

"And Yet, It Moves": Intergenerational Mobility in Italy*

Paolo Acciari[†] Alberto Polo[§] Giovanni L. Violante[¶]

March 16, 2021

Abstract

We link tax returns across two generations to provide the first estimate of intergenerational income mobility in Italy based on administrative data. Italy emerges as less immobile than previously depicted by studies using proxies for economic status or survey data with imputation procedures. This conclusion is robust with respect to a number of common concerns when using administrative data. The expected rank of a child born from parents with income below the median is 0.44, and the probability that a child reaches the top quintile of the national income distribution starting from a family in the bottom quintile is 0.10. Upward mobility is higher for sons, first-born children, children of self-employed parents, and for those who migrate once adults. We uncover substantial geographical variation in the degree of upward mobility. Provinces in Northern Italy, the richest area of the country, display levels much higher than those in the South. This regional variation is strongly correlated with local labor market conditions, indicators of family instability, and measures of school quality.

Keywords: *Intergenerational Income Mobility, Inequality, Italy, Geographical Variation.*

JEL Codes: *J31, J61, J62, R1.*

*The opinions expressed in this working paper 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, and Enrico Moretti for sharing their data.

[†]Ministry of Economy and Finance of Italy

[§]Bank of England and CfM (LSE)

[¶]Princeton University, CEBI, CEPR, IFS, IZA and NBER

1 Introduction

Income mobility across generations is a key socio-economic 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 et al. \(1999\)](#) 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 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 et al. \(2015\)](#) 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 socio-economic factors are strongly correlated with upward mobility across regions, while controlling for the common institutional framework.

¹Some of these studies are [Mocetti \(2007\)](#), [Piraino \(2007\)](#), [Barbieri et al. \(2018\)](#) and [Cannari and D'Alessio \(2018\)](#).

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 two cohorts of Italians born, respectively, between 1942-1963 and 1979-1983 and we match parents and children through their social security numbers. Our final dataset contains nearly 1,720,000 parents-child pairs with detailed income information for three years in each cohort - 1998-99-00 for parents, 2016-17-18 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 45th 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%. For a child in the bottom quintile, the probability of rising to the top quintile is 11%. We also find that upward mobility is larger for sons, for first-born children, for children of self-employed parents and for children who, once adult, migrate to other regions within Italy. By exploiting variation within siblings, we conclude that around 20% of variation in child outcomes is attributable to parental background.

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 a differential in their fathers' earnings of around 36,700 Euros. An RRS of 0.22 means that, when adults, these children will be on average still two deciles apart, a gap that translates into nearly

²The 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).

6,100 Euros of annual earnings. The median rank-rank slope 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-squared 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 90th percentile, the bottom quarter of these children will be below the 35th percentile of their own national distribution. Among children from families at the 10th percentile, the top quarter of them will be above the 60th 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 zero, while above the 10th 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; (ii) possible distortions arising from tax evasion for the self-employed. Correcting for these sources of bias 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 which 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 Nordic European countries. We also use an alternative measure of mobility (namely, the probability that the son earns at least 50% more than the father) that accounts for the lower level of income inequality in Italy compared to the United States, and hence the shorter distance between percentiles. Through the lenses of this alternative index, we conclude that upward mobility in Italy is higher than in the United States only in the bottom tercile of the income distribution.

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 North East), 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 et al. \(2015\)](#) 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 which 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 what socio-economic 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 socio-economic 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% 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 2 defines the measures of intergenerational mobility used in the analysis. Section 3 describes the dataset, outlines the sample selection procedure, and provides some descriptive statistics. Section 4 discusses our findings on the degree of intergenerational mobility at the national level. Section 5 tackles potential sources of bias in our baseline estimates. Section 6 compares mobility outcomes in Italy to those in other countries. Section 7 documents the patterns of geographical variation in upward mobility across provinces in Italy. Section 8 explores the correlation between upward mobility and local characteristics of provinces that could account for geographical variation. Section 9 concludes.

2 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.

2.1 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

$$R_i = \alpha + \beta R_i^P + \varepsilon_i \quad (1)$$

where the constant α measures the expected rank of a child born from parents at the bottom of the income distribution ($R_i^P = 0$), 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

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 (β^{91-100}). Next, we run it on the bottom 90 pct and obtain β^{1-90} .

³By 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 then define

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

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.

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

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 pct differential in parental income translates into a 10 pct 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

2.2 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.

We follow [Chetty et al. \(2014\)](#) and report measures of absolute upward mobility (AUM),

⁴RRS 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, $SD(\log y_i) > SD(\log y_i^P)$, then the effect of parental income on child income is larger and this is reflected in a higher IGE.

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

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:

$$\bar{R}^{50} = \mathbb{E} [R_i | R_i^P \leq 50]. \quad (4)$$

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 25th percentile of the national income distribution, i.e., $\mathbb{E} [R_i | R_i^P = 25]$, which can be easily computed from (1) as $\alpha + 25 \cdot \beta = 50 - 25 \cdot \beta$.

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:

$$Q1Q5 = \Pr \{R_i > 80 | R_i^P \leq 20\}. \quad (5)$$

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

2.3 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

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

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 (\bar{R}_g^{50}) 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 needs 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:

$$\bar{R}_g^{50} = \alpha_g + 25 \cdot \beta_g. \quad (7)$$

This is the measure of absolute upward mobility that we use throughout our regional analysis. Using directly the 25th 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.

2.4 Indicators for cross-country comparisons

The key challenge in comparing the level of intergenerational mobility in Italy to that of other countries is that the national income distribution differs markedly across countries and such differences can obscure the interpretation of the results.⁶

To overcome this problem, when we compare Italy to another country we compute, at each parental income rank r , the probability that the income of the child (y) is higher than the income of the parents (y^P) by a factor $\Phi > 1$, once the two distributions have been rescaled to control for differentials in aggregate growth over the period. Specifically, at each r , we calculate

$$\Pr \left\{ y_i \geq \Phi y_i^P \cdot \left(\frac{\bar{y}}{\bar{y}^P} \right) \mid R_i^P = r \right\}. \quad (8)$$

By measuring upward mobility this way, we quell the valid concern that different degrees of income dispersion across countries may bias our findings.⁷

⁶Consider, for example, country A and country B , where country A is a lot less unequal than B . When using indices of positional mobility, such as the RRS, there is a mechanical force towards measuring more mobility in A than in B . The reason is that it takes a lower absolute change in income in country A to move up (or down) any given percentile compared to B , since in A the income distribution is more compressed.

⁷This measure has been used previously by Chetty et al. (2017) to compare mobility across cohorts.

3 Data

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

3.1 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.⁸ 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 or by 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).⁹

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).¹⁰ The income variables include total gross (before-tax) income and all its components, i.e. income from dependent labor, self-employment (divided

⁸In 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, corrects abnormally high values, etc. In 2012, 100% of tax returns were transmitted electronically to the tax authorities.

⁹Standard deductions such as allowances for children and dependent spouses are applied by the withholding agent.

¹⁰The 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.

into professional and entrepreneurial income), unemployment benefits, financial assets, housing, land and farms.^{11,12} Finally, we also have information on individual tax liabilities.

3.2 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 Appendix shows the corresponding section of the tax form.¹³

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 return 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.¹⁴ 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 (SHIW), 93% 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

¹¹Some income sources we do not observe are fellowships/scholarships, child/family benefits, social assistance transfers.

¹²The 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 free-lance journalist is considered, for tax purposes, a professional. An owner of a firm which 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).

¹³Specifically, 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.

¹⁴Even 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.

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.¹⁵

3.2.1 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 500 euros, (iv) with financial income exclusively from interest, dividends, and capital gains. Because of (i-iii), the very poor are not covered in our sample.¹⁶

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 5, we bring in additional data to deal with all these sources of bias.

¹⁵By 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.

¹⁶Individuals 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.

3.3 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% of the fathers are married. The data show that in 1998 the father is the top earner in nearly 90% of the families in the sample, and around 60% of families have two sources of income. The median gross income of fathers is nearly 19,000 Euros in 1998 and that of mothers is approximately half of it.¹⁷ 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% 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.¹⁸ However, daughters' income in 2017 is 70% 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% of children, compensation for dependent labor is the major component, followed by self-employment income for 17% of the child population. Fewer than 6% 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 Appendix. The table shows that, as expected, the income distribution is markedly right skewed: the top 1% of the distribution accounts for 6.5% of total income in 2017.¹⁹

3.4 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 (SHIW),

¹⁷Italy adopted the Euro in 2002, thus incomes for 1998-99-00 are expressed in Italian Liras in the database. We transformed them in Euros, the currency in which they are reported in 2016-17-18.

¹⁸Over 1998-2017 average annual nominal GDP growth per capita in Italy was 2%, but zero in real terms.

¹⁹In the Italian population, without the age restrictions we impose, this share is close to 9% in 2017. For comparison, in the U.S., the top 1% income share is around 19% in the same year.

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 fathers and mothers: these are the same cohorts of parents in our tax return data.²⁰ 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.

4 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 scatter plot of the mean percentile rank of children as a function of their parents' rank, $\mathbb{E} [R_i | R_i^P = r]$.

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 very bottom where it flattens out and at the very top where it bends upward.²¹

Running the OLS regression in (1), we estimate that a one percentage point (ppt) increase

²⁰We 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 et al. (1997), we drop individuals who *i*) report zero or missing income in the survey, *ii*) report housing asset income below 500 Euros, and *iii*) report only interest income. These criteria exclude from SHIW individuals who do not file tax returns.

²¹This 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 et al., 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.

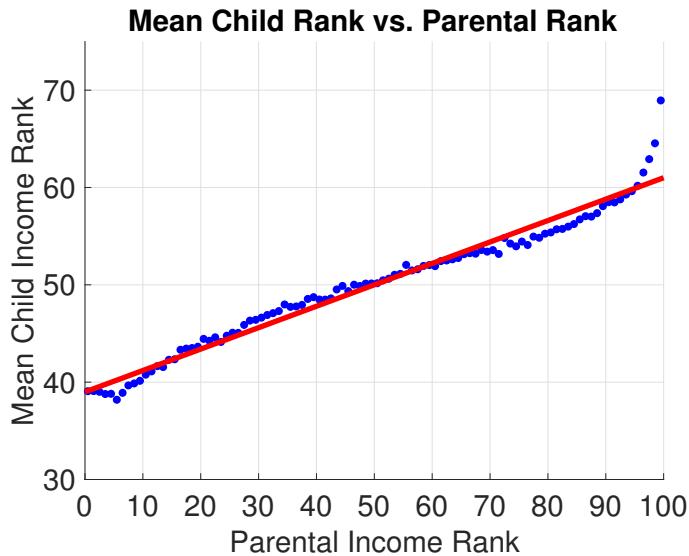


Figure 1: Child mean rank conditional on parental income rank. Blue dots: data. Red line: linear fit. The constant of the red line is 39 and the slope 0.22.

in parental rank is associated with a 0.22 ppt 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 90th percentile of the income distribution and one at the 10th 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 (TMR) is 4.8, a reflection of the fact that in Figure 1, the rank-rank curve steepens towards the upper end. Thus, ranks persist a lot more at the top of the income distribution.²² 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 45th percentile of her income distribution. The Q1Q5 index implies that, at the national level, children born in the bottom quintile of the income distribution

²²One 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.

face an 11 percent probability of belonging to the top quintile as adults.

Tables 7 and 8 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.²³

4.1 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 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 indices constructed by the Bank of Italy for the year 2006 ([Cannari and Iuzzolino, 2009](#)). We explore four different indices: the first one is a price index that allows for regional dispersion in prices of food, clothing, and furniture only. The other three indices 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 indices 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 indices is very high, ranging from 0.53 to 0.99.²⁴

Overall, these price indices 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 9 shows that our national results are quite robust to alternative deflating procedures. Our esti-

²³This 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% drops by more than half, from 0.13 to 0.06.

²⁴Table A2 in the Appendix reports all pairwise correlations between the indices.

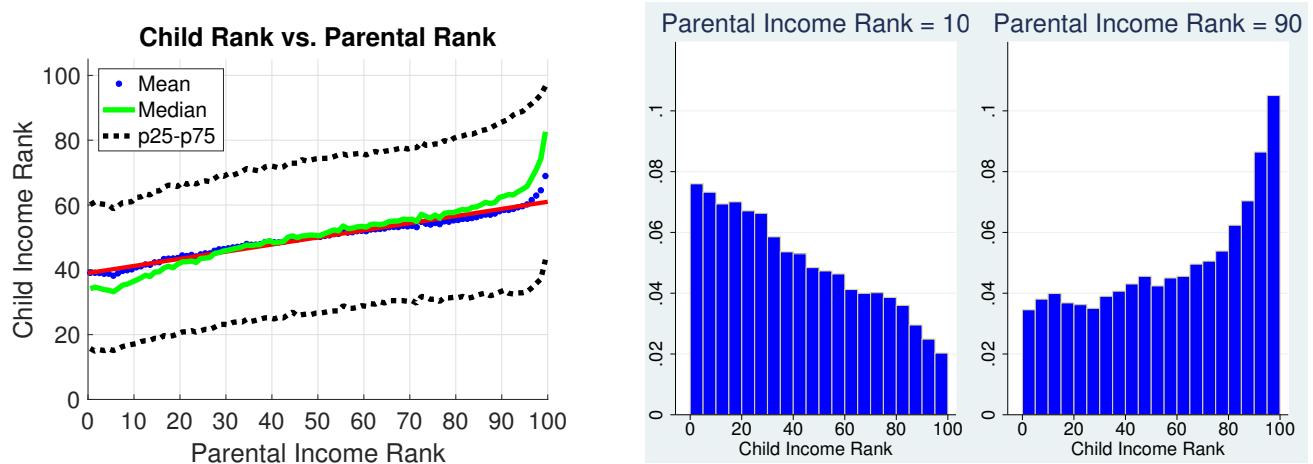


Figure 2: Left panel (a): mean child rank, median child rank, and interquartile range of child rank conditional on parental income rank. Right panel (b): conditional distributions of child rank at parental income percentiles 10 and 90.

mates of relative mobility rise somewhat when we use these regional deflators, but indices of absolute mobility are nearly identical.²⁵ In light of the robustness, in what follows, we keep using nominal income.

4.2 Median regression and conditional distributions of child ranks

Figure 2(a) 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 Figure 2(b). 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.

Figure 2(a) 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

²⁵A 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.

conditional distribution of child ranks, averaged across parents' ranks, is 0.48.²⁶ For example, at the 10th percentile of parental income, one quarter of children have incomes above the 60th percentile, and at the 90th percentile of the parents' distribution, one quarter of children have incomes below the 35th percentile.²⁷ 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% of the within-parental-rank variation.²⁸ 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 1998 (the location where the children grew up) accounts for much of the explained within-rank variation: between 1/5 and 1/2, with the exception of the very top of the parental income distribution. In Section 7, 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.

²⁶Interestingly, [Boserup et al. \(2017\)](#) also report for Denmark an interquartile range of the conditional distribution of child income ranks around 0.4.

²⁷The 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.

²⁸This 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.

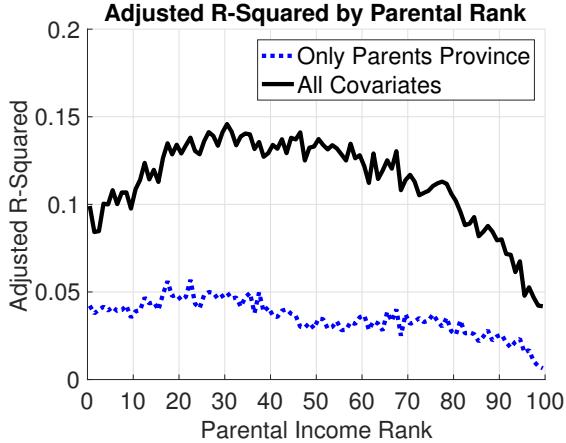


Figure 3: Adjusted R-squared of the 100 within-parental rank regressions of individual child rank in the national distribution on various parental and child covariates.

4.3 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 2. 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% 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 0.29.²⁹ The overall IGE is remarkably close to our estimate of the RRS.³⁰

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.

²⁹This sharp drop in curvature at the bottom of the income distribution emerges for the US as well ([Chetty et al., 2014](#)).

³⁰[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 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).

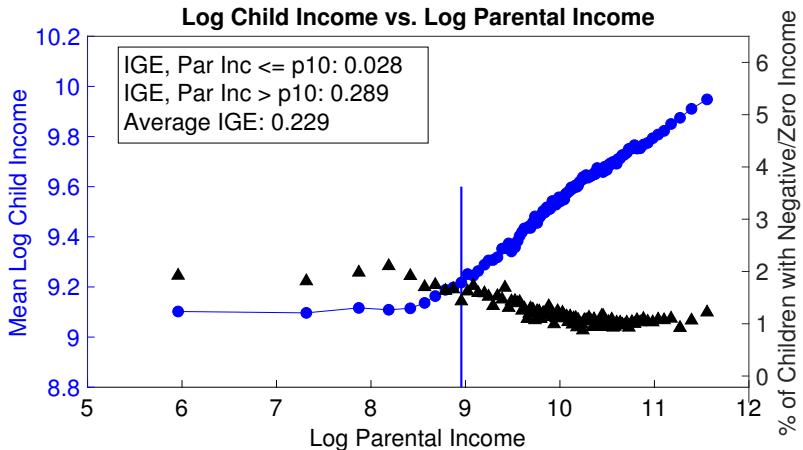


Figure 4: Relation between log income of the child and log income of the parents. The vertical line denotes the 10th percentile. The figure also reports the fraction of children with negative or zero income at each bin (i.e., the percentage of observations dropped).

4.3.1 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 et al., 2018](#); [Cannari and D'Alessio, 2018](#)).³¹

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.

³¹This gap between methodologies exists also for the U.S. For example, when [Björklund and Jäntti \(1997\)](#) use this same imputation procedure on U.S. 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.

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 0.50.³²

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.

4.4 Analysis for different population subgroups

Tables 3 to 5 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 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 pct in 2019).

When we condition on the major source of parental income (labor, self-employment, and capital) in Table 4, we find more pronounced upward mobility for the children of self-employed professionals (e.g., artisans, architects, lawyers, doctors, pharmacists): the Q1Q5 is 1.5 times larger than for children of wage earners.

Table 5 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 57th

³²It is also precisely estimated (S.E. = 0.006 based on 200 bootstrapped replications), given that our sample size is over 300 times larger than that of existing studies.

percentile, relative to a baseline value equal to the 45th 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).³³

4.5 Siblings: the roles of common parental background and birth order

Even though our dataset does not contain explicit identifiers for siblings, we can easily recover siblings by matching children that 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 two children.³⁴ 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 two children.

We exploit this sample of siblings to explore the role of common parental background vs. exposure to lifetime shocks for upward mobility. We find that the mean absolute rank differential between siblings is 0.27. Since the mean absolute difference of a uniform distribution is 0.33, we conclude that according to this metric parental background accounts for almost 1/5 of the total variation of individual outcomes.³⁵ This share is well within the range of the estimates reported by [Jäntti and Jenkins \(2015, Table 9\)](#) for a number of countries.

We go one step further into the analysis by asking whether birth order matters: is there any systematic difference between earlier-born and later-born children in terms of upward mobility?³⁶ Table 6 shows the results of our analysis. We find that the median rank of earlier-born

³³Tables A3 to A5 in the 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 (which is the father in 87% of the cases). Tables A6 to A9 in the 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.

³⁴By 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.

³⁵The more traditional estimate of the relative importance of parental background, *siblings'* income correlation, is 0.24 in our data and hence quite close.

³⁶Some of the children we label as ‘earlier born’ may have older siblings we do not observe because of our age

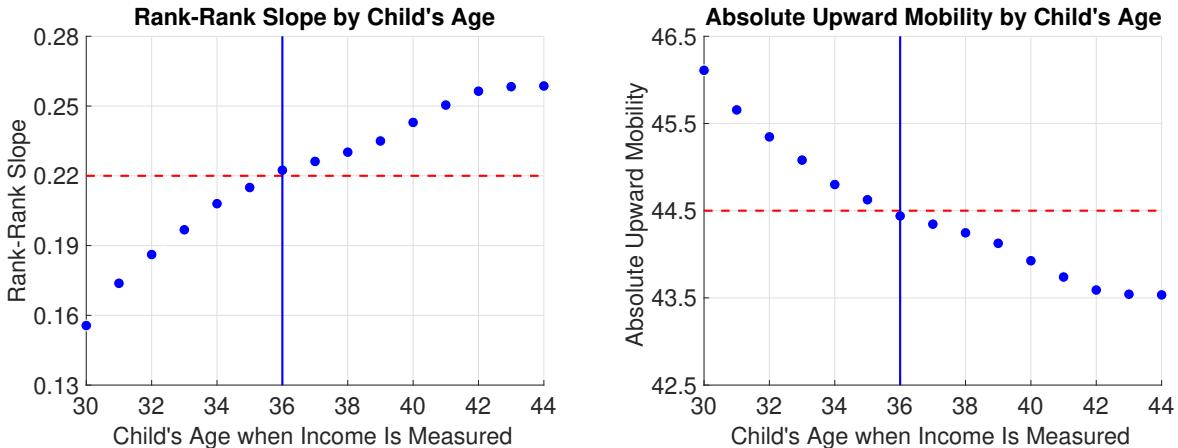


Figure 5: Estimates of RRS (left panel) and AUM (right panel) based on different 3-year age ranges for children. The x axis reports the midpoint of the age range measured in the year in which child income is observed. The vertical blue line indicates the midpoint for our sample, while the dashed red line indicates the estimate of RRS/AUM in the sample.

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.

5 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 few years of data available for each cohort and that we observe fathers and children at different ages; (ii) the omission of certain types of taxpayers from our sample and the under-reporting of income for those in the sample because of tax evasion.³⁷

5.1 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

restrictions on the sample. These cases, however, only arise in families with at least 3 children, which are only 6% of the total number of Italian families. Thus the vast majority of earlier born are first born.

³⁷In the 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.

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 which, 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 3-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 US and Sweden, the RRS levels off in the early 30s.³⁸ 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 this age profile, and assume a time-stationary persistent-transitory process for log residual income, a common representation in the literature on income dynamics:

$$\begin{aligned}\log y_{i,a}^j &= \kappa_{i,a}^j + \epsilon_{i,a}^j \\ \kappa_{i,a}^j &= \rho \kappa_{i,a-1}^j + \eta_{i,a}^j,\end{aligned}$$

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 $(\kappa_{i,0}^s, \kappa_{i,0}^f)$ of a son-father pair are drawn from a bivariate Normal distribution with correlation ϱ_0 . Table 10 reports

³⁸See Figure III.A in Chetty et al. (2014) and Figure 1c in Nybom and Stuhler (2017).

estimates for the structural model parameters $\{\rho, \sigma_\eta, \sigma_\varepsilon, \sigma_{\kappa_0}\}$.³⁹

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 three 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 $\varrho_0 = 0.39$ reproduces almost exactly our estimated values of the rank-rank slope, AUM and Q1Q5 for fathers and sons (Table A6 in the Appendix) when we use the simulated 3-year proxy for income. The last step is computing the 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% 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 made kindly available to us by Hoffmann and Malacrino (2019).⁴⁰ 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 3-year average earnings

³⁹In the Appendix we provide more details on the identification and estimation strategy.

⁴⁰INPS 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 non-farm private sector. The data represent a 6.6% sample of the population. The notion of income is gross labor earnings. See Hoffmann and Malacrino (2019) for details.

$y_{i,a}$ to lifetime average earnings \bar{y}_i :

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

where a denotes a 3-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 parameters estimates for this model.

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' 3-year average earnings $y_i^{[35,37]}$ on fathers' 3-year average earnings $y_i^{[47,49]}$ matches the corresponding estimate in our data (0.226), yielding $\lambda = 0.281$.

Tables 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.

5.2 Omission of certain taxpayers and tax evasion

5.2.1 Omission of taxpayers who filed form 770

In Section 3 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 which 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 770.⁴¹ 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.

5.2.2 Tax evasion

Italy is notoriously a country with high tax evasion. In recent years, the size of the non-observed 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 propensity 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.

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% for self-employment income and up to 80% 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% of reported gross income in the lowest decile to 8% 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,

⁴¹Specifically, this means looking for taxpayers who have only labor income and rental income below 568 Euros 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.).

if self-employed parents under-report income and their children do not, for example because they are dependent workers (or the other way around), mobility would be overestimated.

5.3 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 towards 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 770 taxpayers 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.⁴² We obtain a RRS of 0.30, an IGE of 0.31, an AUM index of 0.43 and a Q1Q5 of 0.09.

6 Cross-country comparison

How does this estimated level of intergenerational mobility in Italy compare to that estimated in other countries? We limit our analysis to studies which 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 et al., 2017). Schnelle (2015) reports a RRS of 0.19 for Norway. Deutscher and Mazumder (2019) measures a RRS of 0.21 and an AUM index of 0.45 for Australia. Heidrich (2017) estimates a RRS of 0.24 for Sweden. Corak (2017) reports a RRS of 0.24 and AUM index of 0.44 for Canada. The United States have 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.⁴³

As we argued in Section 2, these cross-country comparisons of positional mobility can be misleading when countries differ in their level of income inequality. The OECD Income Distribution Database reports Gini indices of gross income around 0.30 for Denmark, Norway, and

⁴²To 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.

⁴³A caveat, when comparing these measures, is that for many of these countries child income corresponds to family income, whereas for Italy it is individual income. Chetty et al. (2014) compute the RRS for both definitions of child income. Their estimates show that the RRS computed based on family income of the child is 20 pct higher than the RRS estimated on individual child income. Applying this correction coefficient to our estimate, the RRS for Italy would be closer to that of the United States.

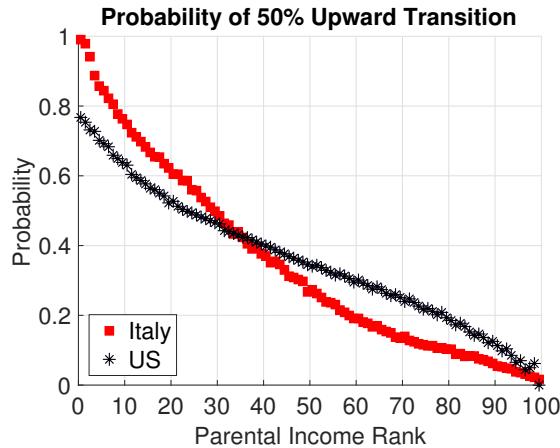


Figure 6: Probability that the income of the child is at least 50% higher than parental income. For the United States, income is family income for both children and parents. For Italy, it is individual income of sons and their fathers.

Sweden, around 0.36 for Canada and Italy and 0.44 for the United States. Higher inequality in the United States compared to Italy translates mechanically in higher rank persistence, but this does not mean that mobility in the United States is necessarily lower: children from US low income families can increase their income, relative to their parents, even more than their Italian counterparts, and yet climb fewer percentiles in the distribution compared to Italian children.

To circumvent this problem, we explore the behavior of the measure in equation (8), the probability that child income is at least 50 percent higher than parental income, after normalizing the parent and child distributions so that they have the same mean to correct for aggregate growth.

Figure 6 depicts this probability of upward transition for both the United States and Italy at each percentile of the distribution of father's income. The plot shows a stark pattern: this index of upward mobility is significantly higher for Italy among poor households. However, after the 35th percentile or so, thus for the middle class as well, upward mobility becomes stronger in the United States.

7 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.⁴⁴

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 4 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 4.1.⁴⁵

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-ten and bottom-ten provinces among the largest fifty 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 7 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 North-East 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 AUM among the four macro areas (North-East, North-West, Center, South and Islands), accounts

⁴⁴Over 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 Appendix contains a map of the Italian provinces and Tables A10-A12 list all the provinces, their population, region and macro area (North-East, North-West, 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.

⁴⁵We 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 indices we could use are all highly correlated. Figure A3 in the 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).

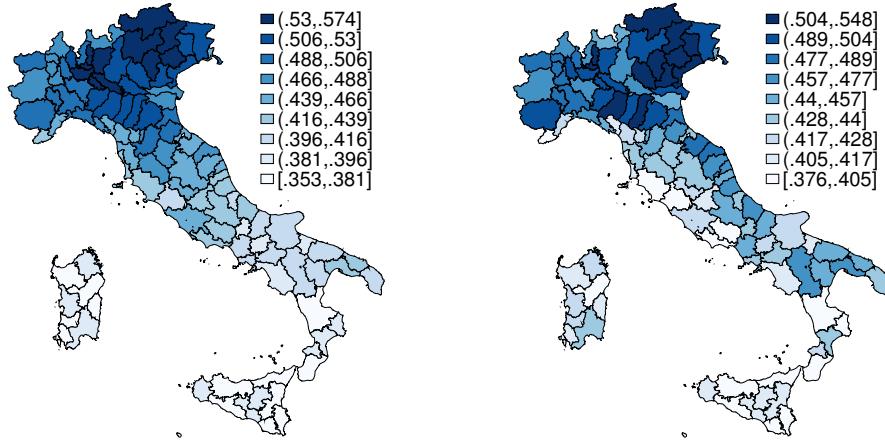


Figure 7: Heat map of Absolute Upward Mobility across provinces. Dark areas are more mobile. Left-panel (a): AUM computed on nominal income. Right-panel (b): AUM computed based on PPP-adjusted income (with regional price indices)

for 82% of the total variation of AUM across the 110 provinces. For example, the highest-ranked province for AUM not in the North-East or North-West is Firenze (Tuscany) at position 34/110. The lowest-ranked province for AUM not in the South and Islands macro-region is Latina (Lazio) at position 73/110.⁴⁶

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 7,200 Euros (nearly 40% of median annual income in 2017), and hence, it translates into substantial gaps in children’s lifetime incomes.

Panel (b) of Figure 7 reveals that, once adjusting for different cost-of-living levels across regions, the predominance of the North-East 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 8. The least mobile provinces have transition rates around 6-7% and the most mobile around 21-22%, i.e. larger by a factor of three. As evident from the heat maps, the correlation between AUM and Q1Q5 indicators is very strong

⁴⁶Tables A13-A15 in the Appendix show the ranking and all mobility measures – including RRS and IGE – for the 110 provinces.

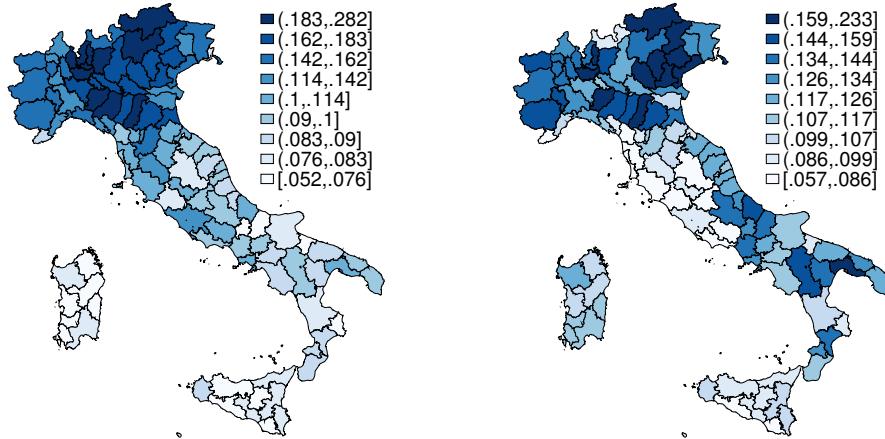


Figure 8: Heat map of Q1Q5 across provinces. Dark areas are more mobile. Left-panel (a): Q1Q5 computed on nominal income. Right-panel (b): Q1Q5 computed based on PPP-adjusted income (with regional price indices)

(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 Reggio-Calabria (Calabria) —the province with the highest RRS among the largest provinces by population — it translates into a gap twice as big, i.e. 22 percentiles. Large differences across provinces arise also in the IGE: the cross-province correlation between RRS and IGE is 0.73.

Figure 9 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). 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 Appendix display the rank-rank relationship for the top and bottom eight provinces.

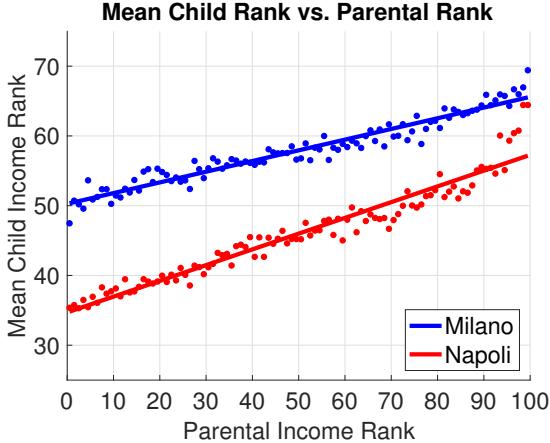


Figure 9: Child rank-parental rank relationship for children who grew up in Milano and Napoli. The RRS are respectively, 0.16 and 0.22, and AUM indicators are 0.54 and 0.41.

We conclude that areas in Northern Italy (especially the regions in the North East), 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.

Recall that, as explained in Section 2, while at the national level RRS and AUM are tightly linked, at the level of a province it needs 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 2, using the 25th 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.

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.

7.1 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% 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 corre-

lated with the AUM (Table V in Chetty et al. (2014)).⁴⁷ 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.

7.2 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 et al. (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.⁴⁸

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 socio-economic indicators.⁴⁹ 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

⁴⁷The top income share instead is not significantly correlated with relative mobility in Chetty et al. (2014).

⁴⁸This indicator is effective only when applied to the sub-population 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.

⁴⁹Güell et al. (2018) use data from a single cross-section of official tax returns in Italy for the year 2005 which contain surnames. The origin of these data is peculiar. They appeared online on the website of the Italian Ministry of Finance on 30 April 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.

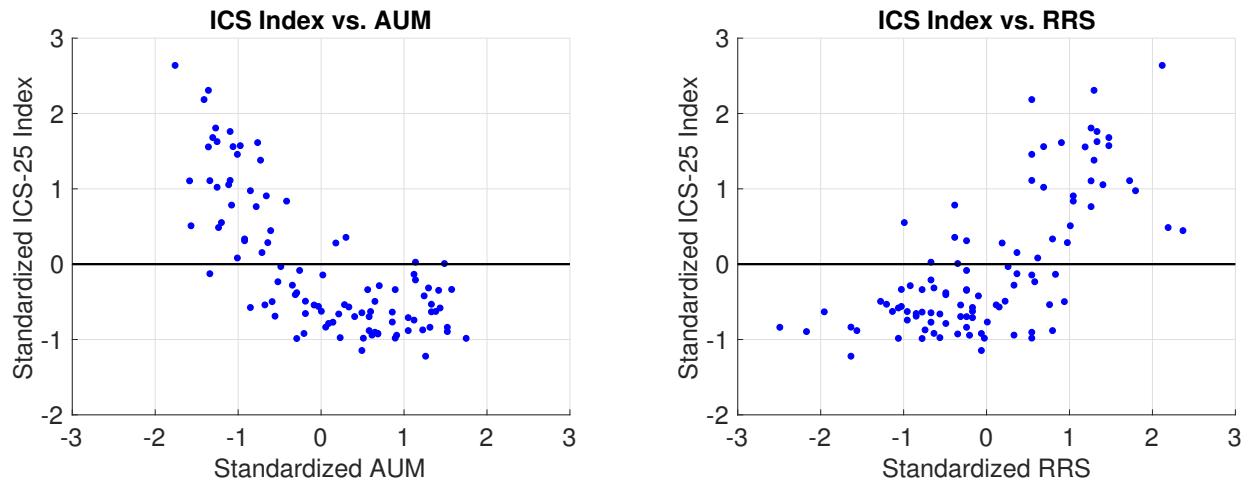


Figure 10: Cross-province relationship between the ICS-25 index and the AUM index (left panel) and between the ICS-25 index and the RRS (right panel).

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.⁵⁰ This is good news for the ICS index. A closer inspection, however, reveals some shortcomings. Figure 10, which 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 10, 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.

⁵⁰We have excluded the three Italian provinces which 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.

8 What correlates with upward mobility?

The goal of this section is to take a first step towards understanding what local characteristics can account for the divergence in upward mobility across Italian provinces that we documented in Section 7. 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.⁵¹

We start from a large set of correlates based on the literature. The list includes numerous local socio-economic 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); (v) life expectancy; (vi) openness to trade and migration; (vii) social capital (e.g., the indicators proposed by Guiso et al. (2004) and Guiso et al. (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, 2007). This set of indicators is organized by school level and by broad category. School levels are four: pre-school, 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.⁵² A detailed description of our set of raw indicators is provided in Tables A16 and A17 in the Appendix.

⁵¹Compared 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.

⁵²The 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.

We begin by documenting unconditional cross-province correlations between two measures of intergenerational mobility, AUM and Q1Q5, and these socio-economic 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.⁵³ Crime and life expectancy statistics do 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.⁵⁴ 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 socio-economic 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 weighted in the index in proportion to the overall variance it explains. We retain 4 principal components for school quality, 2 principal components for crime, and 1 principal component for all other indicators considered (family instability, labor market conditions, life expectancy, economic openness, social capital, educational attainment).⁵⁵ We

⁵³This is an index measuring beliefs in one's own ability to complete tasks and reach goals among children that Guiso et al. (2016) take as a proxy of a local culture of individual empowerment.

⁵⁴It 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 pre-school teachers and it decreases monotonically to 0.33 for high-school teachers but remains statistically significant at 1% level.

⁵⁵When constructing the disaggregated indices for school quality, we retain 1 principal component for school resources and structures, school organization and services, teachers' composition, students' grades and test scores, while we retain 3 to 4 principal components when constructing the indices of school quality by school level (pre-school, primary school, middle school, and high school).

also include a measure of productivity.⁵⁶

Table 24 presents the results of the multivariate regression of Absolute Upward Mobility (AUM) 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%). 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 sub-categories, we find that the factor summarizing school organization and services is the one which 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 pre-school and primary school quality.

With respect to other correlates, family instability retains a sizable correlation, while crime has no systematic relationship with upward mobility.⁵⁷ The social capital factor (which loads positively on measures of blood donation, number of non-profit organizations, voter turnout in the election of the Italian House of 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.

9 Conclusions

Our paper is the first to estimate intergenerational income mobility in Italy. None of the publicly available surveys for Italy spans two generations. For this reason, thus far, the literature on intergenerational mobility in Italy has used other socio-economic outcomes such as education and occupation, other proxies such as the Informational Content of Surnames, or imputation

⁵⁶Tables A18–A22 in the Appendix describe the underlying markers used to construct each index in this conditional correlation analysis, and the source of each marker.

⁵⁷Somewhat puzzling, life expectancy remains significant but not with the expected sign, although it loses significance when controlling for macro-area fixed effects.

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 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 North-East 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, once we adjust for the different levels of income inequality, Italy displays more upward mobility in the bottom tercile, but less everywhere else.

When we exploited within-country geographical variation to correlate a battery of socio-economic 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 et al. \(2018\)](#) in their comparative study of intergenerational mobility and preferences for redistributive policies.

References

- ALESINA, A., S. STANTCHEVA, AND E. TESO (2018): "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 108(2), 521–54.
- ALVAREDO, F., AND E. PISANO (2010): "Top Incomes in Italy, 1974–2004," in *Top Incomes: A Global Perspective*, ed. by A. Atkinson, and T. Piketty. Oxford University Press.
- BARBIERI, T., F. BLOISE, AND M. RAITANO (2018): "Intergenerational Earnings Inequality in Italy: New Evidence and Main Mechanisms," *CIRET WP*, No. 1.
- BARONE, G., AND S. MOCETTI (2016): "Intergenerational Mobility in the Very Long Run: Florence 1427–2011," *Bank of Italy WP*, 1060.
- BERMAN, Y. (2017): "Understanding the Mechanical Relationship between Inequality and Intergenerational Mobility," *SSRN* 2796563.
- BJÖRKLUND, A., AND M. JÄNTTI (1997): "Intergenerational Income Mobility in Sweden Compared to the United States," *The American Economic Review*, 87(5), 1009–1018.
- BLACK, S., AND P. DEVEREUX (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card. North Holland Press.
- BLUNDELL, J., AND E. RISA (2019): "Income and Family Background: Are We Using the Right Models?," *SSRN Working Paper* 3269576.
- BOERI, T., A. ICHINO, E. MORETTI, AND J. POSCH (2018): "National Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces," Discussion paper, University of California Berkeley.
- BOSERUP, S. H., W. KOPCZUK, AND C. T. KREINER (2017): "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations," Discussion paper, University of Copenhagen.
- BRATBERG, E., J. DAVIS, B. MAZUMDER, M. NYBOM, D. D. SCHNITZLEIN, AND K. VAAGE (2017): "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US," *The Scandinavian Journal of Economics*, 119(1), 72–101.
- BRAUN, S. T., AND J. STUHLER (2017): "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany," *The Economic Journal*, 128(609), 576–611.
- CANNARI, L., V. CERIANI, AND G. D'ALESSIO (1997): "Il Recupero degli Imponibili Sottratti a Tassazione," *Bank of Italy, Ricerche Quantitative per la Politica Economica*.
- CANNARI, L., AND G. D'ALESSIO (2018): "Education, Income and Wealth: Persistence Across Generations in Italy," *Bank of Italy, Questioni di Economia e Finanza, Occasional Paper* No. 476.
- CANNARI, L., AND G. IUZZOLINO (2009): "Consumer Price Levels in Northern and Southern Italy," *Bank of Italy, Occasional Paper* No. 49.

- CHECCHI, D., A. ICHINO, AND A. RUSTICHINI (1999): "More Equal but Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the US," *Journal of Public Economics*, 74(3), 351–393.
- CHETTY, R., D. GRUSKY, M. HELL, N. HENDREN, R. MANDUCA, AND J. NARANG (2017): "The Fading American Dream: Trends in Absolute Income Mobility Since 1940," *Science*, 356, 398–406.
- CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 129(4), 1553–1623.
- CORAK, M. (2017): "Divided Landscapes of Economic Opportunity: The Canadian Geography of Intergenerational Income Mobility," *University of Chicago, Human Capital and Economic Opportunity Working Paper*, No. 2017-043.
- CORAK, M., AND A. HEISZ (1999): "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data," *Journal of Human Resources*, pp. 504–533.
- DAHL, M. W., AND T. DELEIRE (2008): "The Association Between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data," Discussion paper, University of Wisconsin-Madison, Institute for Research on Poverty.
- DEUTSCHER, N., AND B. MAZUMDER (2019): "Intergenerational Mobility across Australia and the Stability of Regional Estimates," *Mimeo, Australian National University*.
- FIORIO, C. V., AND F. D'AMURI (2005): "Workers' Tax Evasion in Italy," *Giornale degli Economisti e Annali di Economia*, 64(2/3), 247–270.
- GÜELL, M., J. V. R. MORA, AND C. I. TELMER (2015): "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating," *The Review of Economic Studies*, 82(2), 693–735.
- GÜELL, M., M. PELLIZZARI, G. PICA, AND J. V. R. MORA (2018): "Correlating Social Mobility and Economic Outcomes," *The Economic Journal*, 128(July), F353–F403.
- GUISO, L., P. SAPIENZA, AND L. ZINGALES (2004): "The Role of Social Capital in Financial Development," *American Economic Review*, 94(3), 526–556.
- (2016): "Long-term Persistence," *Journal of the European Economic Association*, 14(6), 1401–1436.
- HAIDER, S., AND G. SOLON (2006): "Life-cycle Variation in the Association Between Current and Lifetime Earnings," *American Economic Review*, 96(4), 1308–1320.
- HEIDRICH, S. (2017): "Intergenerational Mobility in Sweden: A Regional Perspective," *Journal of Population Economics*, 30(4), 1241–1280.
- HOFFMANN, E. B., AND D. MALACRINO (2019): "Employment Time and the Cyclical of Earnings Growth," *Journal of Public Economics*, 169, 160–171.
- JÄNTTI, M., AND S. P. JENKINS (2015): "Income Mobility," in *Handbook of Income Distribution*, vol. 2, pp. 807–935. Elsevier.

- LINDAHL, M., M. PALME, S. S. MASSIH, AND A. 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. R., AND R. ZIZZA (2012): "Personal Income Tax Evasion in Italy: An Estimate by Tax-payer Type," in *Tax Evasion and the Shadow Economy*, ed. by M. Pickhardt, and A. Prinz. Edward Elgar Publishing.
- MAZUMDER, B. (2016): "Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data," in *Inequality: Causes and Consequences*, pp. 83–129. Emerald Group Publishing Limited.
- MOCETTI, S. (2007): "Intergenerational Earnings Mobility in Italy," *The BE Journal of Economic Analysis & Policy*, 7(2), 5.
- MORETTI, E. (2013): "Real Wage Inequality," *American Economic Journal: Applied Economics*, 5(1), 65–103.
- NYBOM, M., AND J. STUHLER (2017): "Biases in Standard Measures of Intergenerational Income Dependence," *Journal of Human Resources*, 52(3), 800–825.
- PIRAINO, P. (2007): "Comparable Estimates of Intergenerational Income Mobility in Italy," *The BE Journal of Economic Analysis & Policy*, 7.
- PISATI, M., AND A. SCHIZZEROTTO (2004): "The Italian Mobility Regime: 1985-97," in *Social Mobility in Europe*, ed. by R. Breen. Oxford University Press.
- SCHNELLE, K. (2015): "Intergenerational Mobility in Norway: Transition Probabilities and Directional Rank Mobility," *Mimeo, University of Bergen*.
- SOLON, G. (1992): "Intergenerational Income Mobility in the United States," *The American Economic Review*, pp. 393–408.
- (1999): "Intergenerational Mobility in the Labor Market," *Handbook of Labor Economics*, 3, 1761–1800.
- TUTTOSCUOLA (2007): "Primo Rapporto Sulla Qualita' della Scuola," Discussion paper, Editoriale Tuttoscuola.

Statistic	N	Value
Parents in 1998		
Mean father's age	1,600,529	48
Mean mother's age	1,196,230	45
% Married father	1,577,095	92.8%
% Father residing in North	1,600,529	42.8%
% Father residing in Center	1,600,529	19.6%
% Father residing in South	1,600,529	37.6%
% Father residing in same region as were born	1,600,529	81.0%
% Father residing in same province as were born	1,600,529	73.5%
% Father born abroad	1,600,529	2.0%
% Families top-earner = father	1,600,529	87.1%
% Families with two positive incomes	1,719,483	60.9%
Corr. 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
% Female	1,719,483	48.4%
% Residing in North	1,635,680	48.5%
% Residing in Center	1,635,680	21.5%
% Residing in South	1,635,680	30.0%
% 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
% 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%

Table 1: Descriptive statistics for the final sample. Income is nominal and expressed in Euros. See Table A1 in the Appendix for more detail on the income distribution and its components.

Children												
Net Wage	Mean	SD	p99	p95	p90	p75	p50	p25	p10	p5	p1	
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												
Net Wage	Mean	SD	p99	p95	p90	p75	p50	p25	p10	p5	p1	
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												
Net Wage	Mean	SD	p99	p95	p90	p75	p50	p25	p10	p5	p1	
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	

Table 2: 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.

	Core	Male	Female	Married Father	Div/Sep Father	2 Earners
Dependent Variable: Child Income Rank						
Parental income rank	0.220 (0.0007)	0.222 (0.0010)	0.227 (0.0010)	0.225 (0.0008)	0.177 (0.0040)	0.219 (0.0010)
Constant	0.390 (0.0004)	0.449 (0.0006)	0.323 (0.0006)	0.388 (0.0005)	0.394 (0.0024)	0.392 (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
N	1,719,483	887,401	832,082	1,464,143	58,048	1,184,767

Table 3: National indicators of intergenerational mobility for the core sample and various subgroups. Standard errors 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.

	Only Wage	> 2/3 Wage	> 2/3 Entr.	> 2/3 Prof.	> 2/3 Cap.
Dependent Variable: Child Income Rank					
Parental income rank	0.231 (0.0012)	0.231 (0.0011)	0.228 (0.0017)	0.216 (0.0038)	0.211 (0.0047)
Constant	0.378 (0.0007)	0.379 (0.0007)	0.386 (0.0008)	0.404 (0.0029)	0.396 (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
N	860,931	991,014	299,909	57,047	64,637

Table 4: National indicators of intergenerational mobility for the core sample and various sub-groups. Standard errors 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).

	Father Top	Mother Top	Mover Reg., Father	Mover Reg.	Father Born Abroad
Dependent Variable: Child Income Rank					
Parental Income Rank	0.226 (0.0008)	0.193 (0.0017)	0.183 (0.0020)	0.129 (0.0022)	0.195 (0.0056)
Constant	0.387 (0.0005)	0.400 (0.0009)	0.426 (0.0013)	0.541 (0.0013)	0.410 (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
N	1,383,653	335,830	273,142	178,696	31,335

Table 5: National indicators of intergenerational mobility for the core sample and various sub-groups. Standard errors 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.

Parental Rank	Median Child Rank		95% CI for the Difference
	Earlier Born	Later Born	
100	65.2	62.3	[1.7, 4.4]
80	56.6	54.8	[0.4, 2.8]
60	51.9	50.1	[0.6, 3.2]
40	44.4	43.6	[-0.1, 2.0]
20	36.1	35.8	[-0.9, 1.4]

Table 6: Median rank of first-born children vs. later-born children by parental income ventile.

Parental Quintile	Child 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

Table 7: National quintile transition matrix (%)

Parental Decile	Child 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

Table 8: National decile transition matrix (%)

	Real 1	Real 2	Real 6	Real 9	Real "Moretti"
Dependent Variable: Child Income Rank					
Parental Income Rank	0.211 (0.0008)	0.195 (0.0008)	0.190 (0.0008)	0.191 (0.0008)	0.199 (0.0008)
Constant	0.394 (0.0004)	0.402 (0.0004)	0.404 (0.0004)	0.404 (0.0004)	0.400 (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
N	1,634,970	1,634,970	1,634,970	1,634,970	1,634,970

Table 9: National indicators of intergenerational mobility for the core sample based on PPP-adjusted income. Standard errors 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 \(2009\)](#). Specifically, Index 1 is constructed under the assumption that food, apparel and furniture are the only consumption categories who 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% 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% 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 \(2013\)](#).

σ_{κ_0}	ρ	σ_η	σ_ϵ
0.588	0.983	0.107	0.128

Table 10: Parameter estimates of persistent-transitory process for log residual earnings.

	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

Table 11: Average mobility statistics using 3-year average income or lifetime income based on simulations of the persistent-transitory model.

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

Table 12: Parameter estimates of log-linear model for 3-year average and lifetime earnings estimated on INPS data.

	RRS	AUM	Q1Q5	IGE
3-year average	0.226	0.444	0.116	0.267
Lifetime earnings	0.269	0.433	0.101	0.281

Table 13: Average mobility statistics using 3-year average or lifetime earnings based on simulations of the log-linear model.

	Core	With reweighting for missing 770	With adjustment for tax evasion	Combined
				Dependent Variable: Child Income Rank
Parental income rank	0.220 (0.0007)	0.212 (0.0006)	0.260 (0.0007)	0.246 (0.0005)
Constant	0.390 (0.0004)	0.399 (0.0003)	0.370 (0.0004)	0.378 (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
N	1,719,483	3,191,802	1,719,483	3,191,802

Table 14: National indicators of intergenerational mobility. Correction for missing 770 forms and tax evasion. Standard errors in parentheses. See the main text for details.

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
Padova	934,216	0.530	0.168	0.137	0.142
Venezia	863,133	0.530	0.156	0.116	0.111
Modena	700,913	0.529	0.198	0.164	0.169

Table 15: Top 10 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
Catania	1,090,101	0.376	0.078	0.207	0.203
Siracusa	404,271	0.376	0.078	0.204	0.200
Palermo	1,249,577	0.373	0.076	0.186	0.205
Cosenza	734,656	0.363	0.077	0.206	0.176

Table 16: Bottom 10 provinces by Absolute Upward Mobility.

	AUM	AUM	AUM	AUM	AUM
Gini Coefficient	-0.416 (0.0875)				
Gini Bottom 99%		-0.613 (0.0760)			
Top 1% Income Share			0.552 (0.0802)		
90-10 Ratio				-0.650 (0.0731)	
SD log Income					-0.649 (0.0732)
<i>R</i> ²	0.173	0.376	0.305	0.423	0.421
N	110	110	110	110	110

Table 17: Absolute Upward Mobility and income inequality. Standard errors in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

	RRS	RRS	RRS	RRS	RRS
Gini Coefficient	0.509 (0.0829)				
Gini Bottom 99%		0.608 (0.0764)			
Top 1% Income Share			-0.245 (0.0933)		
90-10 Ratio				0.639 (0.0741)	
SD log Income					0.622 (0.0753)
<i>R</i> ²	0.259	0.370	0.060	0.408	0.387
N	110	110	110	110	110

Table 18: Rank-Rank Slope and income inequality. Standard errors in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

	IGE	IGE	IGE	IGE	IGE
Gini Coefficient	0.222 (0.0938)				
Gini Bottom 99%		0.282 (0.0923)			
Top 1% 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
N	110	110	110	110	110

Table 19: Inter-Generational Elasticity and income inequality. Standard errors in parentheses. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

	ICS-15	ICS-20	ICS-25	ICS-30
AUM	-0.764 (0.0000)	-0.752 (0.0000)	-0.744 (0.0000)	-0.740 (0.0000)
Q1Q5	-0.609 (0.0000)	-0.590 (0.0000)	-0.579 (0.0000)	-0.567 (0.0000)
RRS	0.705 (0.0000)	0.707 (0.0000)	0.695 (0.0000)	0.700 (0.0000)
IGE	0.419 (0.0000)	0.428 (0.0000)	0.416 (0.0000)	0.424 (0.0000)

Table 20: Cross-province correlation of ICS indices from [Güell et al. \(2018\)](#) with mobility measures estimated on our data.

	AUM Baseline	AUM Macroarea FE	Q1Q5 Baseline	Q1Q5 Macroarea FE
<i>Productivity</i>				
Value added per resident	0.666*** (0.0000)	0.142** (0.0114)	0.605*** (0.0000)	0.196** (0.0121)
Manufacturing share of value added	0.688*** (0.0000)	0.261*** (0.0000)	0.565*** (0.0000)	0.172** (0.0102)
Public works started	0.388*** (0.0000)	0.124*** (0.0061)	0.483*** (0.0000)	0.263*** (0.0000)
Public works completed	0.266*** (0.0056)	0.107** (0.0113)	0.411*** (0.0000)	0.274*** (0.0000)
<i>Labor market</i>				
Unemployment rate	-0.811*** (0.0000)	-0.310*** (0.0001)	-0.654*** (0.0000)	-0.208* (0.0609)
Youth unemployment rate (age 15-24)	-0.849*** (0.0000)	-0.406*** (0.0000)	-0.693*** (0.0000)	-0.303*** (0.0070)
Long-term unemployment rate	-0.772*** (0.0000)	-0.220*** (0.0030)	-0.625*** (0.0000)	-0.165 (0.1163)
Employment rate (college degree or higher)	0.753*** (0.0000)	0.210*** (0.0012)	0.704*** (0.0000)	0.271*** (0.0027)
Labor force participation	0.771*** (0.0000)	0.200*** (0.0047)	0.644*** (0.0000)	0.185* (0.0636)
Female labor force participation	0.747*** (0.0000)	0.168** (0.0157)	0.606*** (0.0000)	0.124 (0.2058)
<i>Educational attainment</i>				
Share of illiterates	-0.436*** (0.0000)	0.020 (0.6842)	-0.354*** (0.0001)	-0.009 (0.8822)
Education level achieved	0.526*** (0.0000)	0.091* (0.0646)	0.497*** (0.0000)	0.177*** (0.0038)
School dropouts	-0.295*** (0.0020)	-0.032 (0.4908)	-0.159 (0.1004)	0.033 (0.6045)
<i>Social capital</i>				
Blood bags collected per resident (GSZ 2004)	0.481*** (0.0000)	0.068 (0.1554)	0.451*** (0.0000)	0.111* (0.0952)
Self-efficacy measure (GSZ 2016)	0.607*** (0.0000)	0.113** (0.0353)	0.617*** (0.0000)	0.265*** (0.0003)
Voter turnout, House of Representatives	0.722*** (0.0000)	0.205*** (0.0018)	0.545*** (0.0000)	0.100 (0.2869)
Voter turnout, European Parliament	0.485*** (0.0000)	0.108** (0.0357)	0.320*** (0.0007)	0.018 (0.7999)
Recycling to total waste ratio	0.798*** (0.0000)	0.304*** (0.0000)	0.706*** (0.0000)	0.306*** (0.0013)

Table 21: Unconditional correlation of AUM and Q1Q5 with various indicators (part 1/3). With and without macro-area fixed effects. p -values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	AUM Baseline	AUM Macroarea FE	Q1Q5 Baseline	Q1Q5 Macroarea FE
<i>Crime</i>				
Thefts	0.276*** (0.0040)	-0.103** (0.0248)	0.272*** (0.0044)	-0.052 (0.4226)
Violent crimes	-0.106 (0.2789)	-0.034 (0.4247)	-0.065 (0.5051)	-0.021 (0.7178)
Distraints	-0.237** (0.0138)	-0.049 (0.2653)	-0.196** (0.0417)	-0.007 (0.9116)
Scam offenses	0.520*** (0.0000)	0.158*** (0.0006)	0.508*** (0.0000)	0.189*** (0.0036)
<i>Life expectancy</i>				
Female life expectancy at birth	0.473*** (0.0000)	0.146*** (0.0031)	0.347*** (0.0002)	0.115 (0.1003)
Male life expectancy at birth	-0.102 (0.2952)	0.089* (0.0609)	-0.171* (0.0770)	0.056 (0.4001)
Old-age index (residents above age 65 / below age 15)	0.448*** (0.0000)	-0.174*** (0.0020)	0.274*** (0.0042)	-0.289*** (0.0002)
Number of suicides	0.409*** (0.0000)	-0.048 (0.3311)	0.354*** (0.0002)	-0.045 (0.5121)
<i>Family instability</i>				
Divorce rate	0.577*** (0.0000)	-0.216*** (0.0009)	0.488*** (0.0000)	-0.222** (0.0159)
Children in custody due to divorce	0.520*** (0.0000)	-0.206*** (0.0006)	0.420*** (0.0000)	-0.234*** (0.0056)
Separation rate	0.564*** (0.0000)	-0.115* (0.0741)	0.466*** (0.0000)	-0.112 (0.2136)
Children in custody due to separation	0.500*** (0.0000)	-0.137** (0.0201)	0.383*** (0.0000)	-0.185** (0.0240)
<i>Openness</i>				
Trade (exports+imports)	0.347*** (0.0002)	0.114*** (0.0070)	0.412*** (0.0000)	0.206*** (0.0004)
Net inter-province migration	0.671*** (0.0000)	0.084 (0.2249)	0.493*** (0.0000)	-0.019 (0.8410)
Foreign-born residents	0.786*** (0.0000)	0.319*** (0.0000)	0.647*** (0.0000)	0.305*** (0.0022)
Inflow of graduates from other province or abroad	0.591*** (0.0000)	0.179*** (0.0003)	0.553*** (0.0000)	0.268*** (0.0001)

Table 22: Unconditional correlation of AUM and Q1Q5 with various indicators (part 2/3). With and without macro-area fixed effects. p -values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	AUM Baseline	AUM Macroarea FE	Q1Q5 Baseline	Q1Q5 Macroarea FE
<i>Quality of schools</i>				
Local government spending in education	0.096 (0.3202)	0.006 (0.8965)	0.123 (0.1778)	0.035 (0.5144)
School assets	0.266*** (0.0052)	-0.014 (0.7588)	0.212** (0.0191)	-0.027 (0.6337)
Availability of teaching materials and technologies	0.280*** (0.0033)	0.064 (0.1420)	0.295*** (0.0009)	0.107** (0.0485)
Quality of school buildings	0.393*** (0.0000)	0.029 (0.5262)	0.303*** (0.0007)	-0.005 (0.9352)
Students using canteen, school bus and other services	0.689*** (0.0000)	0.189*** (0.0016)	0.565*** (0.0000)	0.183** (0.0158)
Students per class	-0.707*** (0.0000)	-0.208*** (0.0005)	-0.576*** (0.0000)	-0.205*** (0.0072)
Extended teaching-time availability	-0.053 (0.5871)	0.005 (0.9105)	0.067 (0.4664)	0.093* (0.0783)
Processing time of teachers' rankings	0.137 (0.1572)	-0.104** (0.0233)	0.105 (0.2484)	-0.106* (0.0654)
Teaching hours per class	-0.238** (0.0130)	-0.096** (0.0278)	-0.125 (0.1710)	-0.049 (0.3796)
Teachers below 40 years old: pre-school	0.641*** (0.0000)	0.123** (0.0301)	0.605*** (0.0000)	0.189*** (0.0080)
Teachers below 40 years old: primary school	0.474*** (0.0000)	0.086* (0.0711)	0.449*** (0.0000)	0.145** (0.0157)
Teachers below 40 years old: middle school	0.319*** (0.0007)	0.064 (0.1675)	0.345*** (0.0001)	0.086 (0.1406)
Teachers below 40 years old: high school	0.011 (0.9092)	-0.006 (0.8981)	0.105 (0.2540)	0.021 (0.7156)
Teachers under temporary contracts: pre-school	0.663*** (0.0000)	0.197*** (0.0002)	0.557*** (0.0000)	0.212*** (0.0017)
Teachers under temporary contracts: primary school	0.740*** (0.0000)	0.169** (0.0102)	0.651*** (0.0000)	0.252*** (0.0021)
Teachers under temporary contracts: middle school	0.439*** (0.0000)	0.036 (0.4634)	0.357*** (0.0000)	0.053 (0.3870)
Teachers under temporary contracts: high school	0.392*** (0.0000)	-0.031 (0.5167)	0.308*** (0.0005)	-0.044 (0.4607)
Tenure and stability of teachers' position	0.253*** (0.0080)	0.112** (0.0122)	0.267*** (0.0029)	0.092 (0.1021)
Students repeating school year	0.177* (0.0660)	0.086** (0.0403)	0.079 (0.3849)	0.007 (0.8913)
INVALSI test scores for primary and middle school	0.510*** (0.0000)	0.074 (0.1311)	0.405*** (0.0000)	0.046 (0.4549)
INVALSI test scores for high school	0.559*** (0.0000)	0.109** (0.0307)	0.437*** (0.0000)	0.074 (0.2477)

Table 23: Unconditional correlation of AUM and Q1Q5 with various indicators (part 3/3). With and without macro-area fixed effects. p -values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	AUM	AUM	AUM	AUM	AUM	AUM
Value added per resident	0.062 (0.273)	0.097* (0.087)	0.063 (0.275)	0.080 (0.103)	0.089* (0.054)	0.074 (0.137)
Crime	0.061 (0.141)	0.010 (0.851)	0.084* (0.054)	0.050 (0.167)	-0.050 (0.252)	0.069* (0.067)
Educational attainment	0.023 (0.631)	0.162*** (0.009)	0.013 (0.809)	0.004 (0.917)	0.106** (0.044)	-0.016 (0.744)
Family instability	-0.122** (0.046)	-0.108* (0.068)	-0.141** (0.031)	-0.237*** (0.000)	-0.283*** (0.000)	-0.259*** (0.000)
Strong labor market	0.668*** (0.000)	0.545*** (0.000)	0.599*** (0.000)	0.400*** (0.000)	0.251** (0.019)	0.380*** (0.001)
Life expectancy	-0.167*** (0.000)	-0.144*** (0.004)	-0.131*** (0.008)	-0.040 (0.396)	-0.014 (0.746)	-0.032 (0.487)
Economic openness	-0.042 (0.493)	-0.067 (0.276)	-0.033 (0.605)	-0.007 (0.892)	-0.084 (0.103)	-0.018 (0.743)
Social capital	0.081 (0.295)	0.086 (0.280)	0.115 (0.150)	0.031 (0.646)	-0.015 (0.825)	0.070 (0.323)
School quality		0.309*** (0.000)			0.145* (0.052)	
Pre-school quality			0.336*** (0.001)			0.250*** (0.003)
Primary school quality			0.124* (0.055)			0.146*** (0.008)
Middle school quality			0.125** (0.015)			-0.015 (0.763)
High school quality			-0.046 (0.448)			-0.064 (0.200)
School structures and resources				-0.008 (0.867)		-0.023 (0.589)
School organization and services				0.249*** (0.001)		0.202*** (0.004)
Teachers' composition				0.086 (0.112)		0.020 (0.681)
Students' grades and test scores				0.158*** (0.004)		0.084* (0.083)
Macroarea fixed effects					Yes	Yes
Observations	100	99	100	100	99	100
Adjusted R^2	0.84	0.85	0.85	0.88	0.90	0.89

Table 24: Conditional correlates of Absolute Upward Mobility. p -values in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are normalized to have mean 0 and standard deviation of 1, so coefficients can be interpreted as correlations.

APPENDIX FOR ONLINE PUBLICATION

originale
UNICO
99
 Persone fisiche

MINISTERO
 DELLE FINANZE
 Modello Unificato Compen
 periodo d'imposta 1998

019

CODICE FISCALE

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

REDDITI

FAMILIARI A CARICO

QUADRO RP

Oneri e spese

Mod. N.

--	--

FAMILIARI A CARICO

Relazione di parentela Codice fiscale
 (Indicare il codice fiscale del coniuge anche se non fiscalmente a carico) N. mesi a carico Percentuale di detrazione spettante

1	C																		
2	F	A																	
3	F	A																	
4	F	A																	
5	F	A																	
6	F	A																	
7	F	A																	
8	F	A																	

Figure A1: Form Modello Unico for 1999. Section on deduction for dependent relatives (familiari a carico). The top line indicates the SSN (codice fiscale) of the taxpayer. In this section, a taxpayer must indicate the name, relationship (relazione di parentela), SSN (codice fiscale), number of months (N. mesi a carico) and percentage of deduction claimed (percentuale di detrazione spettante) of the relative for whom a dependency deduction is claimed. The name of the spouse (Coniuge, C) must be indicated in all cases. The taxpayer must indicate whether the dependent is a child (Figlio, F) or other relative (Altro, A).

Variable	N	Mean	SD	p99.9	p99	p95	p90	p75	p50	p25	p10	p5	p1
Parental income 1998	1,719,483	29,957	35,926	354,495	140,329	75,336	55,533	36,825	23,173	14,309	7,028	3,703	197
Top-earner's income 1998	1,719,483	23,740	30,804	305,540	114,317	59,290	42,026	26,433	18,952	12,924	6,349	3,397	224
Father's income 1998													
Total	1,600,529	23,381	31,227	307,194	114,858	59,544	42,142	26,347	18,628	12,306	5,418	1,970	0
Wage	1,600,529	14,017	17,509	140,380	72,062	42,177	31,270	21,866	13,214	0	0	0	0
Entrepreneurship	1,600,529	4,024	14,515	151,783	52,847	21,615	13,633	0	0	0	0	0	0
Other self-employment	1,600,529	2,690	17,673	203,211	64,557	13,009	0	0	0	0	0	0	0
Capital	1,600,529	1,134	28,245	74,692	15,510	3,871	1,824	540	137	0	0	0	0
Mother's income 1998													
Total	1,196,230	11,777	14,577	138,382	56,956	29,475	22,829	17,669	10,188	1,696	208	81	0
Wage	1,196,230	7,681	10,237	62,159	36,937	22,560	19,835	15,452	1,907	0	0	0	0
Entrepreneurship	1,196,230	1,923	8,497	97,746	33,231	12,561	5,974	0	0	0	0	0	0
Other self-employment	1,196,230	558	5,709	68,689	18,035	0	0	0	0	0	0	0	0
Capital	1,196,230	929	6,459	51,023	12,523	3,581	1,769	558	167	0	0	0	0
Child's income 2017													
Total	1,635,680	21,034	23,948	204,561	84,078	48,944	37,800	27,427	19,641	10,257	3,050	638	-2,503
Wage	1,635,680	17,124	19,318	138,741	67,245	41,877	33,520	25,037	16,848	2,654	0	0	0
Entrepreneurship	1,635,680	1,335	11,897	111,788	38,550	11,140	0	0	0	0	0	0	-4,405
Other self-employment	1,635,680	1,731	10,329	110,386	38,083	10,680	525	0	0	0	0	0	0
Capital	1,635,680	1,206	9,375	64,527	18,667	5,766	1,819	531	99	0	0	0	0

Table A1: Summary statistics on the income distribution and its components for parents and children. Income is nominal and expressed in Euros. "Top-earner" is the parent - father or mother - who earns more income.

	Index 1	Index 2	Index 6	Index 9	Index "Moretti"
Index 1	1.000				
Index 2	0.681	1.000			
Index 6	0.823	0.899	1.000		
Index 9	0.809	0.905	0.999	1.000	
Index "Moretti"	0.538	0.837	0.692	0.695	1.000

Table A2: Cross-region correlation of cost of living indexes from [Cannari and Iuzzolino \(2009\)](#) and [Boeri et al. \(2018\)](#).

Child's Outcome	Parent's Income Definition	Core	Core	Core	Male	Male	Male
Income rank	Parental income rank	0.220 (0.0007)			0.222 (0.0010)		
Income rank	Top-earner income rank		0.211 (0.0007)			0.220 (0.0010)	
Income rank	Father income rank			0.202 (0.0008)			0.218 (0.0011)
Constant		0.390 (0.0004)	0.394 (0.0004)	0.400 (0.0004)	0.449 (0.0006)	0.450 (0.0006)	0.452 (0.0006)
AUM		0.445	0.447	0.451	0.504	0.505	0.507
Q1Q5		0.112	0.113	0.122	0.156	0.158	0.167
N		1,719,483	1,719,483	1,644,653	887,401	887,401	849,921
Child's Outcome	Parent's Income Definition	Female	Female	Female	Married Father	Married Father	Married Father
Income rank	Parental income rank	0.227 (0.0010)			0.225 (0.0008)		
Income rank	Top-earner income rank		0.209 (0.0010)			0.215 (0.0008)	
Income rank	Father income rank			0.192 (0.0010)			0.206 (0.0008)
Constant		0.323 (0.0006)	0.332 (0.0006)	0.342 (0.0006)	0.388 (0.0005)	0.394 (0.0005)	0.399 (0.0005)
AUM		0.379	0.384	0.390	0.445	0.448	0.451
Q1Q5		0.061	0.063	0.072	0.110	0.113	0.123
N		832,082	832,082	794,732	1,464,143	1,464,143	1,464,143
Child's Outcome	Parent's Income Definition	Div/Sep Father	Div/Sep Father	Div/Sep Father	2 Earners	2 Earners	2 Earners
Income rank	Parental income rank	0.177 (0.0040)			0.219 (0.0010)		
Income rank	Top-earner income rank		0.168 (0.0040)			0.209 (0.0009)	
Income rank	Father income rank			0.153 (0.0040)			0.195 (0.0009)
Constant		0.394 (0.0024)	0.395 (0.0025)	0.404 (0.0025)	0.392 (0.0006)	0.405 (0.0006)	0.415 (0.0006)
AUM		0.438	0.437	0.443	0.446	0.457	0.464
Q1Q5		0.118	0.121	0.129	0.117	0.122	0.135
N		58,048	58,048	58,048	1,184,767	1,184,767	1,184,767

Table A3: Additional national indicators of intergenerational mobility for the core sample and various subgroups (part 1/3). Standard errors in parentheses. Results reported for three definition of parental income: family income, top-earner income, and father income. See Table 3 for details on the subgroups.

Child's Outcome	Parent's Income Definition	Father Top	Father Top	Father Top	Mother Top	Mother Top	Mother Top
Income rank	Parental income rank	0.226 (0.0008)			0.193 (0.0017)		
Income rank	Top-earner income rank		0.215 (0.0008)			0.202 (0.0018)	
Income rank	Father income rank			0.224 (0.0009)			0.229 (0.0023)
Constant		0.387 (0.0005)	0.391 (0.0005)	0.381 (0.0005)	0.400 (0.0009)	0.405 (0.0009)	0.431 (0.0008)
AUM		0.444	0.444	0.437	0.448	0.455	0.489
Q1Q5		0.108	0.110	0.106	0.122	0.124	0.147
N		1,383,653	1,383,653	1,383,653	335,830	335,830	261,000
Child's Outcome	Parent's Income Definition	> 2/3 Wage	> 2/3 Wage	> 2/3 Wage	> 2/3 Entr.	> 2/3 Entr.	> 2/3 Entr.
Income rank	Parental income rank	0.231 (0.0011)			0.228 (0.0017)		
Income rank	Top-earner income rank		0.225 (0.0011)			0.217 (0.0017)	
Income rank	Father income rank			0.207 (0.0011)			0.217 (0.0018)
Constant		0.379 (0.0007)	0.380 (0.0007)	0.392 (0.0007)	0.386 (0.0008)	0.390 (0.0008)	0.389 (0.0008)
AUM		0.437	0.437	0.443	0.443	0.444	0.444
Q1Q5		0.093	0.092	0.114	0.111	0.114	0.115
N		991,014	991,014	950,815	299,909	299,909	291,230
Child's Outcome	Parent's Income Definition	> 2/3 Prof.	> 2/3 Prof.	> 2/3 Prof.	> 2/3 Cap.	> 2/3 Cap.	> 2/3 Cap.
Income rank	Parental income rank	0.216 (0.0038)			0.211 (0.0047)		
Income rank	Top-earner income rank		0.209 (0.0038)			0.203 (0.0046)	
Income rank	Father income rank			0.201 (0.0040)			0.207 (0.0050)
Constant		0.404 (0.0029)	0.403 (0.0030)	0.410 (0.0031)	0.396 (0.0013)	0.398 (0.0013)	0.395 (0.0014)
AUM		0.458	0.455	0.461	0.449	0.448	0.446
Q1Q5		0.141	0.140	0.152	0.127	0.127	0.127
N		57,047	57,047	55,019	64,637	64,637	59,549

Table A4: Additional national indicators of intergenerational mobility for the core sample and various subgroups (part 2/3). Standard errors in parentheses. Results reported for three definition of parental income: family income, top-earner income, and father income. See Tables 4 and 5 for details on the subgroups.

Child's Outcome	Parent's Income Definition	Lawyers & co.	Lawyers & co.	Lawyers & co.	Mover	Mover	Mover
Income rank	Parental income rank	0.216 (0.0048)			0.155 (0.0018)		
Income rank	Top-earner income rank		0.206 (0.0048)			0.150 (0.0018)	
Income rank	Father income rank			0.180 (0.0045)			0.141 (0.0018)
Constant		0.379 (0.0037)	0.386 (0.0036)	0.411 (0.0033)	0.507 (0.0011)	0.510 (0.0011)	0.516 (0.0011)
AUM		0.433	0.437	0.456	0.546	0.547	0.551
Q1Q5		0.136	0.136	0.157	0.205	0.208	0.219
N		49,366	49,366	48,438	277,462	277,462	277,462

Table A5: Additional national indicators of intergenerational mobility for the core sample and various subgroups (part 3/3). Standard errors in parentheses. Results reported for three definition of parental income: family income, top-earner income, and father income. "Lawyers & co." restricts to children for which the occupation of at least one parent in 1998 involved professional services (ATECO code 74 - including lawyers, accountants, consultants, architects, ...). "Mover" refers to the restriction to children whose province of residence in 2017 is different from the province of residence of their fathers in 1998.

	Sons	Married	Div/Sep	Mover Reg., Father	Mover Reg.	Father Born Abroad
Dependent Variable: Son's Income Rank						
Father's income rank	0.226 (0.0011)	0.230 (0.0011)	0.165 (0.0056)	0.182 (0.0028)	0.125 (0.0030)	0.214 (0.0078)
Constant	0.389 (0.0006)	0.388 (0.0007)	0.394 (0.0036)	0.430 (0.0017)	0.565 (0.0017)	0.403 (0.0046)
AUM	0.445	0.445	0.435	0.476	0.596	0.456
Q1Q5	0.112	0.112	0.130	0.149	0.217	0.144
TMR	4.835	4.818	7.973	4.371	3.827	3.382
N	849,921	756,878	29,449	138,157	92,340	16,049

Table A6: National indicators of intergenerational mobility for fathers and sons (part 1/2). Standard errors in parentheses. See Tables 3 and 5 for details on the subgroups.

	Only Wage	> 2/3 Wage	> 2/3 Entr.	> 2/3 Prof.	> 2/3 Cap.
Dependent Variable: Son's Income Rank					
Father's income rank	0.234 (0.0020)	0.231 (0.0015)	0.227 (0.0022)	0.216 (0.0031)	0.197 (0.0046)
Constant	0.366 (0.0012)	0.382 (0.0009)	0.387 (0.0010)	0.401 (0.0019)	0.395 (0.0014)
AUM	0.425	0.439	0.444	0.455	0.444
Q1Q5	0.084	0.092	0.114	0.121	0.123
TMR	3.501	3.523	5.559	7.093	5.672
N	262,017	505,498	194,191	62,121	65,661

Table A7: National indicators of intergenerational mobility for fathers and sons (part 2/2). Standard errors in parentheses. See Table 4 for details on the subgroups.

Father's Quintile	Son's Quintile				
	1st	2nd	3rd	4th	5th
1st	28.25	24.08	19.27	16.20	12.21
2nd	21.30	21.97	21.98	19.89	14.86
3rd	17.75	19.92	21.89	22.48	17.96
4th	16.84	18.30	20.05	22.46	22.36
5th	15.21	15.47	16.84	19.27	33.21

Table A8: National quintile transition matrix - fathers and sons (%)

Father's Decile	Son's Decile									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
1st	15.84	13.94	12.92	11.26	9.75	8.81	7.98	7.61	6.74	5.16
2nd	13.65	13.07	12.50	11.49	10.37	9.61	8.57	8.25	7.26	5.25
3rd	10.97	11.38	11.37	11.16	11.10	10.52	9.93	9.16	8.21	6.21
4th	9.92	10.32	10.65	10.76	11.22	11.13	10.73	9.95	8.86	6.45
5th	8.97	9.19	10.01	10.35	10.95	11.35	11.26	10.89	9.71	7.32
6th	8.43	8.91	9.42	10.05	10.44	11.05	11.53	11.28	10.51	8.38
7th	8.42	8.70	9.11	9.55	9.90	10.61	11.13	11.39	11.59	9.59
8th	8.10	8.45	8.69	9.24	9.53	10.05	10.81	11.58	12.29	11.24
9th	7.56	7.97	7.92	8.48	9.08	9.49	10.12	10.98	13.04	15.37
10th	7.40	7.49	7.11	7.44	7.58	7.52	8.19	9.24	12.31	25.71

Table A9: National decile transition matrix - fathers and sons (%)

Identification and Estimation of the Income Process

Recall the income process specified in Section 5.1. For each individual in the sample, we have three consecutive observations at dates $t, t - 1, t - 2$. The cross-sectional moments we use for estimation are:

$$\text{var}(y_t) = \text{var}(\kappa_t) + \sigma_\varepsilon$$

and

$$\begin{aligned}\text{cov}(y_t, y_{t-1}) &= \rho \text{var}(\kappa_{t-1}) \\ \text{cov}(y_{t-1}, y_{t-2}) &= \rho \text{var}(\kappa_{t-2}) \\ \text{cov}(y_t, y_{t-2}) &= \rho^2 \text{var}(\kappa_{t-2}),\end{aligned}$$

where

$$\text{var}(\kappa_t) = \rho^2 \text{var}(\kappa_{t-1}) + \sigma_n = \rho^{2t} \sigma_{\kappa_0} + \sigma_n \sum_{j=1}^t \rho^{2(j-1)},$$

Identification can be achieved in different ways. For example:

$$\begin{aligned}\rho &= \frac{\text{cov}(y_t, y_{t-2})}{\text{cov}(y_{t-1}, y_{t-2})} \\ \sigma_\varepsilon &= \text{var}(y_{t-1}) - \rho^{-1} \text{cov}(y_t, y_{t-1}) \\ \sigma_\eta &= \text{var}(y_{t-1}) - \text{cov}(y_t, y_{t-2}) - \sigma_\varepsilon \\ \sigma_{\kappa_0} &= \left[\text{var}(y_t) - \sigma_\varepsilon - \sigma_n \sum_{j=1}^t \rho^{2(j-1)} \right] \cdot \rho^{-2t}.\end{aligned}$$

To avoid contamination by outliers, we trim the top and bottom 5% of the distribution of log-changes. The estimation methodology follows a standard minimum distance approach (Chamberlain, 1984).

References

CHAMBERLAIN, G. (1984). "Panel Data." *Handbook of Econometrics* 2, 1247-1318.

Correction for the omission of poor children

We discussed in Section 3 how individuals with very low or no income are not required to file. Our dataset contains records of parents with missing information on children. Under the assumption that these children did not file taxes in because their income fell below this threshold level, we can gauge the effect of omitting poor children from our sample.

We have 876,055 parents with missing information on children in 2017, except for their year of birth. Applying the same age selection criteria for parents and children as in the core sample, we are left with 450,740 records. For each of these records, we set the children's income in 2017 from missing to zero, add them to the core sample and re-estimate the national mobility measures, computing average incomes three years of data as we do in the core analysis. The impact of this correction is small: the RRS decreases from 0.22 to 0.215, AUM increases from 0.445 to 0.446 and the Q1Q5 decreases from 0.112 to 0.106.

Correction for missing capital income

We have also implemented a correction for some forms of capital income that are missing from tax returns. In order to impute missing capital income from tax returns we proceed as follows. We start with Bank of Italy's SHIW data and define two categories of capital income which are missing in our data: interest income earned on government bonds, and other capital income (from deposits, stocks, bonds, and mutual funds). This distinction is necessary because SHIW data are reported after-tax, and the tax withheld on government bonds is lower. We convert data into before-tax income by using a tax rate of 12.5% for government bonds, and 26% for other capital income. For each of these two categories, we regress capital income on dummy variables that capture i) quantile in the distribution of total income, ii) age above/below 45 years old, iii) gender, iv) year. Using the estimated coefficients, we impute capital income for each child, parent and year in our dataset, relying on the assumption that the tax system being monotonic implies that the individual rank in the before-tax and after-tax income distribution is the same. We take into account that dividends are subject to different tax rules if holdings represent a substantial share of the stock of a firm (qualified holdings) or not. We estimate that in 2015 (2000) approximately 36% (33%) of all dividends were earned on non-qualified holdings of stocks, while the remaining 64% (67%) corresponded to qualified holdings. Dividends from qualified holdings are included in tax returns while dividends from non-qualified holdings are not, thus this is the portion of capital income we are missing. We also account for the fact that starting 2004 only approximately 50% of dividends from qualified holdings have to be reported on tax returns. This correction leaves our estimates of relative and absolute mobility virtually identical.



Figure A2: Map of the Italian provinces

Province	Population in 2010	Region	Macro area
Genova	882,718	Liguria	
Imperia	222,648	Liguria	
La Spezia	223,516	Liguria	
Savona	287,906	Liguria	
Bergamo	1,100,000	Lombardia	
Brescia	1,300,000	Lombardia	
Como	594,988	Lombardia	
Cremona	363,606	Lombardia	
Lecco	340,167	Lombardia	
Lodi	227,655	Lombardia	
Mantova	415,442	Lombardia	
Milano	3,200,000	Lombardia	
Monza-Brianza	849,636	Lombardia	North-West
Pavia	548,307	Lombardia	
Sondrio	183,169	Lombardia	
Varese	883,285	Lombardia	
Alessandria	440,613	Piemonte	
Asti	221,687	Piemonte	
Biella	185,768	Piemonte	
Cuneo	592,303	Piemonte	
Novara	371,802	Piemonte	
Torino	2,300,000	Piemonte	
Verbania	163,247	Piemonte	
Vercelli	179,562	Piemonte	
Aosta	128,230	Valle d'Aosta	

Table A10: Administrative divisions of Italy (part 1/3)

Province	Population in 2010	Region	Macro area
Bologna	991,924	Emilia	North-East
Ferrara	359,994	Emilia	
Forli-Cesena	395,489	Emilia	
Modena	700,913	Emilia	
Parma	442,120	Emilia	
Piacenza	289,875	Emilia	
Ravenna	392,458	Emilia	
Reggio-Emilia	530,343	Emilia	
Rimini	329,302	Emilia	
Gorizia	142,407	Friuli Venezia Giulia	
Pordenone	315,323	Friuli Venezia Giulia	
Trieste	236,556	Friuli Venezia Giulia	
Udine	541,522	Friuli Venezia Giulia	
Bolzano	507,657	Trentino	
Trento	529,457	Trentino	
Belluno	213,474	Veneto	
Padova	934,216	Veneto	
Rovigo	247,884	Veneto	
Treviso	888,249	Veneto	
Venezia	863,133	Veneto	
Verona	920,158	Veneto	
Vicenza	870,740	Veneto	
Frosinone	498,167	Lazio	Center
Latina	555,692	Lazio	
Rieti	160,467	Lazio	
Roma	4,200,000	Lazio	
Viterbo	320,294	Lazio	
Ancona	481,028	Marche	
Ascoli Piceno	214,068	Marche	
Fermo	177,914	Marche	
Macerata	325,362	Marche	
Pesaro-Urbino	366,963	Marche	
Arezzo	349,651	Toscana	
Firenze	998,098	Toscana	
Grosseto	228,157	Toscana	
Livorno	342,955	Toscana	
Lucca	393,795	Toscana	
Massa-Carrara	203,901	Toscana	
Pisa	417,782	Toscana	
Pistoia	293,061	Toscana	
Prato	249,775	Toscana	
Siena	272,638	Toscana	
Perugia	671,821	Umbria	
Terni	234,665	Umbria	

Table A11: Administrative divisions of Italy (part 2/3)

Province	Population in 2010	Region	Macro area
Chieti	397,123	Abruzzo	
L'Aquila	309,820	Abruzzo	
Pescara	323,184	Abruzzo	
Teramo	312,239	Abruzzo	
Matera	203,726	Basilicata	
Potenza	383,791	Basilicata	
Catanzaro	368,597	Calabria	
Cosenza	734,656	Calabria	
Crotone	174,605	Calabria	
Reggio-Calabria	566,977	Calabria	
Vibo Valentia	166,560	Calabria	
Avellino	439,137	Campania	
Benevento	287,874	Campania	
Caserta	916,467	Campania	
Napoli	3,100,000	Campania	
Salerno	1,100,000	Campania	
Campobasso	231,086	Molise	
Isernia	88,694	Molise	
Bari	1,300,000	Puglia	
Barletta-Trani-Andria	392,863	Puglia	
Brindisi	403,229	Puglia	South and Islands
Foggia	640,836	Puglia	
Lecce	815,597	Puglia	
Taranto	580,028	Puglia	
Cagliari	563,180	Sardegna	
Carbonia - Iglesias	129,840	Sardegna	
Medio Campidano	102,409	Sardegna	
Nuoro	160,677	Sardegna	
Ogliastra	57,965	Sardegna	
Olbia - Tempio	157,859	Sardegna	
Oristano	166,244	Sardegna	
Sassari	337,237	Sardegna	
Agrigento	454,002	Sicilia	
Caltanissetta	271,729	Sicilia	
Catania	1,100,000	Sicilia	
Enna	172,485	Sicilia	
Messina	653,737	Sicilia	
Palermo	1,200,000	Sicilia	
Ragusa	318,549	Sicilia	
Siracusa	404,271	Sicilia	
Trapani	436,624	Sicilia	

Table A12: Administrative divisions of Italy (part 3/3)

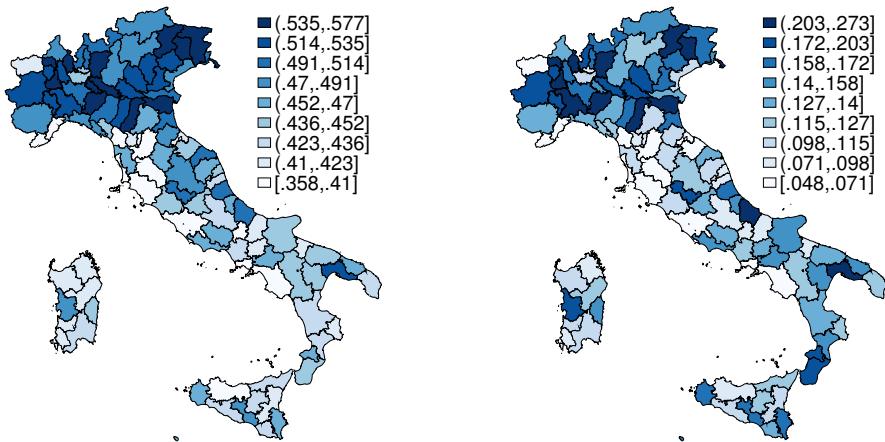


Figure A3: Heat map of Absolute Upward Mobility and Q1Q5 based on PPP-adjusted income (with province-level price indices). Dark areas are more mobile. Left-panel (a): AUM. Right-panel (b): Q1Q5

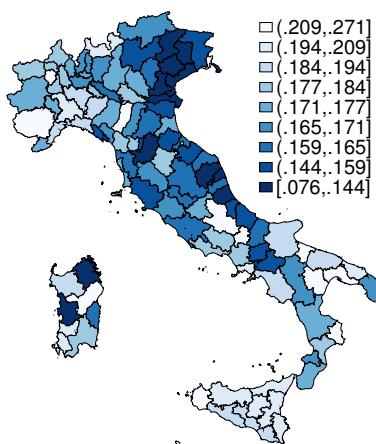


Figure A4: Heat map of the IGE based on nominal income. Dark areas are more mobile.

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	IGE	N
1	Bolzano	507,657	0.574	0.282	0.152	0.169	13,765
2	Lecco	340,167	0.554	0.213	0.141	0.166	10,294
3	Monza-Brianza	849,636	0.545	0.204	0.160	0.175	25,131
4	Bergamo	1,098,740	0.544	0.200	0.142	0.171	30,294
5	Belluno	213,474	0.541	0.168	0.101	0.076	6,395
6	Treviso	888,249	0.541	0.183	0.110	0.127	25,283
7	Milano	3,156,694	0.539	0.224	0.161	0.167	81,305
8	Cremona	363,606	0.536	0.183	0.141	0.161	9,877
9	Lodi	227,655	0.535	0.167	0.164	0.181	5,936
10	Trento	529,457	0.534	0.198	0.127	0.152	15,420
11	Vicenza	870,740	0.533	0.167	0.139	0.161	25,365
12	Padova	934,216	0.530	0.168	0.137	0.142	27,000
13	Venezia	863,133	0.530	0.156	0.116	0.111	23,354
14	Modena	700,913	0.529	0.198	0.164	0.169	17,407
15	Parma	442,120	0.528	0.184	0.153	0.172	10,422
16	Pordenone	315,323	0.526	0.141	0.125	0.144	8,789
17	Piacenza	289,875	0.525	0.213	0.168	0.187	7,031
18	Bologna	991,924	0.524	0.178	0.150	0.160	21,805
19	Brescia	1,256,025	0.519	0.166	0.152	0.174	31,963
20	Verona	920,158	0.519	0.164	0.152	0.176	26,389
21	Mantova	415,442	0.518	0.166	0.144	0.180	9,966
22	Reggio-Emilia	530,343	0.518	0.163	0.194	0.219	12,874
23	Udine	541,522	0.514	0.143	0.127	0.146	15,270
24	Varese	883,285	0.514	0.187	0.166	0.173	22,860
25	Novara	371,802	0.506	0.144	0.165	0.183	10,199
26	Cuneo	592,303	0.505	0.154	0.186	0.223	16,910
27	Ravenna	392,458	0.505	0.164	0.149	0.175	10,053
28	Aosta	128,230	0.503	0.159	0.166	0.182	3,484
29	Como	594,988	0.503	0.188	0.149	0.166	15,593
30	Pavia	548,307	0.503	0.168	0.171	0.190	12,320
31	Trieste	236,556	0.494	0.162	0.145	0.143	4,835
32	Asti	221,687	0.493	0.160	0.161	0.177	5,277
33	Forli-Cesena	395,489	0.493	0.111	0.153	0.148	11,593
34	Firenze	998,098	0.491	0.145	0.135	0.130	25,176
35	Genova	882,718	0.491	0.152	0.186	0.186	20,965
36	Alessandria	440,613	0.489	0.137	0.180	0.209	10,371
37	Rovigo	247,884	0.488	0.122	0.144	0.144	7,789
38	Biella	185,768	0.487	0.150	0.193	0.225	4,835
39	Vercelli	179,562	0.487	0.142	0.164	0.180	4,639
40	Prato	249,775	0.486	0.138	0.164	0.149	7,276

Table A13: Inter-generational mobility across provinces (part 1/3)

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	IGE	N
41	Torino	2,302,353	0.483	0.145	0.170	0.176	60,748
42	Ferrara	359,994	0.482	0.119	0.152	0.158	9,325
43	Sondrio	183,169	0.482	0.151	0.169	0.195	5,819
44	La Spezia	223,516	0.477	0.135	0.162	0.154	5,403
45	Arezzo	349,651	0.473	0.114	0.175	0.184	10,094
46	Siena	272,638	0.471	0.132	0.160	0.168	7,594
47	Verbania	163,247	0.470	0.134	0.192	0.181	4,117
48	Ancona	481,028	0.467	0.091	0.155	0.165	15,265
49	Pisa	417,782	0.466	0.107	0.155	0.160	11,761
50	Gorizia	142,407	0.464	0.081	0.176	0.237	3,211
51	Pesaro-Urbino	366,963	0.462	0.091	0.152	0.154	11,478
52	Rimini	329,302	0.459	0.109	0.157	0.165	9,423
53	Macerata	325,362	0.457	0.088	0.125	0.123	10,322
54	Savona	287,906	0.455	0.127	0.186	0.185	6,362
55	Pistoia	293,061	0.454	0.112	0.166	0.183	7,628
56	Perugia	671,821	0.452	0.083	0.142	0.161	21,445
57	Lucca	393,795	0.449	0.098	0.162	0.180	10,532
58	Fermo	177,914	0.446	0.093	0.139	0.142	5,635
59	Roma	4,194,068	0.443	0.122	0.177	0.164	106,656
60	Terni	234,665	0.443	0.101	0.147	0.161	7,305
61	Massa-Carrara	203,901	0.442	0.125	0.169	0.169	5,559
62	Livorno	342,955	0.439	0.092	0.164	0.175	9,371
63	Ascoli Piceno	214,068	0.437	0.095	0.149	0.161	7,456
64	Chieti	397,123	0.437	0.106	0.157	0.145	14,193
65	Grosseto	228,157	0.436	0.103	0.157	0.145	6,175
66	Rieti	160,467	0.434	0.093	0.180	0.176	5,096
67	Taranto	580,028	0.430	0.110	0.200	0.218	22,231
68	Frosinone	498,167	0.426	0.103	0.178	0.184	18,724
69	L'Aquila	309,820	0.424	0.096	0.187	0.217	11,452
70	Teramo	312,239	0.422	0.089	0.147	0.142	10,602
71	Imperia	222,648	0.420	0.085	0.197	0.177	4,642
72	Brindisi	403,229	0.419	0.097	0.237	0.205	14,405
73	Latina	555,692	0.417	0.100	0.198	0.184	17,842
74	Caserta	916,467	0.416	0.100	0.200	0.181	24,638
75	Viterbo	320,294	0.415	0.083	0.174	0.167	10,057
76	Potenza	383,791	0.413	0.092	0.181	0.170	14,398
77	Bari	1,258,706	0.412	0.089	0.207	0.194	41,057
78	Lecce	815,597	0.410	0.095	0.196	0.168	27,909
79	Matera	203,726	0.409	0.090	0.206	0.213	8,689
80	Napoli	3,080,873	0.405	0.102	0.221	0.195	80,126

Table A14: Inter-generational mobility across provinces (part 2/3)

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	IGE	N
81	Pescara	323,184	0.405	0.090	0.166	0.157	9,966
82	Avellino	439,137	0.401	0.089	0.193	0.159	13,994
83	Campobasso	231,086	0.401	0.076	0.164	0.164	8,422
84	Foggia	640,836	0.398	0.082	0.212	0.191	23,885
85	Benevento	287,874	0.396	0.100	0.186	0.151	9,099
86	Vibo Valentia	166,560	0.396	0.088	0.188	0.167	5,142
87	Trapani	436,624	0.393	0.085	0.190	0.226	14,269
88	Cagliari	563,180	0.392	0.082	0.160	0.179	16,490
89	Caltanissetta	271,729	0.391	0.082	0.208	0.187	9,860
90	Catanzaro	368,597	0.391	0.090	0.186	0.180	11,640
91	Salerno	1,109,705	0.390	0.086	0.210	0.194	33,592
92	Barletta-Trani-Andria	392,863	0.387	0.069	0.238	0.226	13,680
93	Olbia - Tempio	157,859	0.386	0.071	0.138	0.116	4,114
94	Oristano	166,244	0.385	0.070	0.143	0.135	5,844
95	Isernia	88,694	0.383	0.068	0.232	0.219	3,215
96	Enna	172,485	0.382	0.068	0.190	0.201	5,987
97	Crotone	174,605	0.382	0.073	0.208	0.219	4,923
98	Agrigento	454,002	0.381	0.071	0.206	0.191	13,877
99	Medio Campidano	102,409	0.380	0.065	0.152	0.206	3,784
100	Carbonia - Iglesias	129,840	0.379	0.076	0.167	0.201	4,454
101	Messina	653,737	0.379	0.080	0.212	0.207	20,322
102	Reggio-Calabria	566,977	0.377	0.086	0.219	0.176	15,336
103	Sassari	337,237	0.377	0.077	0.181	0.194	10,457
104	Catania	1,090,101	0.376	0.078	0.207	0.203	29,084
105	Siracusa	404,271	0.376	0.078	0.204	0.200	12,717
106	Palermo	1,249,577	0.373	0.076	0.186	0.205	33,949
107	Ogliastra	57,965	0.372	0.052	0.180	0.161	1,872
108	Nuoro	160,677	0.364	0.064	0.199	0.271	5,201
109	Cosenza	734,656	0.363	0.077	0.206	0.176	23,178
110	Ragusa	318,549	0.353	0.063	0.230	0.192	10,214

Table A15: Inter-generational mobility across provinces (part 3/3)

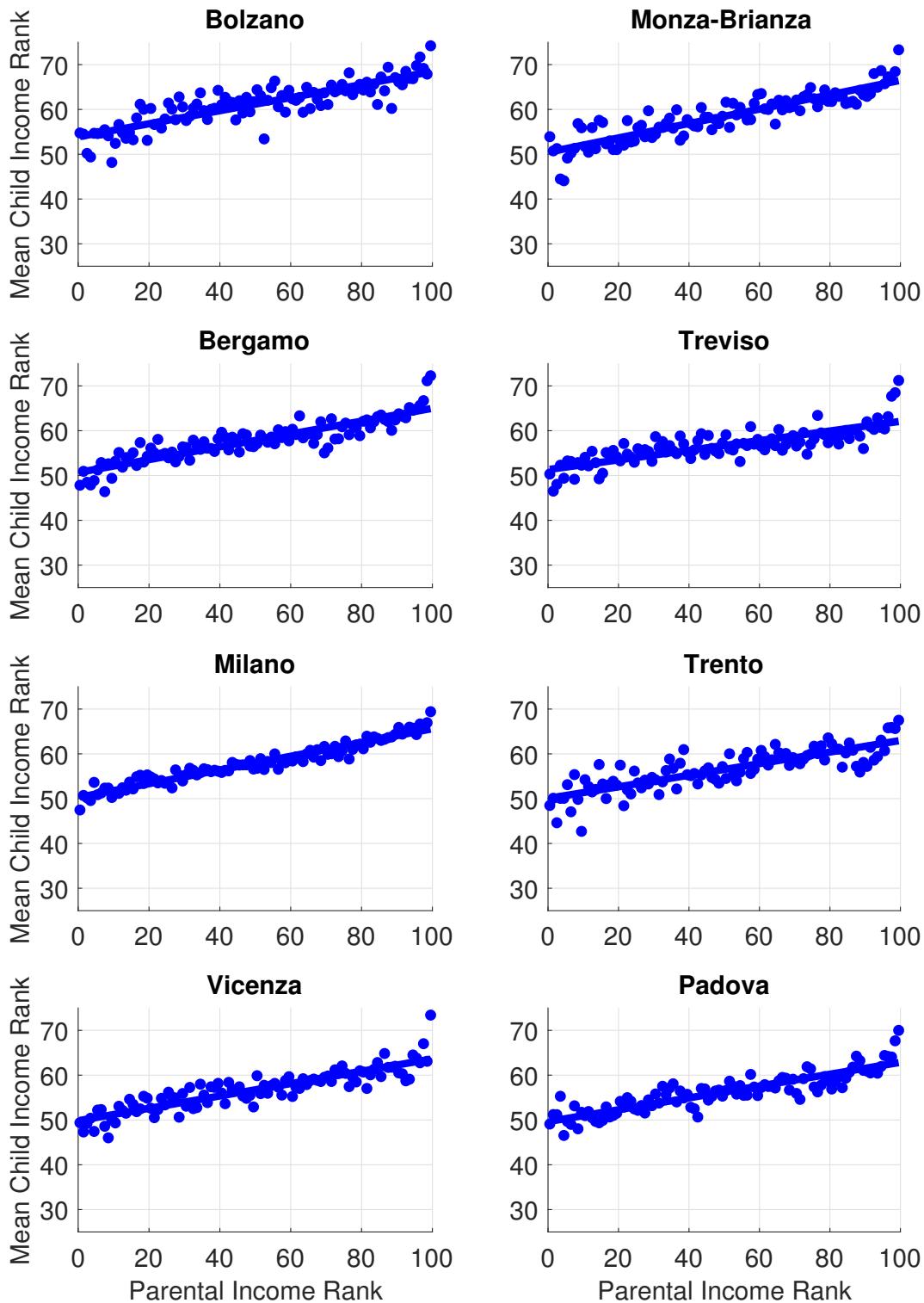


Figure A5: Rank-rank relationship for children who grew up in top provinces for AUM.

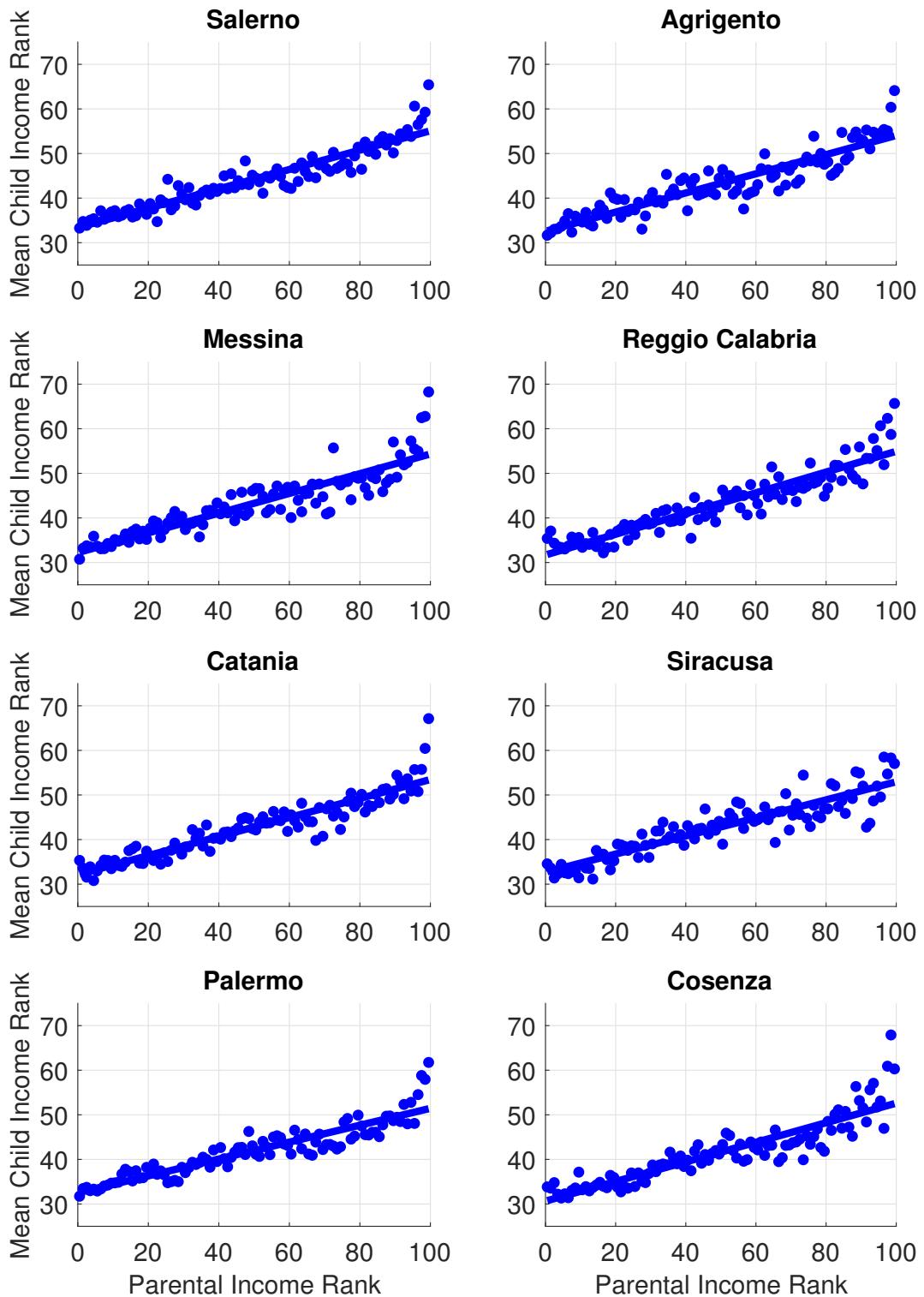


Figure A6: Rank-rank relationship for children who grew up in bottom provinces for AUM.

Variable	p90	p50	p10
<i>Productivity</i>			
Value added per resident (Euros)	22,342	17,121	10,676
Manufacturing share of value added	29%	18%	7.5%
Public works started (1,000 Euros)	267,223	67,500	29,925
Public works completed (1,000 Euros)	166,361	71,023	32,809
<i>Labor market</i>			
Unemployment rate	23%	8.2%	3.7%
Female unemployment rate	36%	12%	5.8%
Male unemployment rate	17%	5.4%	2.1%
Youth unemployment rate (age 15-24)	61%	25%	9.9%
Long-term unemployment rate (12 months or more)	9.2%	2.1%	0.96%
Female employment rate	39%	32%	17%
Male employment rate	63%	57%	48%
Youth employment rate (age 15-24)	43%	30%	12%
Employment rate (high school diploma)	82%	76%	60%
Employment rate (college degree or higher)	87%	81%	73%
Labor force participation	66%	61%	52%
Female labor force participation	57%	48%	32%
Male labor force participation	77%	73%	69%
Youth labor force participation (age 15-24)	49%	38%	28%
<i>Educational attainment</i>			
Share of illiterates	24%	20%	17%
School dropouts (per 100 residents)	32	22	14
<i>Social capital</i>			
Blood bags collected per resident (GSZ 2004)	0.055	0.018	0
Non-profit organizations per resident	7.3	4.5	2.7
Voter turnout, House of Representatives	89%	85%	76%
Voter turnout, European Parliament	81%	75%	63%
Recycling to total waste ratio	31%	14%	1%
<i>Crime</i>			
Burglaries (per 100,000 residents)	804	545	290
Smuggling offenses (per 100,000 residents)	65	1.6	0
GTAs (per 100,000 residents)	592	187	87
Petty thefts (per 100,000 residents)	463	93	17
Thefts with tear (per 100,000 residents)	80	21	4.5
Murders (per 100,000 residents)	2.4	0.92	0
Distraints (per 1,000 residents aged 18 or more)	12	6.1	3.4
Exploitation-of-prostitution offenses (per 100,000 residents)	9.7	2.9	0.44
Protests of checks (per 1,000 residents)	11	5.3	2.6
Drug dealing offenses (per 100,000 residents)	139	60	30
Scam offenses (per 100,000 residents)	138	66	33

Table A16: Statistics on province-level indicators used in the analyses of mobility and local characteristics (part 1/2)

Variable	p90	p50	p10
<i>Life expectancy</i>			
Female life expectancy at birth (years)	83	82	81
Male life expectancy at birth (years)	77	76	75
Female life expectancy at 65 (years)	21	20	19
Male life expectancy at 65 (years)	17	16	15
Old-age index (ratio of residents above age 65 to below age 15)	2.18	1.45	0.91
Number of suicides (per 100,000 residents)	10	6	1.5
<i>Family instability</i>			
Divorce rate	1	0.56	0.21
Divorces (per 100 marriages)	25	12	3.9
Divorce rate of married women less than 35 years old	57	28	9.1
Divorce rate of married women aged 35-44	72	36	13
Children in custody due to divorce (per 100 residents aged 0-17)	0.31	0.17	0.058
Separation rate	1.7	1.2	0.46
Separations (per 100 marriages)	37	24	9.3
Separation rate of married women less than 35 years old	15	10	3.8
Separation rate of married women aged 35-44	10	6.8	2.8
Children in custody due to separation (per 100 residents aged 0-17)	0.81	0.53	0.2
<i>Openness</i>			
Trade (exports+imports, 1,000 Euros)	9,063,175	1,865,859	128,255
Net inter-province migration (per 1,000 residents)	5.2	1.3	-4.3
Rate of inter-province mobility of resident population	26	18	14
Foreign-born residents (per 1,000 residents)	57	35	9
Inflow of graduates (per 100 graduates who left)	221	111	47
<i>Quality of schools</i>			
Teachers below 40 years old: pre-school	26%	18%	7.4%
Teachers below 40 years old: primary school	27%	22%	13%
Teachers below 40 years old: middle school	13%	7.8%	4.7%
Teachers below 40 years old: high school	12%	7.6%	5.3%
Teachers under temporary contracts: pre-school	15%	9.1%	1.5%
Teachers under temporary contracts: primary school	14%	9.1%	2.6%
Teachers under temporary contracts: middle school	27%	21%	15%
Teachers under temporary contracts: high school	28%	22%	15%

Table A17: Statistics on province-level indicators used in the analyses of mobility and local characteristics (part 2/2)

Variable	Source
Value added per resident	ISTAT
Crime:	
Burglaries (per 100,000 residents)	ISTAT
Smuggling offenses (per 100,000 residents)	ISTAT
GTAs (per 100,000 residents)	ISTAT
Petty thefts (per 100,000 residents)	ISTAT
Thefts with tear (per 100,000 residents)	ISTAT
Murders (per 100,000 residents)	ISTAT
Distraints (per 1,000 residents aged 18 or more)	ISTAT
Exploitation-of-prostitution offenses (per 100,000 residents)	ISTAT
Protests of checks (per 1,000 residents)	ISTAT
Drug dealing offenses (per 100,000 residents)	ISTAT
Scam offenses (per 100,000 residents)	ISTAT
Educational attainment:	
Students repeating school year	Tuttoscuola
School dropouts	Tuttoscuola
Education level achieved	Tuttoscuola
Family instability:	
Divorces (per 100 marriages)	ISTAT
Separations (per 100 marriages)	ISTAT
Children in custody due to divorce (per 100 residents aged 0-17)	ISTAT
Children in custody due to separation (per 100 residents aged 0-17)	ISTAT
Separation rate of married women less than 35 years old	ISTAT
Separation rate of married women aged 35-44	ISTAT
Divorce rate of married women less than 35 years old	ISTAT
Divorce rate of married women aged 35-44	ISTAT
Strong labor market:	
Youth unemployment rate (age 15-24)	ISTAT
Female unemployment rate	ISTAT
Male unemployment rate	ISTAT
Long-term unemployment rate (12 months or more)	ISTAT
Youth employment rate (age 15-24)	ISTAT
Female employment rate	ISTAT
Male employment rate	ISTAT
Youth labor force participation (age 15-24)	ISTAT
Female labor force participation	ISTAT
Male labor force participation	ISTAT
Employment rate (high school diploma)	ISTAT
Employment rate (college degree or higher)	ISTAT

Table A18: Indicators used in the principal component analysis of Section 7 (part 1/5)

Variable	Source
Life expectancy:	
Female life expectancy at birth	ISTAT
Male life expectancy at birth	ISTAT
Female life expectancy at 65	ISTAT
Male life expectancy at 65	ISTAT
Economic openness:	
Trade (exports+imports)	ISTAT
Rate of inter-province mobility of resident population	ISTAT
Net inter-province migration (per 1,000 residents)	ISTAT
Social capital:	
Blood bags collected per resident (average within province)	GSZ (2004)
Measure of self-efficacy (average within province)	GSZ (2016)
Non-profit organizations per resident (average within province)	GSZ (2016)
Voter turnout, House of Representatives	ISTAT
School quality:	
School assets	Tuttoscuola
Local government spending in education	Tuttoscuola
Availability of teaching materials and technologies	Tuttoscuola
Quality of school buildings	Tuttoscuola
Students per class	Tuttoscuola
Students using canteen, school bus and other services	Tuttoscuola
Extended teaching-time availability	Tuttoscuola
Processing time of teachers' rankings	Tuttoscuola
Teaching hours per class	Tuttoscuola
Share of male and young teachers	Tuttoscuola
Teachers employed under temporary contracts	Tuttoscuola
Tenure and stability of teachers' position	Tuttoscuola
INVALSI scores for primary and middle schools	Tuttoscuola
INVALSI scores for high school	Tuttoscuola
End-of-year grades at middle and high school	Tuttoscuola
High-school graduation exam	Tuttoscuola

Table A19: Indicators used in the principal component analysis of Section 7 (part 2/5)

Variable	Source
Pre-school quality:	
Local government spending in pre-school education	Tuttoscuola
Students per class	Tuttoscuola
Students per class at private pre-school	Tuttoscuola
Share of students using canteen	Tuttoscuola
Share of students using canteen at private pre-school	Tuttoscuola
Share of students using school bus	Tuttoscuola
Share of students using school bus at private pre-school	Tuttoscuola
Share of students participating in earlier school start time	Tuttoscuola
Availability of saturday classes at private pre-school	Tuttoscuola
Share of students attending more than 40h/week at private pre-school	Tuttoscuola
Processing time of teachers' rankings (band I)	Tuttoscuola
Processing time of teachers' rankings (band III)	Tuttoscuola
Teaching hours per class	Tuttoscuola
Ratio of students to support teachers	Tuttoscuola
Share of male teachers	Tuttoscuola
Share of teachers below 40 years old	Tuttoscuola
Share of teachers employed under temporary contracts	Tuttoscuola
Share of teachers retiring/resigning	Tuttoscuola
Share of teachers transferred to other school/position	Tuttoscuola
Primary school quality:	
Local government spending in primary school education	Tuttoscuola
Students per class	Tuttoscuola
Share of classes with students of different age	Tuttoscuola
Share of primary schools with canteen	Tuttoscuola
Share of students using canteen	Tuttoscuola
Share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of students using extended teaching time	Tuttoscuola
Processing time of teachers' rankings (band I)	Tuttoscuola
Processing time of teachers' rankings (band III)	Tuttoscuola
Teaching hours per class	Tuttoscuola
Ratio of students to support teachers	Tuttoscuola
Share of male teachers	Tuttoscuola
Share of teachers below 40 years old	Tuttoscuola
Share of teachers employed under temporary contracts	Tuttoscuola
Share of teachers retiring/resigning	Tuttoscuola
Share of teachers transferred to other school/position	Tuttoscuola
Share of students repeating school year	Tuttoscuola

Table A20: Indicators used in the principal component analysis of Section 7 (part 3/5)

Variable	Source
Middle school quality:	
Local government spending in middle school education	Tuttoscuola
Students per class	Tuttoscuola
Share of students using canteen	Tuttoscuola
Share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of students using extended teaching time	Tuttoscuola
Processing time of teachers' rankings (band I)	Tuttoscuola
Processing time of teachers' rankings (band III)	Tuttoscuola
Teaching hours per class	Tuttoscuola
Ratio of students to support teachers	Tuttoscuola
Share of male teachers	Tuttoscuola
Share of teachers below 40 years old	Tuttoscuola
Share of teachers employed under temporary contracts	Tuttoscuola
Share of teachers retiring/resigning	Tuttoscuola
Share of teachers transferred to other school/position	Tuttoscuola
Share of students repeating school year	Tuttoscuola
High school quality:	
Value of books, teaching/office materials, and equipment available to high schools	Tuttoscuola
Students per class	Tuttoscuola
Share of students using canteen	Tuttoscuola
Share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of classes with extended teaching time	Tuttoscuola
Year-over-year change in share of students using extended teaching time	Tuttoscuola
Processing time of teachers' rankings (band I)	Tuttoscuola
Processing time of teachers' rankings (band III)	Tuttoscuola
Teaching hours per class	Tuttoscuola
Ratio of students to support teachers	Tuttoscuola
Share of male teachers	Tuttoscuola
Share of teachers below 40 years old	Tuttoscuola
Share of teachers employed under temporary contracts	Tuttoscuola
Share of teachers retiring/resigning	Tuttoscuola
Share of teachers transferred to other school/position	Tuttoscuola
Share of students repeating school year	Tuttoscuola
Share of students dropping out	Tuttoscuola
Share of students continuing to college	Tuttoscuola

Table A21: Indicators used in the principal component analysis of Section 7 (part 4/5)

Variable	Source
School structures and resources:	
School assets	Tuttoscuola
Local government spending in education	Tuttoscuola
Availability of teaching materials and technologies	Tuttoscuola
Quality of school buildings	Tuttoscuola
School organization and services:	
Students per class	Tuttoscuola
Students using canteen, school bus and other services	Tuttoscuola
Extended teaching-time availability	Tuttoscuola
Processing time of teachers' rankings	Tuttoscuola
Teachers' composition:	
Teaching hours per class	Tuttoscuola
Share of male and young teachers	Tuttoscuola
Teachers employed under temporary contracts	Tuttoscuola
Tenure and stability of teachers' position	Tuttoscuola
Students' grades and test scores:	
INVALSI scores for primary and middle schools	Tuttoscuola
INVALSI scores for high school	Tuttoscuola
End-of-year grades at middle and high school	Tuttoscuola
High-school graduation exam	Tuttoscuola

Table A22: Indicators used in the principal component analysis of Section 7 (part 5/5)