 percent sample of the population. The notion of income is gross labor earnings. See Hoffmann and Malacrino (2019) for details.

Table 12—Parameter Estimates of log-Linear Model for 3-Year Average and Lifetime Earnings Estimated on INPS Data

$\beta^{[35,37]}$	$\sigma_\epsilon^{[35,37]}$	$\beta^{[47,49]}$	$\sigma_\epsilon^{[47,49]}$	$\mu_{\bar{y}}$	$\sigma_{\bar{y}}$
0.895	0.186	1.081	0.148	10.32	0.400

Given these parameters, we draw 849,921 pairs of lifetime earnings $\bar{y}_i$ for fathers and sons and, from equation (9), we generate the same number of 3-year average earnings draws $y_{i,a}$ for fathers and sons. In line with our previous simulation exercise, the father-son lifetime income correlation λ is set so that the rank-rank slope of sons' three-year average earnings $y_i^{[35,37]}$ on fathers' three-year average earnings $y_i^{[47,49]}$ matches the corresponding estimate in our data (0.226), yielding $\lambda = 0.281$.

Table 13 shows that mobility estimates on simulated lifetime earnings are remarkably similar to the corresponding estimates from the previous exercise, despite using a different dataset and a different methodology. This close alignment gives us confidence that our correction is accurate.

B. *Omission of Certain Taxpayers and Tax Evasion*

Omission of Taxpayers Who Filed Form 770.—In Section II, we explained that our sample of 730 and MU tax returns does not include individuals who file tax form 770. These missing observations may distort our results because this group of individuals is not necessarily representative of the population in terms of income levels and sources of income.

Here we correct for this selection bias. First, we collect aggregated data on the number of forms 770 submitted by region and by 20 income classes that are published yearly by the Ministry of Economy and Finance. Next, we identify the taxpayers in our dataset of parents who are similar to those who filed form 770.[39] We then split these taxpayers into the same 20 income groups, region by region. Finally, we reweigh each observation in an income/region cell by a factor equal to the ratio of total (MU+730+770) taxpayers to (MU+730) taxpayers in that cell. Table 14 reports the results of this reweighing procedure and shows that our estimates of intergenerational mobility are virtually unchanged after this correction.

Tax Evasion.—Italy is notoriously a country with high tax evasion. In recent years, the size of the nonobserved economy was estimated by the Italian National Statistical Institute to be as large as 12 percent of GDP. A common rate of tax evasion across the population would not affect relative ranks and estimates of mobility, but this benchmark is far from reality. The propension to evade taxes differs significantly across earner categories. For dependent workers and retirees, taxes are withheld at the source. Thus, evasion is nearly impossible for these groups. Self-employment and rental incomes are, instead, much easier to hide.

[39] Specifically, this means looking for taxpayers who have only labor income and rental income below €568 in 1998. The key reason why these taxpayers opted for forms 730 or MU instead of the form 770 is because they claim itemized deductions (e.g., for medical expenses, mortgage interests, charitable donations, etc.).

Table 13—Average Mobility Statistics Using 3-Year Average or Lifetime Earnings Based on Simulations of the log-Linear Model

	RRS	AUM	Q1Q5	IGE
3-year average	0.226	0.444	0.115	0.267
Lifetime earnings	0.269	0.433	0.101	0.281

Table 14—National Indicators of Intergenerational Mobility

	Dependent variable: Child income rank			
	Core	With reweighting for missing 770	With adjustment for tax evasion	Combined
Parental income rank	0.220	0.212	0.260	0.246
	(0.0007)	(0.0006)	(0.0007)	(0.0005)
Constant	0.390	0.399	0.370	0.378
	(0.0004)	(0.0003)	(0.0004)	(0.0003)
AUM	0.445	0.452	0.435	0.439
Q1Q5	0.112	0.111	0.102	0.100
IGE	0.229	0.213	0.278	0.252
Observations	1,719,483	3,191,802	1,719,483	3,191,802

Notes: Correction for missing 770 forms and tax evasion. Standard errors are in parentheses. See the main text for details.

The best available estimates of tax evasion are based on a comparison between survey data (SHIW) and administrative tax return data. The key assumption is that self-reported income in surveys is much closer to the truth since the anonymity of survey respondents is protected by privacy laws. These studies compute average after-tax income from surveys and tax returns for groups of earners with the same type of income, income decile, and region. They confirm negligible rates of tax evasion for dependent workers. Tax evasion rates are, instead, estimated to be higher for other sources of income. Marino and Zizza (2012) estimate average rates of tax evasion around 40 percent for self-employment income and up to 80 percent for rental income. Fiorio and D'Amuri (2005) show that tax evasion rates decrease steeply with the level of income. For example, for the self-employed, tax evasion declines from 70 percent of reported gross income in the lowest decile to 8 percent in the top decile.

We use these sources to inflate the relevant income categories in all our records for both children and parents. The corrected estimates of intergenerational mobility in Table 14 imply a modest increase of the RRS and a smaller reduction in the upward mobility indices. Intuitively, if self-employed parents underreport income and their children do not, for example, because they are dependent workers (or the other way around), mobility would be overestimated.

C. *Taking Stock: Combining the Sources of Bias*

As seen in this section, most of the sources of bias shift our point estimates of intergenerational income mobility toward lower mobility. Here, we combine all of them to assess the overall potential bias in our baseline calculations. When we combine the correction for the missing taxpayers who submit the 770 form and the adjustment for tax evasion, we obtain the estimates in the last column of Table 14. Next, we add the correction for attenuation and life cycle bias obtained from our first strategy.[40] We obtain an RRS of 0.30, an IGE of 0.31, an AUM index of 0.43, and a Q1Q5 of 0.09.

V. Cross-Country Comparison

How does the level of intergenerational mobility in Italy compare to that estimated in other countries? We limit our analysis to studies that used large administrative datasets like ours.

The estimate of the RRS for Denmark is 0.18—corresponding to an AUM index of 0.46 (Boserup, Kopczuk, and Kreiner 2017). Bratberg, Nilsen, and Vaage (2005) reports an RRS of 0.19 for Norway. Heidrich (2017) estimates an RRS of 0.2 for Sweden. Deutscher and Mazumder (2019) measures an RRS of 0.21 and an AUM index of 0.45 for Australia. Meneses (2020) estimates an RRS of 0.21 for Chile. Corak (2017) reports an RRS of 0.24 and AUM index of 0.44 for Canada. The United States has a higher RRS, around 0.34, with an AUM index of 0.41 (Chetty et al. 2014). Thus, Italy's level of intergenerational mobility lies between that of Scandinavian countries and that of the United States, as illustrated in Figure 6, which reproduces the full binned scatterplots for Denmark, Italy, and the United States. An important caveat, when comparing these measures, is the exact definition of income used in the calculations. Sometimes it is individual income, sometimes it is family income. Chetty et al. (2014) compute the RRS for both definitions and show that the RRS computed based on individual income of the child, which is our definition, is 20 percent lower than the RRS estimated using family income of the child.

Next, we explore the mobility indicator Π defined in equation (8), the probability that a son overtakes his father in the level of income at his same age, after adjusting for inflation. We estimate this probability to be 0.53 in our cohorts of children. The corresponding estimate for the same US cohorts of sons and fathers—based on Chetty et al. (2017a)—is almost the same, 0.55.[41] Figure 7 plots this probability at each percentile of father's income for both countries. Comparing the red line in the left panel (Italy) to the US counterpart in the right panel reveals that the similar point estimates for Π mask the fact that Italy has more upward mobility at the bottom and less at the top of the distribution, compared to the United States.

As described in Section I, we run three counterfactuals to control for differences in positional mobility, growth, and inequality between Italy and the United States. First, we confirm the finding of Berman (2020) that changes in the copula have

[40] To do this, we compute the factor of adjustment between baseline and simulation results, and we apply it to our mobility indicators in the last column of Table 14.

[41] Since for Italy we only observe individual income (and not family income) of children, to ensure comparability between Italy and the United States, we focus on fathers and sons in both countries, relying on sons' and fathers' income distributions available from Chetty et al. (2017a). We rescale sons' income distributions in Italy and the United States to match real income per capita growth in each country between 1999 and 2017—the midpoints of the year range for which we observe fathers' and sons' incomes in our data. Fathers and sons are on average 36 in Italy and 30 in the United States when their income is measured.

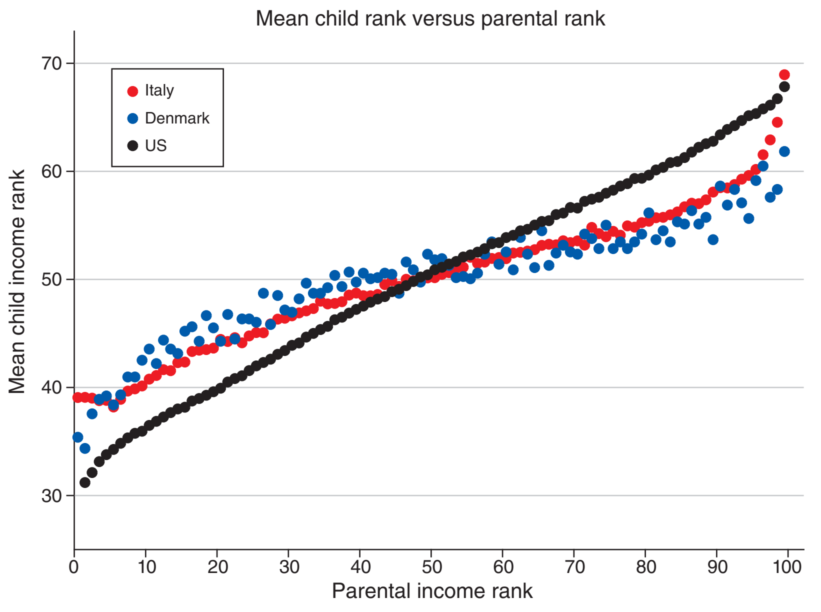

Figure 6. Mean Child Rank Conditional on Parental Income Rank for Denmark, Italy, and the United States

small effects: using the copula estimated by Chetty et al. (2017a) for the United States, the Π index for Italy is basically unchanged.

The other two counterfactuals where we control for income growth and income inequality differentials are, instead, more salient. On the one hand, we find that, if Italy had experienced the same per capita growth as the United States between 1999 and 2017, i.e., 1.2 percent instead of 0.01 percent per year (the blue line in the left panel of Figure 7), Π would have been 0.64 in Italy. On the other hand, if income inequality in Italy had been as in the United States (the purple line in the left panel), Π would have been 0.46. We conclude that lower income growth and lower income inequality in Italy, relative to the United States, are two forces that offset each other almost perfectly and explain why this indicator of mobility ends up being very similar in the two countries.

VI. Geographical Variation

To investigate the geographical variation in intergenerational mobility within the country, we focus on provinces. A province is an administrative division of intermediate level between a municipality and a region.[42]

[42] Over the period 1998–2017, the number of provinces increased from 103 to 110 before decreasing to 107. We use the geographic partition in 110 provinces established in 2009. Figure A2 in the online Appendix contains a map of the Italian provinces, and Tables A10–A12 list all the provinces, their population, region, and macro area (Northeast, Northwest, Center, South, and Islands). The average population of a province was 551,000 as of 2010, but there is large heterogeneity. The largest province, Rome, has over 4 million residents and contains 121 different municipalities. The smallest province, Ogliastra (Sardinia), has less than 60,000 residents and only includes 23 municipalities.

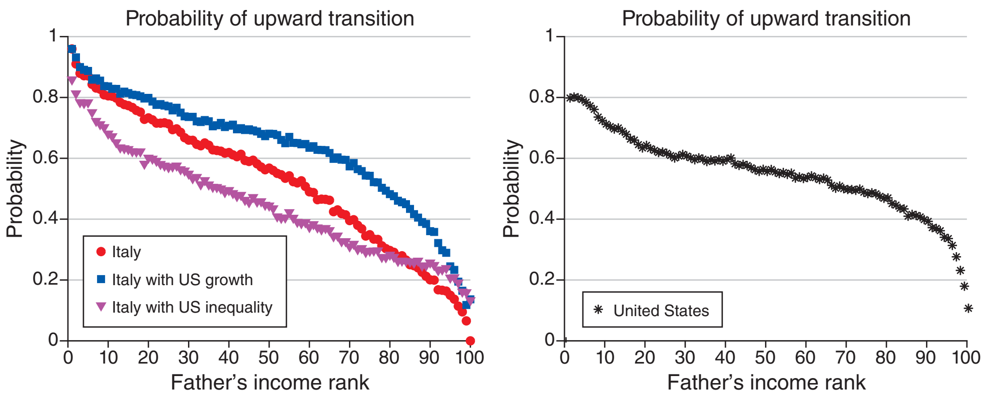

Figure 7. Probability That the Income of a Son Is at Least as High as His Father's Income in Real Terms (Base Year: 2014).

Notes: Left panel: Italy (with counterfactuals described in the text). Right panel: United States.

In order to analyze the province-level variation in mobility measures, we assign each child to the province that her father (or her mother, if her father's information is missing) indicated as the province of residence in his (her) own 1998 tax return. Such province should be interpreted as the area where the children grew up.

We document mobility at the provincial level using the same definitions of parental and child income and the same sample we used for the national analysis of Section III and Table 3, column 1. Income ranks for children and parents are defined with respect to their national distributions. We also report mobility statistics on incomes adjusted for different purchasing power at the regional level, as described in Section IIIA.[43]

The extent of the difference in relative and upward mobility across provinces is summarized by Tables 15 and 16. These tables report mobility measures for the top 10 and bottom 10 provinces among the largest 50 provinces in Italy based on resident population in 2010. A stark pattern starts already emerging from these tables: each of the top ten provinces is in northern Italy, and each of the bottom ten is in the South.

Figure 8 graphically summarizes geographic variation in AUM across all the 110 provinces. In this heat map, darker colors correspond to more mobile areas. Two broad spatial patterns emerge from this figure. First, there is substantial heterogeneity in upward mobility across provinces. The interquartile range of AUM across provinces is 0.1, almost 1.5 times as large as the one estimated by Chetty et al. (2014) across the 700 US commuting zones. Second, upward mobility has a clear North–South gradient and is highest in the Northeast of the country, especially in the regions of Veneto, Lombardia, and Emilia-Romagna. A within-between-macro area variance decomposition for AUM implies the between component, i.e., variation in

[43] We have used Index 9 for these figures, but our conclusions are robust to using other indices. As discussed in that section, the different local price indexes we could use are all highly correlated. Figure A3 in the online Appendix shows geographic variation in AUM and Q1Q5 after adjusting income for different purchasing power at the province level using the index constructed by Boeri et al. (2018).

Table 15—Top Ten Provinces by Absolute Upward Mobility

Province name	Population in 2010	AUM	Q1Q5	RRS	IGE
Bolzano	507,657	0.574	0.282	0.152	0.169
Monza-Brianza	849,636	0.545	0.204	0.160	0.175
Bergamo	1,098,740	0.544	0.200	0.142	0.171
Treviso	888,249	0.541	0.183	0.110	0.127
Milano	3,156,694	0.539	0.224	0.161	0.167
Trento	529,457	0.534	0.198	0.127	0.152
Vicenza	870,740	0.533	0.167	0.139	0.161
Venezia	863,133	0.530	0.156	0.116	0.111
Padova	934,216	0.530	0.168	0.137	0.142
Modena	700,913	0.529	0.198	0.164	0.169

AUM among the five macro areas (Northeast, Northwest, Center, South, and Islands), accounts for 82 percent of the total variation of AUM across the 110 provinces. For example, the highest-ranked province for AUM not in the Northeast or Northwest is Firenze (Tuscany) at position 34/110. The lowest-ranked province for AUM not in the South and Islands macro-region is Viterbo (Lazio) at position 75/110.[44]

Quantitatively, these differences are meaningful. The province with the highest AUM is Bolzano (Trentino-Alto Adige), with a value of 0.57, and the one with the lowest AUM index is Ragusa (Sicily), with a value of 0.35. This expected rank differential corresponds to nearly €7,500 (nearly 40 percent of median annual income in 2017), and hence, it translates into substantial gaps in children's lifetime incomes.

The right panel of Figure 8 reveals that, once adjusting for different cost-of-living levels across regions, the predominance of the Northeast macro area in terms of upward mobility is accentuated. At the same time, a few pockets of upward mobility emerge also in southern Italy.

These differences in upward mobility are equally pronounced for the Q1Q5 index, i.e., the probability that a child from a family in the bottom quintile of the national income distribution makes it to the top quintile, as shown in Figure 9. The least mobile provinces have transition rates around 6–7 percent and the most mobile around 21–22 percent, i.e., larger by a factor of 3. As evident from the heat maps, the correlation between AUM and Q1Q5 indicators is very strong (0.92).

Geographical dispersion is also high for relative mobility. For children growing up in Treviso (Veneto), being born from a family at the bottom of the national income distribution translates into only 11 percentiles of rank differential compared to someone born from parents at the top of the distribution. For children growing up in Brindisi (Puglia)—the province with the highest RRS among the largest provinces by population—it translates into a gap more than twice as big, i.e., 24 percentiles. Large differences across provinces arise also in the IGE: the cross-province correlation between RRS and IGE is 0.73.

Figure 10 plots the full rank-rank relation in two of Italy's largest metropolitan areas, Milano and Napoli. Milano (shown in blue) displays a rank-rank relationship that is both flatter and everywhere higher compared to Napoli (shown in red).

[44] Tables A13–A15 in the online Appendix show the ranking and all mobility measures—including RRS and IGE—for the 110 provinces.

Table 16—Bottom Ten Provinces by Absolute Upward Mobility

Province name	Population in 2010	AUM	Q1Q5	RRS	IGE
Trapani	436,624	0.393	0.085	0.190	0.226
Cagliari	563,180	0.392	0.082	0.160	0.179
Salerno	1,109,705	0.390	0.086	0.210	0.194
Agrigento	454,002	0.381	0.071	0.206	0.191
Messina	653,737	0.379	0.080	0.212	0.207
Reggio-Calabria	566,977	0.377	0.086	0.219	0.176
Siracusa	404,271	0.376	0.078	0.204	0.200
Catania	1,090,101	0.376	0.078	0.207	0.203
Palermo	1,249,577	0.373	0.076	0.186	0.205
Cosenza	734,656	0.363	0.077	0.206	0.176

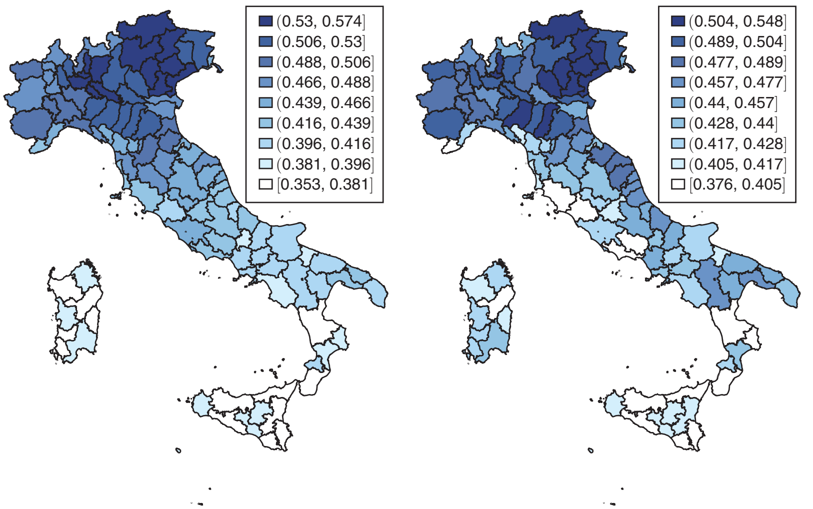

Figure 8. Heat Map of Absolute Upward Mobility across Provinces

Notes: Dark areas are more mobile. Left panel: AUM computed on nominal income. Right panel: AUM computed based on PPP-adjusted income (with regional price indexes).

Through the lenses of a utilitarian planner, Milano dominates Napoli: children who grow up in Milano fare uniformly better across the whole income distribution, and ex post their income distribution is less unequal. Notice also that the rank-rank relationships are quite linear even at the provincial level. We verified that this is true for all the largest provinces—especially so for those in the North. Many of the provinces in the South, with the lowest level of mobility, show a rather sharp increase in slope at the top of the distribution, similar to what we observe at the national level. Figures A5 and A6 in the online Appendix display the rank-rank relationship for the top and bottom eight provinces.

We conclude that areas in northern Italy (especially the regions in the Northeast), relative to the South, are both more egalitarian (higher relative mobility) and more upward mobile (higher absolute mobility). In the North, children from parents with unequal background are more similar in their economic outcomes when adults, and children from poor parents fare better when adults.

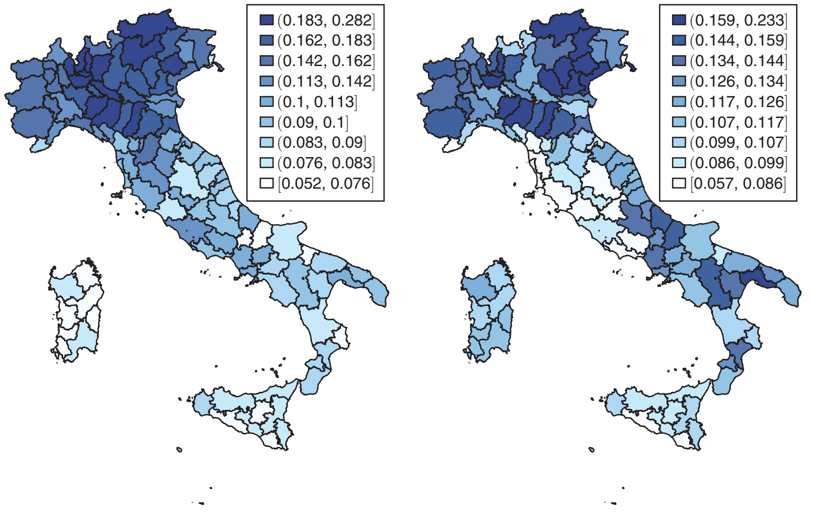

Figure 9. Heat Map of Q1Q5 across Provinces

Notes: Dark areas are more mobile. Left panel: Q1Q5 computed on nominal income. Right panel: Q1Q5 computed based on PPP-adjusted income (with regional price indexes).

Recall that, as explained in Section I, while at the national level RRS and AUM are tightly linked, at the level of a province it need not be the case. Equation (7) makes it clear that the AUM of a province can be higher either because the constant of the regression is high, or because the slope is high. In our data, it turns out that constant and slope are strongly negatively correlated (the correlation coefficient is -0.72), but most of the variation in AUM across provinces is accounted for by the constant terms, suggesting that the province effect materializes mostly through moving every child raised in that province up or down in the national income distribution, independently of their parental rank. This result is consistent with the large observed gap in aggregate income growth between the North and the South, combined with the low degree of geographical mobility from the South to the North. An important implication of this finding is that provinces with high (low) upward mobility also feature low (high) downward mobility: the cross-province correlation between Q1Q5 and Q5Q1, the probability of moving from the top quintile to the bottom quintile of the income distribution, is -0.85.

A possible source of concern with our analysis may be that the stark geographical variation in mobility that we have documented is driven by the large gaps in the level of income between the North and the South of Italy. For example, if the distribution of income in the South is shifted to the left compared to the one in the North, when computing AUM using the national distribution, upward mobility in the South would be mechanically lower. As argued in Section I, using the twenty-fifth percentile of the national distribution of parental income —which is what we do— already largely solves this problem. In order to further address this concern, when we compute indices of mobility of a province, we reweigh observations in each bin (e.g., each decile) by the ratio between the share of fathers in that bin of the national distribution (e.g., 0.1) and the share of fathers in the same bin residing in that province. Under this reweighing scheme, mobility estimates by province are barely affected.

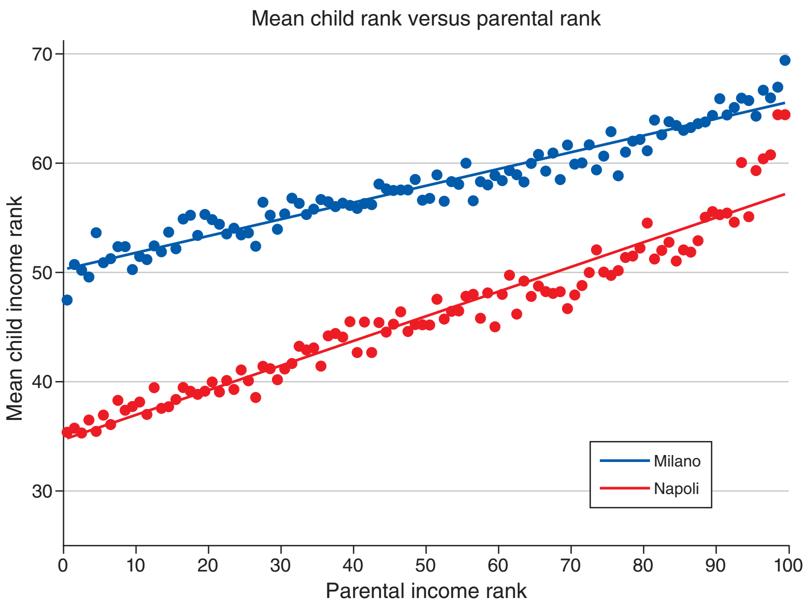

Figure 10. Child Rank–Parental Rank Relationship for Children Who Grew Up in Milano and Napoli

Note: The RRS are, respectively, 0.16 and 0.22, and AUM indicators are 0.54 and 0.41.

A separate question is whether the local income rank matters over and above the national one. Consider two families at the same quantile of the national income distribution, where the first one is at the top of its provincial distribution and the second one is at the bottom. Do the children of the first family fare better in the national distribution, possibly because they have access to better local opportunities? We conclude that they do not: in a regression of national child income rank on national parental income rank, province, and local parental income rank, the latter variable has no statistically significant effect.

A. *Is There a Great Gatsby Curve within Italy?*

The Great Gatsby curve refers to a negatively sloped empirical relation between income inequality and intergenerational mobility. This relationship has been extensively documented using cross-country variation and often interpreted as the outcome of different national institutions. The stark geographical heterogeneity across Italian provinces provides us with a source of variation while controlling for national-level institutions.

In Tables 17, 18, and 19, we correlate absolute and relative mobility, respectively, with various indicators of income inequality. A negative relation emerges for most inequality indices. One reason is that, when inequality is higher, the rungs of the income ladder are further apart and it becomes more difficult to climb them. The fact that the IGE is more weakly correlated with inequality indices than the RRS supports this hypothesis. Somewhat surprisingly, though, the income share of the top 1 percent is positively correlated with all three indices of mobility. This result contrasts with the finding for the United States where the top income share is negatively correlated with the AUM (table V in Chetty et al. 2014).[45] In Italy, the top income share is highly positively correlated with the fraction of entrepreneurs and professionals in the province, and we observed earlier that upward mobility is higher for children of families with a large component of self-employment income.

B. *Comparison with Informational Content of Surname Indices*

Reliable estimates of intergenerational mobility require large datasets that link successive generations, like ours. Such datasets only exist for a handful of countries and, until recently, were not available at all. Therefore, it is always useful to propose alternative approaches to the measurement of intergenerational mobility that have less strict data requirements.

Güell, Rodríguez Mora, and Telmer (2015) introduced a new indicator that overcomes some of these difficulties, the Informational Content of Surnames (ICS). The only data required for this methodology is a cross section of individual records with information on income and on the surname of the individuals. The data are then used to construct an indicator of the extent to which family names capture the variance of income (ICS Index). The basic idea is simple. Surnames are intrinsically irrelevant for the determination of economic outcomes, but they get passed from one generation to the next, alongside other determinants of income such as ability, wealth, and privileges, for example. The more these inherited characteristics matter for economic outcomes, the more information surnames contain on the realization of outcomes (the higher is the ICS index, i.e., the share of cross-sectional income variation explained by surnames), and the lower the degree of social mobility.[46]

In a recent paper, Güell et al. (2018) have constructed ICS indicators of social mobility at the level of Italian provinces and correlated them with many socioeconomic indicators.[47] We have the opportunity to assess how well the ICS index correlates with more direct measures of intergenerational mobility based on rich administrative data. This validation exercise has never been performed before in any country.

In Table 20 we report the cross-province correlation between the various ICS indices calculated by Güell et al. (2018) and our measures of income mobility (AUM, Q1Q5, RRS and IGE). The Spearman rank correlation index is negative and highly statistically significant for all measures.[48] This is good news for the ICS index. A closer inspection, however, reveals some shortcomings. Figure 11, which

[45] The top income share instead is not significantly correlated with relative mobility in Chetty et al. (2014).

[46] This indicator is effective only when applied to the subpopulation with rare last names. In our analysis, we use ICS indices built on last names with less than 15, 20, 25, and 30 individuals per province.

[47] Güell et al. (2018) use data from a single cross section of official tax returns in Italy for the year 2005 that contain surnames. The origin of these data is peculiar. They appeared online on the website of the Italian Ministry of Finance on April 30, 2008. This act was supposed to be part of a general strategy to fight tax evasion through social stigma. The Italian Privacy Authority quickly ordered the Ministry to take down the website, but at that point the data had already been downloaded and became publicly available.

[48] We have excluded the three Italian provinces that are officially bilingual and where many last names are French (Aosta) or German (Bolzano and Trento) because it is likely that significant shares of families with the same last names reside in France, Germany, or Austria. In any case, correlations are very robust to their inclusion.

Table 17—Absolute Upward Mobility and Income Inequality

	AUM	AUM	AUM	AUM	AUM
Gini coefficient	−0.416				
	(0.0875)				
Gini bottom 99 percent		−0.613			
		(0.0760)			
Top 1 percent income share			0.552		
			(0.0802)		
90-10 ratio				−0.650	
				(0.0731)	
SD log income					−0.649
					(0.0732)
R^2	0.173	0.376	0.305	0.423	0.421
Observations	110	110	110	110	110

Notes: Standard errors are in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

contains a scatterplot of standardized measures of ICS with AUM and RRS by province, reveals that the statistical relationship between the ICS and more conventional indicators of mobility is strong only for provinces with high values of the ICS index. The correlation between these two measures for areas with ICS below the mean is essentially zero. In other words, the informational content carried by the ICS about true mobility is very limited when the ICS is low: according to Figure 11, the ICS is a poor proxy of true income mobility in at least half of Italian provinces. In light of this last finding, we conclude that researchers should be cautious when using this index.

VII. What Correlates with Upward Mobility?

The goal of this section is to take a first step toward understanding what local characteristics can account for the divergence in upward mobility across Italian provinces that we documented in Section VI. We do not claim that the correlations we uncover should be interpreted as causal relations, but they surely serve to guide future research on the deeper determinants of intergenerational mobility. A similar analysis has been recently performed by Chetty et al. (2014) for the United States and by Güell et al. (2018) for Italy, using the ICS indicator.[49]

We start from a large set of correlates based on the literature. The list includes numerous local socioeconomic indicators (Sistema di Indicatori Territoriali, or Local Indicator System) compiled by the National Statistical Institute (ISTAT) for seven broad categories: (i) productivity (e.g., value added per resident), (ii) criminal activity of various types (e.g., scams, protested checks, drug offenses, thefts, murders), (iii) family instability (e.g., separations, divorces, children in custody), (iv) labor market conditions (e.g., unemployment rate, labor force participation rate),

[49] Compared to Güell et al. (2018), we also include indicators of school quality that turn out to be very important. Moreover, they only report unconditional correlations, while we go beyond that. Finally, as just explained, the ICS is weakly correlated with true mobility for many provinces.

Table 18—Rank-Rank Slope and Income Inequality

	RRS	RRS	RRS	RRS	RRS
Gini coefficient	0.509				
	(0.0829)				
Gini bottom 99 percent		0.608			
		(0.0764)			
Top 1 percent income share			−0.245		
			(0.0933)		
90-10 ratio				0.639	
				(0.0741)	
SD log income					0.622
					(0.0753)
R^2	0.259	0.370	0.060	0.408	0.387
Observations	110	110	110	110	110

Notes: Standard errors are in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

(v) life expectancy, (vi) openness to trade and migration, and (vii) social capital (e.g., the indicators proposed by Guiso, Sapienza, and Zingales 2004, 2016).

To these, we add indicators of educational attainment and a vast array of markers of school quality obtained from a national report on the performance of the local school system (Tuttoscuola 2007a). This set of indicators is organized by school level and by broad category. School levels are four: preschool, primary school, middle school, and high school. Categories are school resources and structures (e.g., local government spending in education as a share of total spending), school organization and services (e.g., transportation, extended teaching-time availability, administrative efficiency, students per class), teachers' composition (e.g., teaching hours, teachers' age and gender, teachers on temporary contracts), students' test and examination scores.[50] A detailed description of our set of raw indicators is provided in Tables A16 and A17 in the online Appendix.

We begin by documenting unconditional cross-province correlations between two measures of intergenerational mobility, AUM and Q1Q5, and these socioeconomic markers. Tables 21, 22, and 23 summarize the results for a subset of these indicators. In general, most variables correlate with mobility indices with the expected sign, and those that do not are insignificant or mildly significant. The correlation with measures of economic and labor market conditions (in particular youth unemployment, highly skilled employment rate, and labor force participation) is very strong. The educational attainment of residents in a province is also positively associated with upward mobility. Among the social capital indicators, the self-efficacy index is strongly correlated with upward mobility.[51] Crime and life expectancy statistics do

[50] The Tuttoscuola report is based on the geographical partition of Italy into 103 provinces, which had been adopted until 2001. Moreover, quality-of-schools indicators from this report exclude the provinces of Aosta, Bolzano, and Trento due to limited data availability. As a result, whenever we use these school quality indicators, our analysis is limited to 100 provinces.

[51] This is an index measuring beliefs in one's own ability to complete tasks and reach goals among children that Guiso, Sapienza, and Zingales (2016) take as a proxy of a local culture of individual empowerment.

Table 19—Intergenerational Elasticity and Income Inequality

	IGE	IGE	IGE	IGE	IGE
Gini coefficient	0.222				
	(0.0938)				
Gini bottom 99 percent		0.282			
		(0.0923)			
Top 1 percent income share			−0.160		
			(0.0950)		
90-10 ratio				0.310	
				(0.0915)	
SD log income					0.292
					(0.0920)
R^2	0.049	0.080	0.025	0.096	0.085
Observations	110	110	110	110	110

Notes: Standard errors are in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

Table 20—Cross-Province Correlation of ICS Indices from Güell et al. (2018) with Mobility Measures Estimated on Our Data

	ICS-15	ICS-20	ICS-25	ICS-30
AUM	−0.739	−0.727	−0.709	−0.707
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Q1Q5	−0.652	−0.627	−0.609	−0.595
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RRS	0.659	0.654	0.642	0.665
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
IGE	0.419	0.430	0.421	0.452
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

not show a systematic strong association. The fraction of foreign-born and population inflows show a tight positive correlation with upward mobility. Finally, many of the measures of school quality—such as available resources, students per class, share of young teachers, and test scores—are closely associated with upward mobility.[52] Adding macro area fixed effects reduces substantially the values of the correlations, and many of them—in particular, a number of school quality indicators—lose their significance.

A limitation of these unconditional associations is that all socioeconomic variables are strongly correlated among each other. We therefore move to a multivariate conditional correlation analysis. Given the sheer number of possible covariates, we begin by reducing the number of variables into a small number of principal components for each category, which capture a significant portion of the variation of the original variables in the category. Each principal component that we retain is then

[52] It may appear surprising that the share of teachers with temporary contracts shows such robust positive correlation. The reason is that it is very collinear with the share of young teachers: the correlation between the two variables is 0.74 for preschool teachers, and it decreases monotonically to 0.33 for high school teachers but remains statistically significant at the 1 percent level.

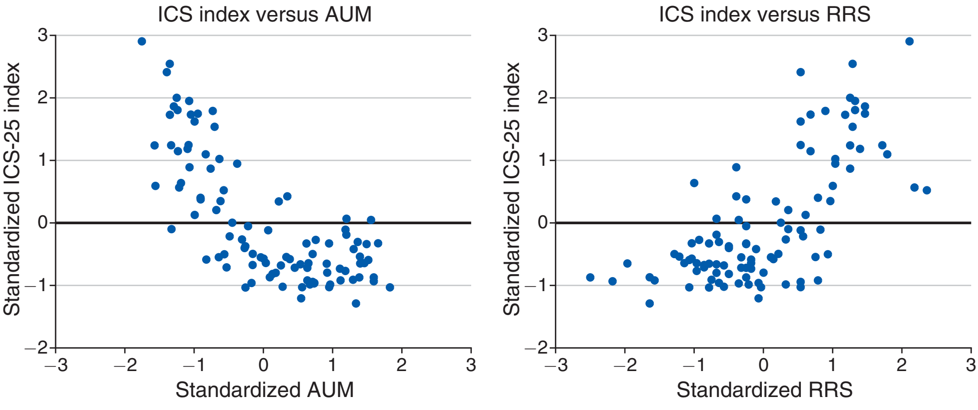

Figure 11. Cross-Province Relationship between the ICS-25 Index and the AUM Index (Left Panel) and between the ICS-25 Index and the RRS (Right Panel)

weighted in the index in proportion to the overall variance it explains. We retain four principal components for school quality, two principal components for crime, and one principal component for all other indicators considered (family instability, labor market conditions, life expectancy, economic openness, social capital, educational attainment).[53] We also include a measure of productivity.[54]

Table 24 presents the results of the multivariate regression of Absolute Upward Mobility on the factors we construct for each dimension of interest. Overall, the included categories explain a very large portion of the variation of AUM (between 85 and 90 percent). The state of the local labor market is the factor with the strongest correlation with AUM. This factor loads positively on various measures of occupation and labor force participation and negatively on measures of unemployment. The second variable in terms of strength of correlation with AUM is school quality. When disaggregating this factor into subcategories, we find that the factor summarizing school organization and services is the one that more strongly correlates with AUM. This factor loads positively on the number of students using canteen, school bus, and other services; extended teaching-time availability; and availability of teachers' rankings, and negatively on the number of pupils per class. Students' test scores is also highly significant. When disaggregating the school quality index into school levels, the strongest effect is found for preschool and primary school quality.

With respect to other correlates, family instability retains a sizable correlation, while crime has no systematic relationship with upward mobility.[55] The social capital factor (which loads positively on measures of blood donation, number of nonprofit organizations, voter turnout in the election of the Italian House of

[53] When constructing the disaggregated indices for school quality, we retain one principal component for school resources and structures, school organization and services, teachers' composition, students' grades and test scores, while we retain three principal components when constructing the indices of school quality by school level (preschool, primary school, middle school, and high school).

[54] Tables A18–A22 in the online Appendix describe the underlying markers used to construct each index in this conditional correlation analysis and the source of each marker.

[55] Somewhat puzzling, life expectancy remains significant but not with the expected sign, although it loses significance when controlling for macro area fixed effects.

Table 21—Unconditional Correlation of AUM and Q1Q5 with Various Indicators (Part 1/3)

	AUM baseline	AUM macro area FE	Q1Q5 baseline	Q1Q5 macro area FE
Productivity				
Value added per resident	0.666	0.142	0.604	0.195
	(0.0000)	(0.0112)	(0.0000)	(0.0123)
Manufacturing share of value added	0.688	0.261	0.564	0.171
	(0.0000)	(0.0000)	(0.0000)	(0.0106)
Public works started	0.388	0.124	0.483	0.262
	(0.0000)	(0.0061)	(0.0000)	(0.0000)
Public works completed	0.266	0.107	0.411	0.274
	(0.0055)	(0.0112)	(0.0000)	(0.0000)
Labor market				
Unemployment rate	−0.811	−0.310	−0.653	−0.205
	(0.0000)	(0.0001)	(0.0000)	(0.0646)
Youth unemployment rate (age 15–24)	−0.849	−0.406	−0.692	−0.300
	(0.0000)	(0.0000)	(0.0000)	(0.0076)
Long-term unemployment rate	−0.772	−0.219	−0.625	−0.163
	(0.0000)	(0.0032)	(0.0000)	(0.1202)
Employment rate (college degree or higher)	0.753	0.210	0.704	0.271
	(0.0000)	(0.0012)	(0.0000)	(0.0027)
Labor force participation	0.771	0.199	0.643	0.183
	(0.0000)	(0.0051)	(0.0000)	(0.0669)
Female labor force participation	0.746	0.167	0.606	0.122
	(0.0000)	(0.0167)	(0.0000)	(0.2128)
Educational attainment				
Share of illiterates	−0.435	0.021	−0.353	−0.007
	(0.0000)	(0.6756)	(0.0001)	(0.9053)
Education level achieved	0.526	0.091	0.496	0.175
	(0.0000)	(0.0663)	(0.0000)	(0.0042)
School dropouts	−0.295	−0.032	−0.159	0.034
	(0.0020)	(0.4896)	(0.1008)	(0.6012)
Social capital				
Blood bags collected per resident (GSZ 2004)	0.481	0.068	0.451	0.111
	(0.0000)	(0.1591)	(0.0000)	(0.0952)
Self-efficacy measure (GSZ 2016)	0.607	0.113	0.617	0.265
	(0.0000)	(0.0348)	(0.0000)	(0.0003)
Voter turnout, House of Representatives	0.722	0.205	0.545	0.099
	(0.0000)	(0.0018)	(0.0000)	(0.2895)
Voter turnout, European Parliament	0.485	0.107	0.321	0.018
	(0.0000)	(0.0363)	(0.0007)	(0.8022)
Recycling to total waste ratio	0.798	0.304	0.706	0.305
	(0.0000)	(0.0000)	(0.0000)	(0.0013)

Notes: With and without macro area fixed effects. p-values are in parentheses.

Table 22—Unconditional Correlation of AUM and Q1Q5 with Various Indicators (Part 2/3)

	AUM baseline	AUM macro area FE	Q1Q5 baseline	Q1Q5 macro area FE
Crime				
Thefts	0.275	−0.103	0.272	−0.052
	(0.0040)	(0.0245)	(0.0044)	(0.4175)
Violent crimes	−0.106	−0.035	−0.066	−0.022
	(0.2763)	(0.4169)	(0.4999)	(0.7084)
Distraints	−0.238	−0.049	−0.197	−0.007
	(0.0137)	(0.2624)	(0.0415)	(0.9078)
Scam offenses	0.520	0.157	0.508	0.188
	(0.0000)	(0.0007)	(0.0000)	(0.0037)
Life expectancy				
Female life expectancy at birth	0.473	0.146	0.346	0.113
	(0.0000)	(0.0033)	(0.0002)	(0.1047)
Male life expectancy at birth	−0.103	0.089	−0.172	0.055
	(0.2918)	(0.0636)	(0.0760)	(0.4072)
Old-age index (residents above age 65/below age 15)	0.448	−0.174	0.274	−0.289
	(0.0000)	(0.0019)	(0.0041)	(0.0002)
Number of suicides	0.409	−0.048	0.353	−0.046
	(0.0000)	(0.3307)	(0.0002)	(0.5040)
Family instability				
Divorce rate	0.577	−0.216	0.488	−0.223
	(0.0000)	(0.0010)	(0.0000)	(0.0155)
Children in custody due to divorce	0.520	−0.206	0.420	−0.235
	(0.0000)	(0.0006)	(0.0000)	(0.0054)
Separation rate	0.564	−0.115	0.466	−0.113
	(0.0000)	(0.0744)	(0.0000)	(0.2086)
Children in custody due to separation	0.500	−0.137	0.383	−0.185
	(0.0000)	(0.0200)	(0.0000)	(0.0236)
Openness				
Trade (exports + imports)	0.347	0.113	0.411	0.205
	(0.0002)	(0.0073)	(0.0000)	(0.0004)
Net interprovince migration	0.671	0.084	0.493	−0.022
	(0.0000)	(0.2255)	(0.0000)	(0.8187)
Foreign-born residents	0.786	0.318	0.646	0.303
	(0.0000)	(0.0000)	(0.0000)	(0.0024)
Inflow of graduates from other province or abroad	0.591	0.179	0.553	0.266
	(0.0000)	(0.0003)	(0.0000)	(0.0001)

Notes: With and without macro area fixed effects. p-values are in parentheses.

Representatives, and self-efficacy) has no longer any statistically significant effect. This appears to be due to the strong correlation between the social capital and state of the labor market indices. Excluding the measure of labor market conditions, social capital regains a positive and statistically significant relationship with AUM.

VIII. Conclusions

Our paper is the first to estimate intergenerational income mobility in Italy. None of the publicly available surveys for Italy span two generations. For this reason, thus far, the literature on intergenerational mobility in Italy has used other socioeconomic

Table 23—Unconditional Correlation of AUM and Q1Q5 with Various Indicators (Part 3/3)

	AUM baseline	AUM macro area FE	Q1Q5 baseline	Q1Q5 macro area FE
Quality of schools				
Local government spending in education	0.096	0.006	0.124	0.035
	(0.3196)	(0.8945)	(0.1756)	(0.5065)
School assets	0.266	−0.014	0.211	−0.028
	(0.0053)	(0.7550)	(0.0198)	(0.6171)
Availability of teaching materials and technologies	0.280	0.064	0.296	0.107
	(0.0033)	(0.1405)	(0.0009)	(0.0478)
Quality of school buildings	0.394	0.030	0.303	−0.006
	(0.0000)	(0.5206)	(0.0007)	(0.9231)
Students using canteen, school bus, and other services	0.689	0.189	0.565	0.182
	(0.0000)	(0.0017)	(0.0000)	(0.0164)
Students per class	−0.707	−0.208	−0.576	−0.205
	(0.0000)	(0.0005)	(0.0000)	(0.0072)
Extended teaching-time availability	−0.053	0.005	0.067	0.094
	(0.5856)	(0.9158)	(0.4615)	(0.0759)
Processing time of teachers' rankings	0.137	−0.104	0.106	−0.106
	(0.1578)	(0.0230)	(0.2472)	(0.0656)
Teaching hours per class	−0.238	−0.097	−0.124	−0.048
	(0.0127)	(0.0265)	(0.1745)	(0.3912)
Teachers below 40 years old: Preschool	0.641	0.123	0.605	0.188
	(0.0000)	(0.0302)	(0.0000)	(0.0083)
Teachers below 40 years old: Primary school	0.474	0.086	0.449	0.145
	(0.0000)	(0.0718)	(0.0000)	(0.0160)
Teachers below 40 years old: Middle school	0.320	0.065	0.346	0.087
	(0.0007)	(0.1619)	(0.0001)	(0.1373)
Teachers below 40 years old: High school	0.012	−0.006	0.106	0.023
	(0.9054)	(0.9060)	(0.2491)	(0.6978)
Teachers under temporary contracts: Preschool	0.663	0.197	0.557	0.212
	(0.0000)	(0.0002)	(0.0000)	(0.0017)
Teachers under temporary contracts: Primary school	0.739	0.168	0.651	0.252
	(0.0000)	(0.0108)	(0.0000)	(0.0021)
Teachers under temporary contracts: Middle school	0.439	0.036	0.357	0.052
	(0.0000)	(0.4556)	(0.0000)	(0.3886)
Teachers under temporary contracts: High school	0.392	−0.031	0.309	−0.044
	(0.0000)	(0.5211)	(0.0005)	(0.4592)
Tenure and stability of teachers' position	0.253	0.112	0.267	0.092
	(0.0080)	(0.0121)	(0.0029)	(0.1015)
Students repeating school year	0.177	0.087	0.079	0.007
	(0.0656)	(0.0396)	(0.3850)	(0.8910)
INVALSI test scores for primary and middle school	0.510	0.074	0.405	0.046
	(0.0000)	(0.1327)	(0.0000)	(0.4551)
INVALSI test scores for high school	0.558	0.109	0.437	0.074
	(0.0000)	(0.0309)	(0.0000)	(0.2477)

Notes: With and without macro area fixed effects. *p*-values are in parentheses.

outcomes such as education and occupation, other proxies such as the Informational Content of Surnames, or imputation procedures to obtain crude estimates of incomes for successive cohorts.

The micro data underlying our empirical analysis are obtained from an administrative database of tax returns where we link two generations through SSNs of

Table 24—Conditional Correlates of Absolute Upward Mobility

	AUM	AUM	AUM	AUM	AUM	AUM
Value added per resident	0.063	0.097	0.064	0.081	0.089	0.075
	(0.267)	(0.086)	(0.270)	(0.100)	(0.053)	(0.134)
Crime	0.061	0.009	0.083	0.049	−0.052	0.069
	(0.146)	(0.869)	(0.057)	(0.173)	(0.241)	(0.071)
Educational attainment	0.023	0.162	0.013	0.004	0.105	−0.015
	(0.631)	(0.010)	(0.809)	(0.917)	(0.045)	(0.747)
Family instability	−0.122	−0.108	−0.140	−0.237	−0.283	−0.259
	(0.046)	(0.069)	(0.032)	(0.000)	(0.000)	(0.000)
Strong labor market	0.666	0.543	0.597	0.398	0.248	0.379
	(0.000)	(0.000)	(0.000)	(0.000)	(0.021)	(0.001)
Life expectancy	−0.168	−0.144	−0.132	−0.040	−0.015	−0.033
	(0.000)	(0.004)	(0.008)	(0.387)	(0.737)	(0.476)
Economic openness	−0.041	−0.067	−0.032	−0.007	−0.084	−0.017
	(0.500)	(0.279)	(0.614)	(0.900)	(0.105)	(0.754)
Social capital	0.082	0.086	0.115	0.032	−0.014	0.070
	(0.291)	(0.279)	(0.149)	(0.638)	(0.828)	(0.321)
School quality	0.309			0.146		
	(0.000)			(0.052)		
Preschool quality		0.337			0.250	
		(0.001)			(0.003)	
Primary school quality		0.125			0.147	
		(0.054)			(0.008)	
Middle school quality		0.124			−0.016	
		(0.016)			(0.744)	
High school quality		−0.045			−0.063	
		(0.455)			(0.205)	
School structures and resources			−0.008			−0.022
			(0.878)			(0.601)
School organization and services			0.248			0.201
			(0.001)			(0.004)
Teachers' composition			0.085			0.020
			(0.115)			(0.688)
Students' grades and test scores			0.157			0.084
			(0.004)			(0.084)
Macro area fixed effects				Yes	Yes	Yes
Observations	100	99	100	100	99	100
Adjusted R^2	0.84	0.85	0.85	0.88	0.90	0.89

Notes: p-values are in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

parents and children. As rich as they are, the data are not perfect, and thus our analysis is not without caveats. Attenuation and life cycle biases due to the short panel dimension and misreporting of income due to tax evasion are the main threats to the credibility of our estimates. In the paper, we dealt as best as we could with these shortcomings, quantified their importance, and established that our main conclusions are quite robust.

Our findings contain some good news and some bad news. On the one hand, they paint a somewhat less pessimistic picture of intergenerational income mobility at the national level compared to many previous studies that represented Italy as a paralyzed society. Specifically, through a simulation exercise, we show that former analyses based on imputation procedures are likely to lead to understated estimates of national mobility. On the other hand, we revealed the presence of acute inequality in the degree of upward mobility within the country: the Northeast appears to be a land of equal and abundant opportunities and the South to be a land where ranks in society endure across generations. Compared to the United States, Italy has a similar share of children who overtake their parents in terms of income. This similarity, though, results from offsetting effects of more modest income growth and lower income inequality in Italy relative to the United States.

When we exploited within-country geographical variation to correlate a battery of socioeconomic indicators with our measures of upward mobility, we uncovered that the quality of early childhood education, children's test scores, indicators of family instability, and local labor market conditions have the strongest association to intergenerational mobility.

Another important result is that, even conditional on a particular percentile of the parental national income distribution, and within a province, economic outcomes of children are vastly different. Looking ahead, explaining this unobserved heterogeneity is one of the main challenges of this literature, and progress in this direction requires richer data on characteristics and choices of parents and children.

The availability of reliable estimates for Italy allows to add one important data point to cross-country analyses, as recently done, for example, by Alesina, Stantcheva, and Teso (2018) in their comparative study of intergenerational mobility and preferences for redistributive policies.

REFERENCES

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante. 2022. "Replication data for: And Yet It Moves: Intergenerational Mobility in Italy." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E151642V1.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108 (2): 521–54.

Alvaredo, Facundo, and Elena Pisano. 2010. "Top Incomes in Italy, 1974–2004." In *Top Incomes: A Global Perspective*, edited by Anthony Atkison and Thomas Piketty, 625–63. Oxford, UK: Oxford University Press.

Bank of Italy. 2019. "Survey on Household Income and Wealth, 07-10-2019 Historical Database (All Waves)." Bank of Italy, Rome (Italy). Multiple Electronic Files.

Barbieri, Teresa, Francesco Bloise, and Michele Raitano. 2018. "Intergenerational Earnings Inequality in Italy: New Evidence and Main Mechanisms." CIRET WP No. 1/2018.

Barone, Guglielmo, and Sauro Mocetti. 2016. "Intergenerational Mobility in the Very Long Run: Florence 1427–2011." Bank of Italy WP 1060.

Berman, Yonatan. 2017. "Understanding the Mechanical Relationship between Inequality and Intergenerational Mobility." Unpublished.

Berman, Yonatan. 2020. "The Long Run Evolution of Absolute Intergenerational Mobility." Unpublished.

Björklund, Anders, and Markus Jäntti. 1997. "Intergenerational Income Mobility in Sweden Compared to the United States." *American Economic Review* 87 (5): 1009–18.

Black, Sandra E., and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." In *Handbook of Labor Economics*, Vol. 4 (Part B), edited by Orley Ashenfelter and David Card, 1487–1541. Amsterdam: North Holland Press.

Blundell, Jack, and Erling Risa. 2019. "Income and Family Background: Are We using the Right Models?" Unpublished.

Boeri, Tito, Andrea Ichino, Enrico Moretti, and Johanna Posch. n.d. "Data on Province-level Consumer Price Index." Unpublished Data (accessed February 22, 2018).

Boeri, Tito, Andrea Ichino, Enrico Moretti, and Johanna Posch. 2018. "National Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces." Unpublished.

Boserup, Simon Halphen, Wojciech Kopczuk, and Claus Thustrup Kreiner. n.d. "Data on Binned Scatterplot for Denmark Provided in Meneses (2020)." Unpublished Data (accessed July 8, 2021).

Boserup, Simon Halphen, Wojciech Kopczuk, and Claus Thustrup Kreiner. 2017. "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations." Unpublished.

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

Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage. 2005. "Intergenerational Earnings Mobility in Norway: Levels and Trends." *Scandinavian Journal of Economics* 107 (3): 419–35.

Braun, Sebastian Till, and Jan Stuhler. 2017. "The Transmission of Inequality across Multiple Generations: Testing Recent Theories with Evidence from Germany." *Economic Journal* 128 (609): 576–611.

Cannari, Luigi, and Giovanni D'Alessio. 2018. "Education, Income, and Wealth: Persistence across Generations in Italy." Questioni di Economia e Finanza Occasional Paper No. 476.

Cannari, Luigi, and Giovanni Iuzzolino. 2009a. "Consumer Price Levels in Northern and Southern Italy." Bank of Italy Occasional Paper No. 49.

Cannari, Luigi, and Giovanni Iuzzolino. 2009b. "Data on Regional Consumer Price Indices for Italy." Bank of Italy.

Cannari, Luigi, Vieri Ceriani, and Giovanni D'Alessio. 1997. "Il Recupero degli Imponibili Sottratti a Tassazione." Bank of Italy Ricerche Quantitative per la Politica Economica.

Checchi, Daniele, Andrea Ichino, and Aldo Rustichini. 1999. "More Equal but Less Mobile?: Education Financing and Intergenerational Mobility in Italy and in the US." *Journal of Public Economics* 74 (3): 351–93.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017a. "The Fading American Dream: Trends in Absolute Income Mobility Since 1940." *Science* 356 (6336): 398–406.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017b. "Replication data and code for: The Fading American Dream: Trends in Absolute Income Mobility Since 1940." American Association for the Advancement of Science [publisher], Opportunity Insights [distributor]. https://opportunityinsights.org/data/.

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

Corak, Miles. 2017. "Divided Landscapes of Economic Opportunity: The Canadian Geography of Intergenerational Income Mobility." HCEO Working Paper No. 2017-043.

Corak, Miles, and Andrew Heisz. 1999. "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data." *Journal of Human Resources* 34 (3): 504–33.

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

Deutscher, Nathan, and Bhashkar Mazumder. 2019. "Intergenerational Mobility across Australia and the Stability of Regional Estimates." Unpublished.

Fiorio, Carlo V., and Francesco D'Amuri. 2005. "Workers' Tax Evasion in Italy." *Giornale degli Economisti e Annali di Economia* 64 (2/3): 247–70.

Güell, Maia, José V. Rodríguez Mora, and Christopher I. Telmer. 2015. "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating." *Review of Economic Studies* 82 (2): 693–735.

Güell, Maia, Michele Pellizzari, Giovanni Pica, and José Vicente Rodríguez Mora. n.d. "Data on Informational Content of Surnames (ICS) Estimates of Mobility." Unpublished Data (accessed June 18, 2015).

Güell, Maia, Michele Pellizzari, Giovanni Pica, and José Vicente Rodríguez Mora. 2018. "Correlating Social Mobility and Economic Outcomes." *Economic Journal* 128 (612): F353–F403.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. n.d. "Data on Indicators of Social Capital." Unpublished Data (accessed July 8, 2015).

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "The Role of Social Capital in Financial Development." *American Economic Review* 94 (3): 526–56.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-Term Persistence." *Journal of the European Economic Association* 14 (6): 1401–36.

Haider, Steven, and Gary Solon. 2006. "Life cycle Variation in the Association Between Current and Lifetime Earnings." American Economic Review 96 (4): 1308–20.

Heidrich, Stefanie. 2017. "Intergenerational Mobility in Sweden: A Regional Perspective." *Journal of Population Economics* 30 (4): 1241–80.

Hoffmann, Eran B., and Davide Malacrino. 2019. "Employment Time and the Cyclicality of Earnings Growth." *Journal of Public Economics* 169: 160–71.

ISTAT. n.d. "Sistema di Indicatori Territoriali." Istat—Istituto Nazionale di Statistica, Rome (Italy). Multiple Electronic Files (accessed July 9–10, 2015).

Jäntti, Markus, and Stephen P. Jenkins. 2015. "Income Mobility." In Handbook of Income Distribution, Vol. 2, 807–935. Amsterdam: Elsevier.

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren. 2015. "Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations." *Journal of Human Resources* 50 (1): 1–33.

Marino, M. Rosaria, and Roberta Zizza. 2012. "Personal Income Tax Evasion in Italy: An Estimate by Taxpayer Type." In *Tax Evasion and the Shadow Economy*, edited by Michael Pickhardt and Aloys Prinz. Northampton, MA: Edward Elgar Publishing.

Mazumder, Bhashkar. 2016. "Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data." In *Inequality: Causes and Consequences*, edited by Lorenzo Cappellari, Solomon W. Polachek, and Konstantinos Tatsiramos, 83–129. Bingley: UK: Emerald Group Publishing Limited.

Meneses, Francisco. 2020. "Analyzing the Roadblocks of Intergenerational Mobility: A Year-to-Year Analysis of a National Cohort of Students." Unpublished.

Ministero dell'Economia e delle Finanze. 2021. "Electronic Database of Personal Income Tax Returns." Dipartimento delle Finanze del Ministero dell'Economia e delle Finanze, Rome (Italy). Multiple Electronic Files.

Mocetti, Sauro. 2007. "Intergenerational Earnings Mobility in Italy." *BE Journal of Economic Analysis and Policy* 7 (2).

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.

Nybom, Martin, and Jan Stuhler. 2017. "Biases in Standard Measures of Intergenerational Income Dependence." *Journal of Human Resources* 52 (3): 800–25.

OECD. 2010. "Main Economic Indicators—Complete Database." Main Economic Indicators (Database). http://dx.doi.org/10.1787/data-00052-en (accessed July 14, 2021).

Opportunity Insights. n.d. "Geography of Mobility: National Mobility Statistics by Parent or Child Income Percentile." https://opportunityinsights.org/data/ (accessed on June 24, 2016).

Piraino, Patrizio. 2007. "Comparable Estimates of Intergenerational Income Mobility in Italy." *BE Journal of Economic Analysis and Policy* 7 (2).

Pisati, Maurizio, and Antonio Schizzerotto. 2004. "The Italian Mobility Regime: 1985–97." In Social Mobility in Europe, edited by Richard Breen, 149–74. Oxford, UK: Oxford University Press.

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

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." *Handbook of Labor Economics* 3: 1761–1800.

Tuttoscuola. 2007a. "Primo Rapporto Sulla Qualita' della Scuola." Editoriale Tuttoscuola.

Tuttoscuola. 2007b. "Data on Educational Attainment and School Quality." Editoriale Tuttoscuola.

U.S. Bureau of Economic Analysis. n.d. "Real Gross Domestic Product per Capita [A939RX0Q048SBEA]". FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/A939RX0Q048SBEA (accessed July 14, 2021).

World Bank. n.d. "Constant GDP per Capita for Italy [NYGDPPCAPKDITA]." FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/NYGDPPCAPKDITA (accessed July 14, 2021).