

NBER WORKING PAPER SERIES

"AND YET IT MOVES":
INTERGENERATIONAL MOBILITY IN ITALY

Paolo Acciari
Alberto Polo
Giovanni L. Violante

Working Paper 25732
<http://www.nber.org/papers/w25732>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2019, Revised February 2020

The opinions expressed in this working paper are those of the authors and do not necessarily reflect those of their institutions, nor of the National Bureau of Economic Research. We are grateful to Alberto Alesina, Damien Capelle, Lorenzo Cappellari, Miles Corak, Nicola Curci, David Lee, Salvatore Lo Bello, Davide Malacrino, Costas Meghir, Salvatore Morelli, and Danny Yagan for useful comments. We thank Simon Boserup for sharing the intergenerational transition matrix for Denmark and Enrico Moretti for sharing his data on province-level cost-of-living indexes.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Paolo Acciari, Alberto Polo, and Giovanni L. Violante. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

"And Yet It Moves": Intergenerational Mobility in Italy
Paolo Acciari, Alberto Polo, and Giovanni L. Violante
NBER Working Paper No. 25732
April 2019, Revised February 2020
JEL No. J31,J61,J62,R1

ABSTRACT

We link tax returns across two generations to provide the first estimate of intergenerational mobility in Italy based on administrative income 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 concerns, both standard when using administrative data and specific to our sample. A 10 percentile increase in parental income is associated with a percentile increase in child income between 2.5 and 3. The expected rank of children born from parents with income below the median is around 0.43. 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 three times as large as those in the South. This regional variation is strongly correlated with local labor market conditions, indicators of family instability, and school quality.

Paolo Acciari
Italian Ministry of Economy and Finance
Via XX Settembre 97
Rome 00187
Italy
paolo.acciari@mef.gov.it

Giovanni L. Violante
Department of Economics
Princeton University
Julis Romo Rabinowitz Building
Princeton, NJ 08540
and NBER
glv2@princeton.edu

Alberto Polo
Department of Economics
New York University
19 west 4th street
New York, NY 10012
ap3562@nyu.edu

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 formidable. Very few publicly available datasets around the world have information that allows 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. There exists, so far, no study 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. [Cecchi 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 Survey of Household Income and Wealth (SHIW) or in the Italian sample 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 mobility has witnessed a strong revival thanks to the fact that economists have been able 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 indexes 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 about what socio-economic factors are strongly correlated with upward mobility across regions, while controlling for the institutional framework.

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

In this paper we add to this recent wave of studies and introduce 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 extracted a sample of two cohorts of Italians born, respectively, between 1942-1963 and 1972-1983 and we matched parents and children through their social security numbers. Our final dataset contains nearly 650,000 parents-child pairs with detailed income information for three years in each cohort - 1998-99-00 for parents, 2011-12-14 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 (thus, on average the 25th percentile) is expected to reach the 44th 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 deciles, for a child from parents in the top decile, the probability of keeping her parents' rank as an adult is 26%. For a child in the bottom decile, the probability of rising to the top decile is less than 4%. We also find that upward mobility is larger for sons, for first-born among siblings, 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 confirm the finding of [Chetty et al. \(2014\)](#) that this relation is markedly linear, except at the very top where it 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. In Italy, the RRS is 0.25. 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 43,500 Euros. An RRS of 0.25 means that, when adults, these children will be on average still two deciles apart,

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).

a gap that translates into nearly 6,500 Euros of annual earnings. The median rank-rank slope is much higher, 0.36. 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.

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, a 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 0.25. However, we uncover that this elasticity varies dramatically across the distribution: at the bottom it is nearly zero, while above the 10th percentile its value exceeds 0.30. This sharp nonlinearity makes it problematic to use the IGE as a single statistic to summarize mobility across the whole distribution. When we replicate imputation procedures used by previous studies to estimate the IGE in Italy, we obtain much higher point estimates, in line with those findings. 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 three main shortcomings of our data are: (i) an attenuation bias due to the short within-individual panel dimension; (ii) a selection bias due to the fact that we can only link parents to children when the latter are claimed as dependent by the former; (iii) possible distortions arising from tax evasion for the self-employed. Correcting for these three sources of bias increases somewhat our estimates of intergenerational rank persistence, e.g. the RRS rises above 0.30. However, measures of upward mobility remain largely unchanged.

When placing our estimates in a comparative context, upward mobility in Italy is higher than in the United States, but lower than in Nordic European countries. However, when using alternative measures of mobility (namely, the probability that the son earns at least 50% more than the father) that account for the different level of income inequality between Italy and the United States, and hence distance between percentiles, 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 by parents below the median). 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. The level of upward mobility in Northern Italy exceeds that of Scandinavia. Provinces with higher upward mobility systematically display also lower downward mobility.

To give a sense of the magnitudes, we compare the province of Milano (ranked 10th in terms of upward mobility) to that of Palermo (ranked 107th), both very large provinces with 3.2M and 1.2M residents, respectively. In Milan, children from parents with income below the median in the national parental income distribution are expected to reach the 54th percentile of their own national income distribution, once adults. Children from parents with the same economic background who grow up in Palermo reach, on average, the 38th percentile. The probability of moving from the bottom quintile to the top quintile of the national income distribution is 0.22 for children who grow up in Milano and less than 0.06 for those raised in Palermo.

We uncover a Great Gatsby curve with a negative slope linking upward mobility and several measures of 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 indexes 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 a large portion of the geographic variation in rates of upward mobility (nearly 90%). The key explanatory variables are the local labor market conditions, indicators of family instability and two specific indexes of school quality: students' grades and test scores, and the quality of pre-school.

These correlations do not imply causality. We make an attempt to identify a causal link by exploiting households who are likely to have moved exogenously because of an earthquake that struck three provinces in the South of Italy in 1980. By comparing movers and stayers, we find statistically significant differences in upward mobility that are almost as large as the differences between mobility indexes of provinces of destination and provinces of origin. This exercise supports the idea that fixed characteristics of the province, over and above selection in the pool of residents, has an independent impact on the degree of intergenerational mobility.

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 mobility

In this section, we briefly discuss the measures of intergenerational 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 parent 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 0 to 99) 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. In other words high relative mobility corresponds to a low value for β .

For example, by computing $\Delta \cdot \beta$ we can answer the question: what is the difference in

expected rank between two children with parents who are Δ percentiles apart in the national income distribution? And, by simple iteration we can ask how many generations it would take, on average, for descendants of families originally Δ percentiles apart to belong to the same percentile of the income distribution, i.e. the value N that solves $\beta^N \Delta = 1$.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} . We then define

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

The higher this ratio, the stronger 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 four advantages of the RRS compared to the IGE. First, the IGE is based on log

³This 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 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.

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.

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 conditional expectation of log child income as a function of parental income, in practice, is often very non-linear. As a result, a single elasticity is not a good summary of intergenerational mobility at all points in the distribution. Conversely, the empirical literature that uses measures of positional mobility showed that the rank-rank relation is remarkably close to linear. Third, 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 Fourth, the RRS is less sensitive than the IGE to lifecycle and attenuation biases (Nybom and Stuhler, 2017).

2.2 Absolute mobility

Absolute mobility indexes 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), 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:

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

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

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

the bottom:

$$Q1Q5 = \Pr \left\{ R_i \geq 80 | R_i^P < 20 \right\}. \quad (5)$$

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 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 not the case for the regional regressions, since the average national rank of residents of a specific region needs not be the median. However, as long as linearity holds, the AUM for region g can 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 on lower values than in rich ones, thus affecting local estimates of the AUM.

2.4 Indicators for cross-country comparisons

We are also interested in comparing the level of intergenerational mobility in Italy with that of other countries. The key challenge in this comparison 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 do cross-country comparisons, we report another measure beyond positional indexes. The measure has been used previously by [Chetty et al. \(2017\)](#) to compare mobility between cohorts in the United States. 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 by their mean to capture 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.

3 Data

In this section, we describe our dataset, outline our sample selection procedure and present some descriptive statistics. We argue that the earnings distribution in our sample is comparable to the distribution obtained from more standard survey data, hence it is representative.

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.⁷ The accuracy of the

⁶Consider, for example, country A and country B , where country A is a lot less unequal than B . When using indexes 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.

⁷In Italy, the tax unit is the individual.

information in this database is preliminarily statistically checked and validated by the Department of 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). However, it is the first time that the dataset is used to link children to their parents in order to construct measures of intergenerational 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 tax returns consist 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 is not required to submit a tax return directly. This is mostly the case of individuals with 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 into professional and entrepreneurial income), pensions, unemployment benefits, financial assets, housing, land and farms.^{11,12}

⁸This process is mainly performed by SOGEI SpA, an in-house company of the Ministry of Economy and Finance, through a series of algorithms that checks 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 in our dataset. These individuals are required to report it in the tax return, as the code describes the specific economic activity that generates the income.

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 extensively 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 includes income from owned firms, i.e., firms where individuals own a share of total equity (partnerships, closely-held companies).

3.2 Sample construction

Our aim is to link parents and children in the database. This operation is possible because, 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. Data for parents are extracted from tax returns of year 1998 by selecting all taxpayers who claimed allowances for children born between 1972 and 1983 (age 15-26 in 1998).¹³ Then through children's SSNs, we match the tax return of the parents to that of the children in tax year 2012. Our cohort of children in 2012 is therefore of age 29-40.

Our initial dataset with information on both child and parents comprises of 2,846,994 matched records. Each record contains information on a child, her father, and her mother.¹⁴ Based only on the tax return, we cannot be sure that the individuals claiming the child deduction are the parents (e.g., they could be older brothers, or grandparents).¹⁵ Therefore, we drop records where the age difference between child and father is less than 15 or larger than 45 which reduces sample size to 2,835,920 records. This criterion minimizes instances where the individual claiming the dependent on the tax return is not the parent.

In what follows, we discuss three potential problems with our sample: (i) selection bias, (ii) attenuation bias, and (iii) life-cycle bias. We explain why these issues are not severe and later, in Section 5, we propose explicit corrections for each of them.

3.2.1 Sample selection bias

The matching procedure outlined above inevitably misses some parent-child pairs for three reasons. First, in 1998 the child's SSN was mandatory only for forms MU and 730. 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 income from housing and land

¹³Year 1998 is the first one in which parents were required to report the SSN of the dependent children on the tax form.

¹⁴From the tax returns information, we can always match spouses unless they are divorced and live apart. We can retrace divorced couples when they both claim a positive percentage of deduction for the same individuals born in 1972-83.

¹⁵While tax filers have to specify if the dependent is a child by checking a box, there could be errors in reporting.

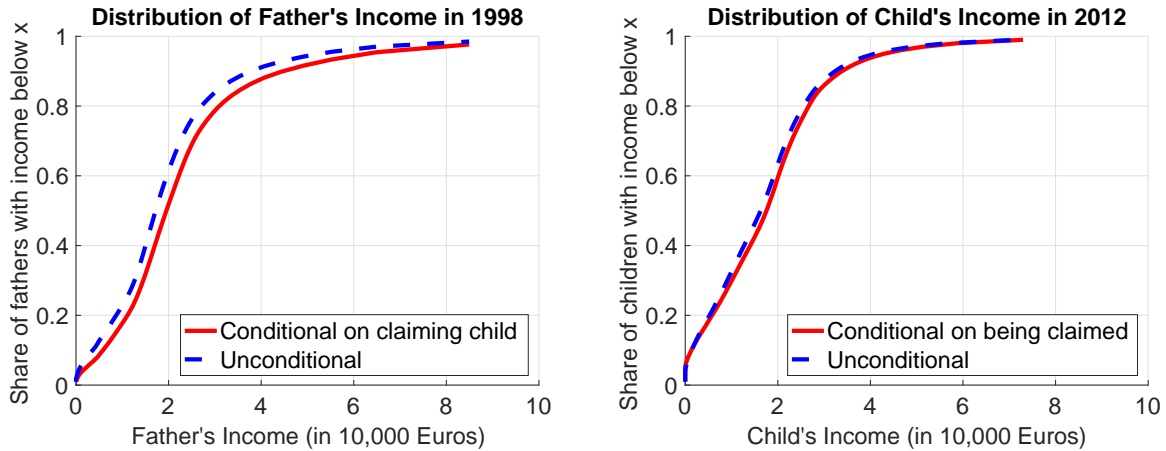


Figure 1: Unconditional and conditional distributions of father's and child's income.

assets below 500 euros, (iv) with income exclusively from interests, dividends, and capital gains deriving from financial assets. Because of (i-iii), the very poor and the extreme tax evaders are not covered in our sample.¹⁶ In Section 5, we bring in additional data to deal with these two possible sources of bias.

The third reason why our sample may not be representative is that a share of individuals aged 15-26 in 1998 were no longer dependent and thus they cannot be matched to parents through tax returns. Our data gives us access to all children aged 29-40 in 2012, but only to parents who claim some of those children as dependent in 1998. From the Ministry of Economy and Finance, we obtained additional tabulations of the income distribution of all taxpayers in the same age group of the linked parents in 1998, i.e. all potential parents of those children. Figure 1 compares the conditional (on having a dependent child) distributions of father income in 1998 and the conditional (on being claimed as dependent) distribution of child income in 2012 to their unconditional counterparts. In both cases, the conditional and unconditional distributions are remarkably similar.

Figure 2 goes one step further and compares the distributions for all children (left panel) and for dependent ones (right panel) across some key characteristics.

¹⁶Individuals in group (iv) are very few because it is unlikely that someone in that group would not have other types of taxable income.

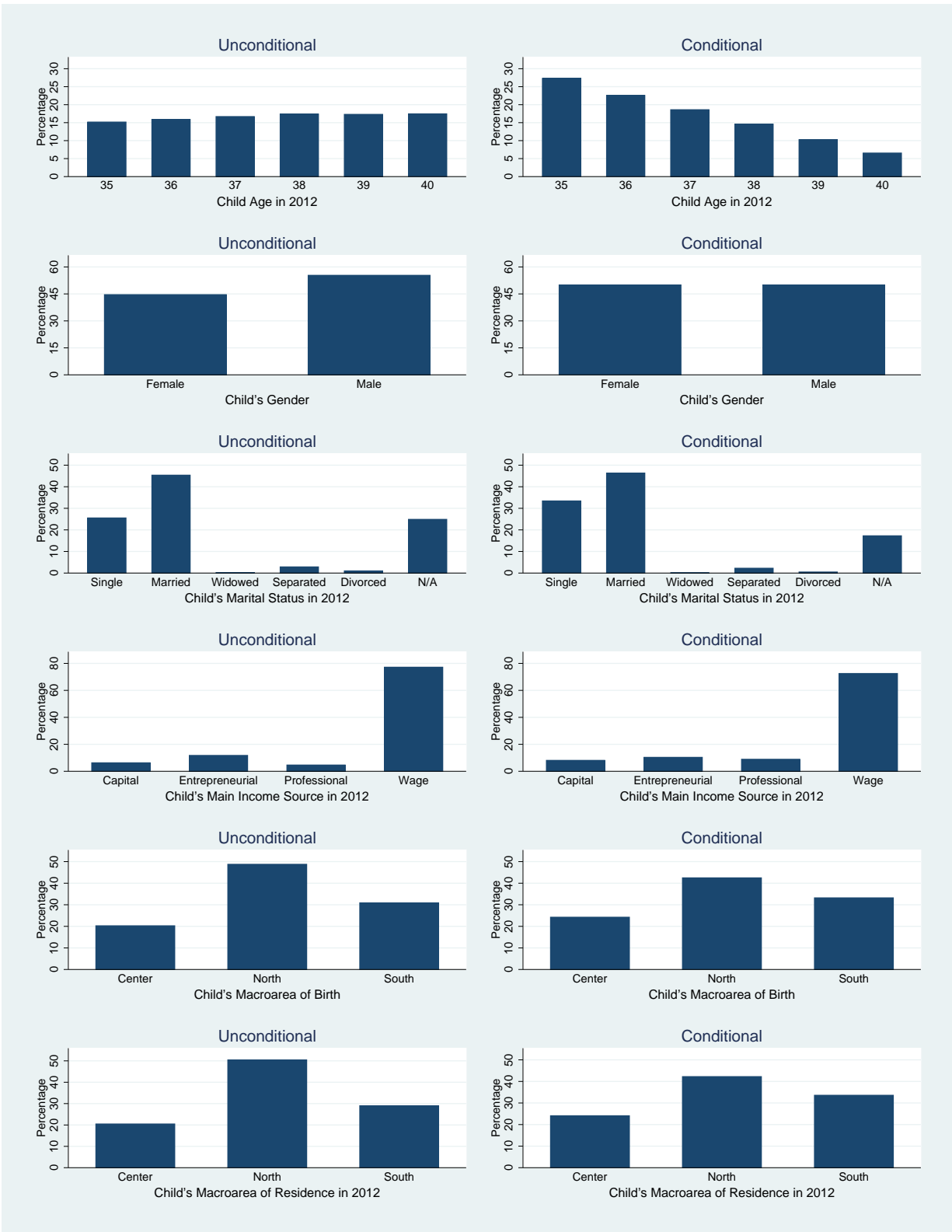


Figure 2: Unconditional and conditional (on being claimed as dependent) distributions of various characteristics for children in 2012.

In general, the distributions are quite close to each other. The only exception is, not surprisingly, age: children claimed as dependent in 1998 are younger on average. However, as we explain in Section 3.2.3 below, in the relevant age range our measures of intergenerational mobility are very stable.¹⁷

This preliminary analysis suggests that our sample of dependent children is representative of the population. In Section 5 we reinforce this conclusion with an additional exercise.

3.2.2 Attenuation bias

After matching children in 2012 to their parents in 1998, we are able to add two years of data of tax returns for both generations, 2011-2014 and 1999-2000 respectively. For these two years we only have total income, not each of its components, and hence throughout most of the analysis our definition of income is total gross income.

As extensively discussed in the literature (e.g., Mazumder, 2016), a short income span means our analysis is susceptible to attenuation bias. As a result, intergenerational mobility can be overestimated: a fraction of the mobility measured could be spuriously due to the fact that transitory shocks attenuate the income correlation between parents and children.

To deal with the attenuation bias, throughout our analysis we use measures of income averaged over the three years for parents and children.¹⁸ Ideally, one would like to ascertain that the correlation between the 3-year average and lifetime income for our sample is high. Even if we do not have access to a longer income span in our own dataset of tax returns, we performed this calculation on the Italian Social Security (INPS) dataset made available to us by Hoffmann and Malacrino (2019). In this long longitudinal sample (1985-2016), the rank correlation between a 3-year average of earnings for individuals of the age of our sons (37) and our fathers (51) and the lifetime income of these same individuals computed between ages 27-55 is 0.93, giving us confidence that we have a strong proxy for permanent income.¹⁹ We return on this point in Section 5, where we also assess the size of the attenuation bias by simulation.

¹⁷Figures A3-A5 in the Appendix confirm these findings hold also for regions.

¹⁸Figure III.B in Chetty et al. (2014) shows that for the US data the attenuation bias is small: their estimates of the RRS using 2 and 15 years of income are very close.

¹⁹INPS 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.

3.2.3 Lifecycle bias

Another problem in using only a short snapshot of lifetime income is that, if we catch children earlier in their working life relative to parents, we may understate income more for those with steep age-income profiles (e.g., the more educated), which also leads to an understated estimate of intergenerational persistence when the age-income slopes are positively correlated across generations.

Figure 3 illustrates this problem in our data. The figure plots the value of the RRS computed on samples with different age cut-off criteria for children and fathers. Panel (a) shows that including children below age 35 would dramatically reduce our estimates of rank persistence, but after age 35 the estimate levels off.²⁰ Clearly, by selecting older children in 2012, the age of fathers in 1998 mechanically increases. Panel (a) shows that the effect of mean parental age on the RRS is very modest until age 55, when it starts becoming considerable.

In light of these findings, to deal with lifecycle bias, we impose two selection criteria. First, we drop children younger than 35 in 2012. This selection reduces the sample size from 2,835,920 to 1,067,195 pairs. The oldest child in our dataset is 40 years old, and hence our final sample covers a 6-year age range for children. Second, we only retain children whose fathers and mothers in 1998 were between the ages of 35 and 55. After imposing this second restriction, the sample size shrinks to 647,662 records. This is our final baseline sample.

3.3 Descriptive statistics

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

On average, fathers (mothers) are 51 (48) 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 2/3 of families have two sources of income. The median gross income of fathers is nearly 21,000 Euros in 1998 and that of mothers is approximately half of it.²¹ The correlation between the two (which excludes zeros) is positive and significant, around 0.2.

²⁰Figure III.A in [Chetty et al. \(2014\)](#) is the counterpart of our Panel (a) for the US and Figure 1c in [Nyblom and Stuhler \(2017\)](#) is the counterpart for Sweden. In both countries, the RRS levels off at around age 30, i.e. five years earlier than in Italy. This is not surprising, since entry in the labor market in Italy is much delayed relative to both US and Sweden.

²¹Italy 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 2011-12-14.

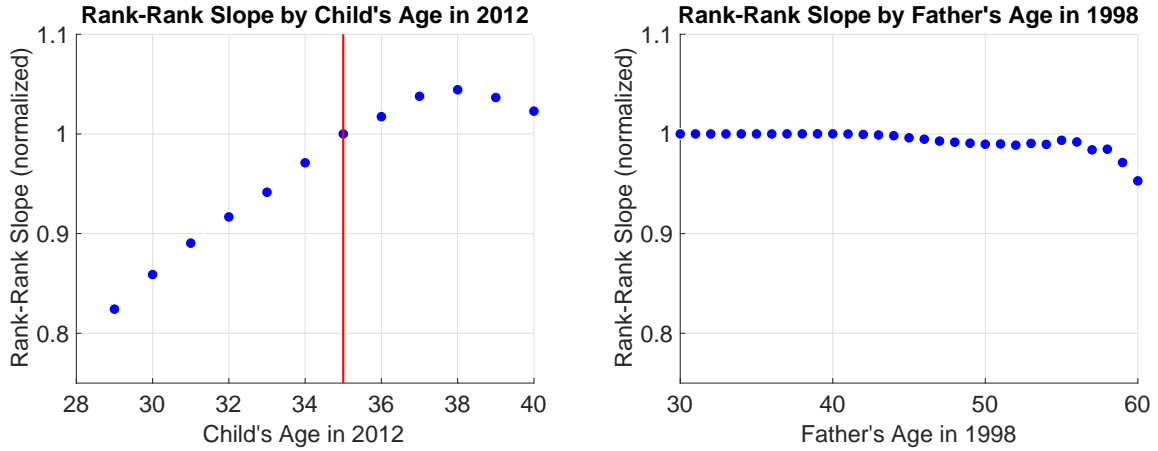


Figure 3: Left panel (a): Estimates of rank-rank slope (RRS) for different age cut-offs for the child. Right panel (b): Estimates of RRS for different age cut-offs for the father. In both cases, the RRS is normalized to one when the cut-off is age 35 (the one used in the analysis).

Children are on average 37 years old in 2012 and less than half are married. Median nominal income of sons is only 4% higher than that of their fathers 14 years earlier. This small difference reflects the age gap between the two groups, but also the dismal aggregate growth of the Italian economy over this period.²² Instead, women's income is substantially higher in 2012 reflecting the steady rise in female employment rate among the newer cohorts. The composition of income reveals that for nearly 72% of children compensation for dependent labor is the major component of income, followed by self-employment for 21% of the children population. Fewer than 6% of children are rentiers—their main source of income being financial and housing capital.

Additional statistics about 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 6.6% of total income in 2012.²³

²²Average annual GDP growth per capita from 1998-2012 is less than 0.5%. Moreover, as in many developed countries, much of the growth accrued to the top earners so it is less visible for median income.

²³In the total Italian population, without the age restrictions we impose, this share is close to 10%. For comparison, in the U.S. this share is around 20%

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), administered by the Bank of Italy. In both datasets we select all individuals (not just dependent) aged 35-40 in 2012. Two additional adjustments are performed to make the two samples more comparable. First, following [Cannari et al. \(1997\)](#), we drop individuals who *i)* report zero or missing income in the survey, *ii)* report housing asset income below 185 Euros, and *iii)* report only interest income. These criteria exclude from SHIW individuals who do not file tax returns. Second, because our dataset comprises only gross income, while SHIW reports after-tax income data, we apply an algorithm supplied by the Italian Ministry of Economy and Finance to reconstruct net wage income (by far, the largest component of income for the vast majority of the population, as mentioned above) in our data.

Table 2 compares statistics on the after tax earnings distribution of in the two datasets. Overall, the two distributions line up well.

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 total pre-tax 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 distribution of parental incomes and the rank of children R_i as their percentile in the distribution of child incomes. Figure 4 presents a binned scatter plot of the mean percentile rank of children as a function of their parent 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 almost perfectly linear, except at the top of the income distribution where it bends

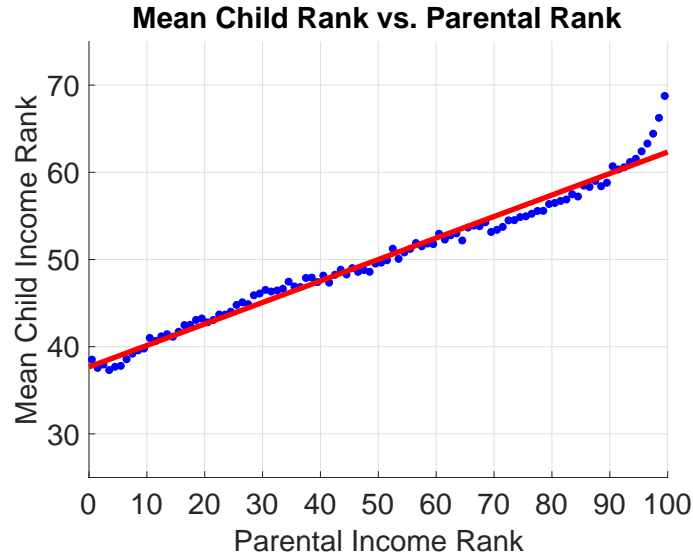


Figure 4: Child mean rank conditional on parental income rank. Blue dots: data. Red line: linear fit. The constant of the red line is 37.7 and the slope 0.246.

upward.²⁴

Running the OLS regression in (1), we estimate that a one percentage point (ppt) increase in parental rank is associated with a 0.25 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 in the top decile of the income distribution and one in the bottom decile, a generation later the child of the rich family is expected to be still two deciles above the child from the poor family. On average, it would take three generations for these differences in initial conditions to be fully offset so that the descendants of the two families (i.e., their great-grandchildren) would be expected to belong to the same percentile.

The first column of Table 3 reports also other measures of mobility. The top mobility ratio (TMR) is 3.7, a reflection of the fact that in Figure 4, the rank-rank curve steepens towards the upper end. Thus, ranks persist a lot more at the top of the income distribution.²⁵ This value of the TMR implies that in the top decile the slope is close to 1: the children of two families, both

²⁴This 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). In these studies, the data also show significant deviations from linearity (less mobility) only at the extremes of the distribution.

²⁵Mechanically, this happens because 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.

in the top decile but at the two extremes of the decile, will still be nearly 10 percentiles apart a generation later.

The AUM index equals 0.44 which means that a child of parents with income below the median is expected to end up in the 44th percentile of her income distribution. The Q1Q5 index implies that, at the national level, the probability of belonging to the top quintile of the income distribution, for children born in the bottom quintile, is 10 percent.

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 25 percent, almost 7 times higher than the probability that a child from a family in the bottom decile ascends to the top one as an adult.

4.1 Robustness to 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. A location-specific deflating procedure might therefore affect the ranks of parents and children in the national distribution.

To examine this issue, we used regional consumer price indexes constructed by the Bank of Italy for the year 2006 (Cannari and Iuzzolino, 2009). We explore four different indexes: the first one is a price index that allows for regional dispersion in prices of food, clothing, and furniture only. The other three indexes include alternative ways of accounting for the location-specific cost of housing services and other services. We also used an alternative deflating procedure based on the province-level price indexes for year 2005 constructed by Boeri et al. (2018) following the methodology proposed by Moretti (2013). The cross-regional correlation between all these cost-of-living indexes is very high, ranging from 0.53 to 0.99.²⁶

Overall, these price indexes reveal cost-of-living differentials of up to 30 percent between the least expensive regions (e.g., Calabria) and the most expensive ones (e.g., Lombardia). Table 9 shows that our national results are quite robust to alternative deflating procedures. If anything, our estimates of relative mobility rise somewhat when we use these regional deflators, but

²⁶Table A2 in the Appendix reports all pairwise correlations between the indexes.

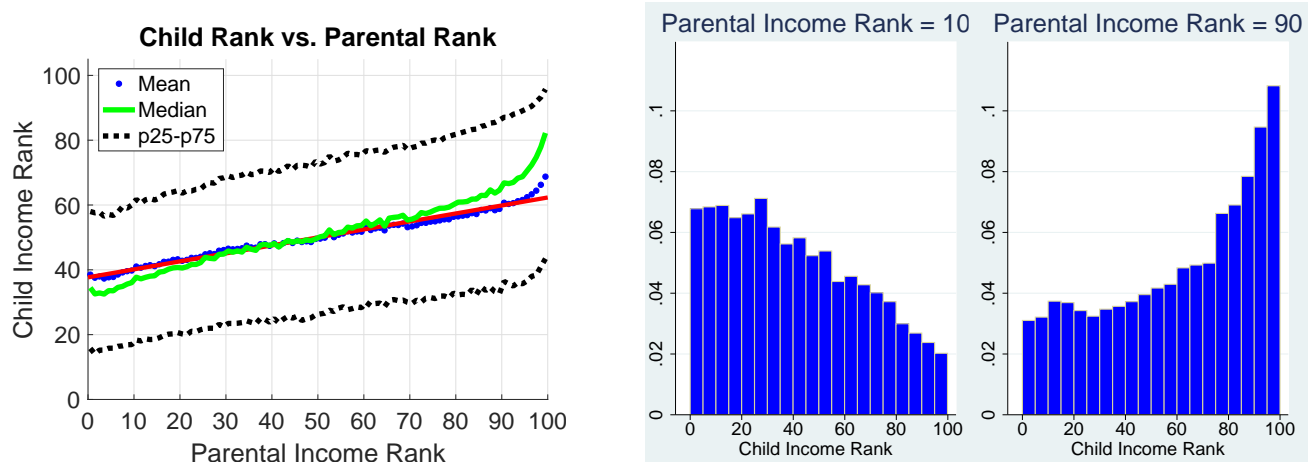


Figure 5: Left panel (a): Mean child rank, median child rank, and interquartile range of child ranks conditional on the parental income rank. Right panel (b): conditional distributions of child ranks at parental income percentiles 10 and 90.

indexes of absolute mobility are nearly identical.²⁷

4.2 Conditional distributions of child ranks

We now return to our baseline sample. Figure 5(a) reports the median child rank and the 25-75 percentile range of child ranks conditional on the parental income rank in addition to the mean rank already reported in Figure 4. The relationship between the median child rank and parental rank remains linear, except at the very top, but it is a lot steeper: the rank-rank slope is 0.36 for the median. The discrepancy between mean and median is due to the fact that the conditional distributions of child ranks are very skewed, as illustrated in Figure 5(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, which induces a steeper positive relationship between parents' rank and conditional child median rank.

Figure 5(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 conditional distribution of child ranks, averaged across parents' ranks, is 0.47.²⁸ For example,

²⁷ A dampening of relative rank persistence is what one would expect if high income areas are the high cost-of-living ones. However, when there is a great deal of within-province income variation, the attenuation is small.

²⁸ Interestingly, Boserup et al. (2017) also report for Denmark an interquartile range of the conditional distribu-

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.²⁹ Put differently, even though the slope of the rank-rank relation is strongly statistically significant, the regression has a very low R^2 , around 0.06.

This finding is not uncommon in the literature and recently has led to interest in methods that combine multiple parental predictors (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: marital status, age, gender, a self-employment dummy, and a geographical mover dummy. Figure 6 shows that, jointly, these variables explain around 15% of the within-parental-rank variation.³⁰ 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: around 1/3 up to the median, and less and less as parental rank rises. 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 remains unexplained, suggesting that there is an enormous amount of unobserved heterogeneity left, even within parental rank and within province.

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. For completeness, we report this measure as well. Figure 7 plots the relation between log income of the child and log income of the parents, and the share of observations dropped among the

tion of child income ranks around 0.4.

²⁹The dispersion in the distribution of ranks is not substantially affected by gender. When considering the income distribution of sons only, the interquartile range is 0.455.

³⁰This 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.

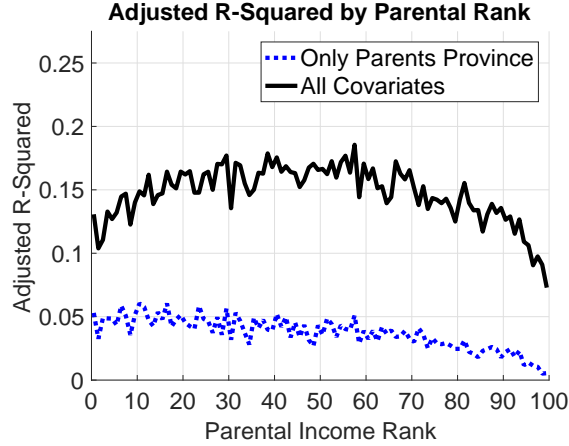


Figure 6: Adjusted R-squared of the 100 within-parental rank regressions of individual child rank in the national distribution on various parental and child covariates.

children's records. There is only 1.5% of observations dropped across the entire distribution, so this does not seem a serious problem for our dataset.³¹

The figure reveals that the relationship is very far from linear, especially if one includes the bottom decile of the distribution. The overall IGE is 0.25, but for the bottom decile it is 0.04 and for the top nine deciles it is 0.32.³² In other words, the IGE in our sample has the serious limitation that it is not a sufficient statistics to characterize mobility across the entire distribution.³³ In this regard, the linearity of the rank-rank relation offers a significant advantage even though, as we showed, deviations from linearity emerge at the top for this measure.

4.3.1 Comparison with existing estimates

Other authors before us 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

³¹ Approximately 0.3% of observations are dropped because parents' income is negative or zero, while the rest is dropped because children's income is negative or zero.

³² 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 2.5 in our dataset).

³³ Chetty et al. (2014) show this is the case for the US as well. In their sample, the relation between log income of parents and children flattens out also at the top, differently from what emerges from Italian data.

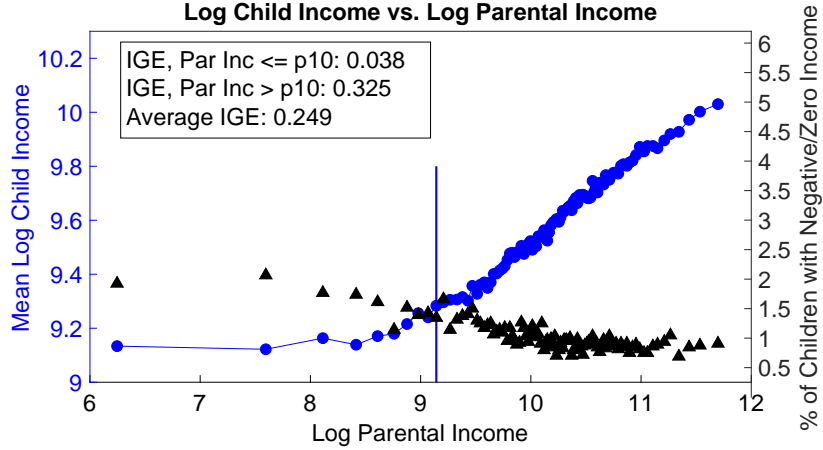


Figure 7: Relation between log income of the child and log income of the parents. The vertical line denotes the 10th percentile. The figure also reports the fraction of children with negative or zero income at each bin (i.e., the percentage of observations dropped).

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 (2,000-3,000 observations) (Barbieri et al., 2018; Cannari and D’Alessio, 2018).³⁴

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 father’s province of birth as an instrument. An F-statistic over 200 for this first stage rules out a weak instrument problem. The IGE estimated from observed child income and imputed father’s income is 0.40. This value is much larger than our actual estimate of IGE and in the range of existing estimates reported above.³⁵

These 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 in-

³⁴This 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.

³⁵It is also much more precisely estimated (S.E. = 0.02), given that our sample size is over 200 times larger than that of existing studies

dependent effect on child's income beyond parental income. In our replication, we chose as instrument a variable that is strongly correlated with father's income and yet the least likely ex-ante, among those available, to separately impact child income.³⁶ Yet, when we regress child income on actual father's income and father's province of birth in our dataset with matched fathers and sons where we observe income for both, we find that the instrument remains strongly significant confirming the source of upward bias. In other words, it appears that the key challenge in this approach is finding a valid instrument.

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 core sample of column (1) in Table 3.

Comparing males and females and focusing on the RRS, it appears that relative mobility is higher for women. 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 almost twice as large. One of the determinants behind this result is that in Italy female labor force participation is still quite low (it was below 40 pct in 2013). However, these differences in upward mobility across genders disappear once we estimate AUM and Q1Q5 separately using gender-specific income distributions for children.

When we condition on the major source of parental income (labor, self-employment, and capital) in Table 4, we find significantly higher upward mobility only for the children of self-employed professionals (e.g., artisans, architects, lawyers, doctors, pharmacists): the Q1Q5 is three times larger than for sons of wage earners. Relative rank persistence at the top is extremely high for this group as well: the TMR is above 6.

Table 5 shows that children who, in 2012, reside in a different region from the one where their parents lived in 1998 (movers) display much higher economic mobility. For example, their Q1Q5 transition rate is 0.17 compared to a baseline value of 0.10, and their AUM index is the 54th percentile, relative to a baseline value equal to the 44th percentile. Therefore, a

³⁶Other available variables are: province of residence, marital status, major source of income, occupational code.

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).

4.4.1 Alternative definition of parental income

Tables 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). In both instances, results are very similar to those based on our baseline definition. Tables A6 to A9 in the Appendix report mobility measures and transition matrices considering only income of male children and income of fathers —thus the most comparable definitions across generations. Mobility estimates are very close to their baseline counterparts. The RRS and the Q1Q5 are 0.257 and 0.105, respectively, compared to 0.246 and 0.099 in the baseline.

4.5 Siblings

Our dataset does not contain explicit identifiers for siblings. Given the information available, we can however 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 70,000 parents with at least two children.³⁷ 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 to children who are 36 or 37 years old in 2012 –around the mean age of children in our dataset–, leaving us with around 18,500 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 differentials between siblings is 0.26. Since the mean absolute difference of a uniform distribution is 0.33, we conclude that according to this metric parental background accounts for roughly 20%

³⁷By 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.

of the total variation of individual outcomes.³⁸ This share is in the range of the estimates of siblings correlation reported by Jäntti and Jenkins (2015, Table 9) for a number of countries.

One can refine the analysis by asking whether there is any systematic difference between first-born and later-born children in terms of upward mobility.³⁹ Table 6 shows the results of our analysis. We find that the expected rank of first-born children is between 0.2 and 1.6 percentiles higher. Interestingly, this gap increases with parental income, and it doubles as one moves from the bottom to the top of the parental income distribution. It is statistically significant, however, only for families at the top of the income distribution.

5 Corrections for potential sources of bias

In this Section, we verify the robustness of our analysis with respect to three potential sources of bias: (i) attenuation bias due to the combination of a short sample and volatile income; (ii) dependence bias stemming from the inability to link children who in 1998 are already independent taxpayers to their parents; (iii) 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.

5.1 Attenuation bias

In Section 3 we already argued that a 3-year average of income measured around ages 37 for children and 50 for parents is very strongly correlated with lifetime income. Here, in order to quantify more precisely the attenuation bias in our estimates, we perform the following simulation. We start by estimating a common life-cycle earnings profile from our data. To do this, we pool fathers and sons in order to have data for the longest possible age range and estimate a quartic polynomial in age.⁴⁰ Next, we residualize earnings with respect to this age profile.

³⁸The more traditional estimate of the relative importance of parental background, siblings correlation, is 0.19 in our data and hence almost identical.

³⁹Note that, as said before, we are restricting to siblings with approximately the same age. Some of the children we label as ‘first born’ may have older siblings who we do not observe because of our age restrictions on the sample. However, these cases would only arise in families with at least 3 children, which are only 6% of the total number of Italian families. If birth order matters, and this is what we find, this source of bias attenuates the true effect because of a few spurious cases among those who we identify as ‘first-born’.

⁴⁰We normalize average earnings of the sons to the value of that of the fathers for the range over which they overlap, 30-40.

We assume a time-stationary persistent-transitory process for log residual earnings, a common representation in the literature on income dynamics:

$$\begin{aligned}\log y_{i,t}^j &= \kappa_{i,t}^j + \epsilon_{i,t}^j \\ \kappa_{i,t}^j &= \rho^j \kappa_{i,t-1}^j + \eta_{i,t}^j, \quad t > 0\end{aligned}$$

where $t = 0, \dots, 30$ is age, $j = f, s$ is an indicator for father or son, $\eta_{i,t}^j$ is a persistent shock with variance σ_{η}^j and the $\epsilon_{i,t}^j$ is an uncorrelated shock with variance σ_{ϵ}^j . The two shocks are orthogonal to each other, i.i.d. across all individuals and Normally distributed. Let $\sigma_{\kappa_0}^j$ be the initial standard deviation of the permanent component at age $t = 0$. 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 .

With a three-year panel for each generation, we can estimate the structural parameters $\{\rho^j, \sigma_{\eta}^j, \sigma_{\epsilon}^j, \sigma_{\kappa_0}^j\}$. We provide more details on the identification and estimation strategy in the Appendix. Table 10 reports these parameter estimates. It is remarkable how much more additional volatility sons' earnings display relative to fathers' earnings. We conjecture this is the outcome of the Italian labor market reforms implemented over the last two decades.⁴¹ These reforms introduced flexibility both at entry (with an expansion of temporary contracts) and at exit, but they affected only new hires as the vast majority of older cohorts was hired under the old permanent contracts that were not affected by these changes in legislation.

The experiment then proceeds as follows. We run 100 simulations of a panel of 323,856 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 50-52 for fathers and 38-40 for sons. In these simulations we use both the estimated deterministic age profile and the stochastic component of income. We set $\rho_0 = 0.78$ in order to reproduce our estimated value of the rank-rank slope in the national distribution using the three-year proxy for income. Then we compute the rank-rank slope and other mobility statistics using average lifetime income.

⁴¹This finding remains true even when controlling for age, i.e. when focusing on fathers who, in 1998, have the same age as children in 2012. The two major reforms in our sample period were the Treu Package (1997) and the Biagi Law (2003).

Table 11 reports average mobility statistics across simulations. The bottom line is that the short sample induces a downward bias in relative mobility of around 20%, i.e. the true RRS would be 0.30 instead of 0.25. Our estimates of the AUM and Q1Q5 are even more robust.

5.2 Dependency bias

Our preliminary analysis of Section 3.2 already suggests that the sample of dependent children is quite representative of the universe of children in most dimensions. Nevertheless, here we make a formal attempt to assess the size of the dependency bias.

Recall that in our dataset we have information on all children aged 29-40 in 2012, but only on the parents who claim some of those children as dependent in 1998. In addition, we have access to detailed tabulations of the income distribution of all taxpayers in the same age group as the linked parents in 1998, i.e. all potential parents of those children.

Our first step is to reweigh the sample of matched fathers in our dataset in order to realign their income distribution to the unconditional one. Second, we merge the sample of non-dependent children with the sample of dependents. We then estimate a Probit model of dependency status for children in 1998 as a function of their observable characteristics (gender, age, province of birth, income percentile, major income component, and interactions). The R-squared of this regression is 0.21. Finally, we use these probabilities to reweigh the sample of linked parent-child pairs. If we let X_i be the set of observables for child i and d_i its dependency status, that observation is reweighed by dividing it by $\Pr(d_i = 1|X_i)$ as predicted by the regression. In essence, this correction is a propensity-score reweighing based on selection on observables.

With this correction, our estimates of RRS, AUM and Q1Q5 are 0.262, 0.435 and 0.098, respectively (Table 12). Therefore, the sample selection induced by the dependence status leads us to moderately understate the degree of rank persistence (0.246 instead of 0.262), but upward mobility measures are essentially unaffected.

5.3 Omission of certain taxpayers and tax evasion

5.3.1 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 representative of the population in terms of income levels and sources of income. Therefore, our rank ordering may be incorrect.

Here we propose a simple correction 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 comparable with those who filed form 770.⁴² 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 taxpayers (730+MU+770) to 730+MU taxpayers in that cell. Table 12 reports the results of this reweighing procedure and shows that our estimates of intergenerational mobility are robust to this correction.

5.3.2 Tax evasion

Italy is notoriously a country with high tax evasion. The size of the non-observed economy was estimated by the Italian National Statistical Institute to be as large as 12.1 percent of GDP in 2014 - the last year considered in our study. 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. Tax evasion rates differ significantly across earner categories. For dependent workers and retirees, taxes are withheld at the source. Thus, evasion is nearly impossible for these categories of earners, if we exclude illegal workers who are not in the sample.⁴³ Self-employment and rental income, instead, is 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

⁴²Specifically, 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 the 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.).

⁴³Tax evasion by illegal dependent workers accounted for 4.5% of total tax evasion in Italy between 2013 and 2015, according to official sources.

surveys is much closer to the truth since survey respondents are protected by anonymity due to privacy laws. These studies compute average after-tax income from survey and tax returns for groups of earners with the same type of income, income decile, and region. They confirm rates of tax evasion near zero for dependent workers.⁴⁴ 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 decline steeply with the level of income. For example, for the self-employed, tax evasion rises from 8% of reported gross income in the top decile to 70% in the lowest decile.

We use these sources to inflate these types of income in all our records for both children and parents. The corrected estimates of intergenerational mobility in Table 12 imply a sensible increase of the RRS and a smaller reduction in the upward indexes. Intuitively, 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.

In the Appendix, we describe two other corrections that leave our estimates barely unchanged: the omission of poor children, and missing capital income.

5.4 Taking stock: combining the sources of bias

As seen in this section and Section 4, some of the sources of bias shift our point estimates upward and others downward. Here, we consider all the major ones to assess the overall potential bias in our baseline calculations for relative and upward mobility.

Table 12 shows that combining the dependency bias, the bias coming from the omission of certain taxpayers, and the bias originating from tax evasion pushes up the RRS to 0.30. This compounded correction has a smaller impact on the index of AUM and the Q1Q5.

One could then further correct these estimates by the attenuation bias computed in our simulations of Section 5.1. Applying the same proportional factors we obtained for the baseline estimates, one would arrive at a RRS of 0.36, an index of AUM of 0.41 and a Q1Q5 of 0.074. These values should be considered a lower bound for our estimates of the level of mobility in Italy for a number of reasons. For example, as seen previously, by deflating income with local

⁴⁴Recall that our comparison of Section 3.4 is in fact based on wage income.

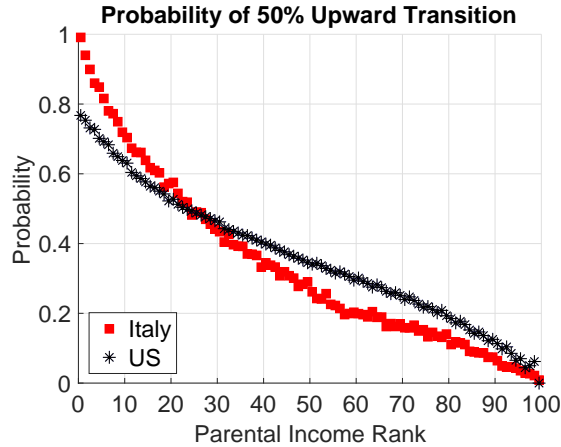


Figure 8: Probability that the income of the child is at least 50% higher than parental income. For the United States income is family income for both child and parents. For Italy, it is individual income of sons and their fathers.

price indexes, our estimates of mobility would indicate higher levels of mobility.

6 Cross-country comparison

In our core sample, the rank-rank slope (RRS) is 0.25 and the index of absolute upward mobility is 0.44. How does this estimated level of intergenerational mobility in Italy compare to that estimated in other countries? We limit our comparison to studies that used large administrative datasets like ours.

The estimate of the RRS for Denmark is 0.18 – corresponding to an AUM index of 0.46 – (Boserup 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.⁴⁵

As we argued in Section 2, these cross-country comparisons of positional mobility can be

⁴⁵ An important 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.

misleading if countries differ in their level of income inequality. The OECD Income Distribution Database reports Gini indexes of gross income around 0.30 for Denmark, Norway, and 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 low income US 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 the income of the child is 50 percent higher than parental income, after normalizing the parent and child distributions so that they have the same mean.⁴⁶

Figure 8 depicts this probability of upward transition for both the United States and Italy at each percentile of the distribution of parental income. The plot shows that this index of upward mobility is significantly higher for Italy among poor households. However, after the third decile or so, thus for the middle class as well, upward mobility becomes stronger for 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. Over the period 1998-2012 the number of provinces increased from 103 to 110. 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 includes 23 municipalities.

⁴⁶This normalization corrects both for aggregate growth between 19980 and 2012 and for the fact that parental income includes income from both spouses while child income is for the individual only.

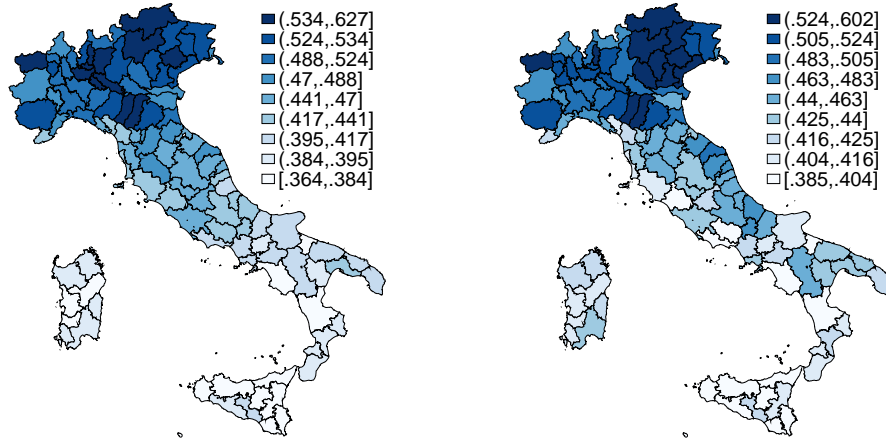


Figure 9: Heat map of Absolute Upward Mobility across provinces. Dark areas are more mobile. Left-panel (a): AUM computed on nominal income. Right-panel (b): AUM computed based on PPP-adjusted income (with regional price indexes)

In order to analyze the province-level variation in mobility measures, we assign each child to the province that her father indicated as the province of residence in his own 1998 tax return. Such province is interpreted as the area where the children grew up, not necessarily as the area where they live as adults.

We document mobility at the provincial level using the same definitions of parental and child income and the same (core) 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 4.1.⁴⁷

The extent of the difference in relative and upward mobility across provinces is summarized by Tables 13 and 14. These tables report mobility measure for the top-ten and bottom-ten provinces among the largest fifty provinces in Italy based on resident population in 2010. The full ranking of the 110 provinces is in the Appendix.

Figure 9 summarizes graphically geographic variation in AUM across the 110 provinces. In this heat map, darker colors correspond to more mobile areas. Two broad spatial patterns

⁴⁷We have used Index 9 for these figures, but our conclusions are robust to using other indexes. As discussed in that section, the different local price indexes we could use are all highly correlated. Figure A6 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).

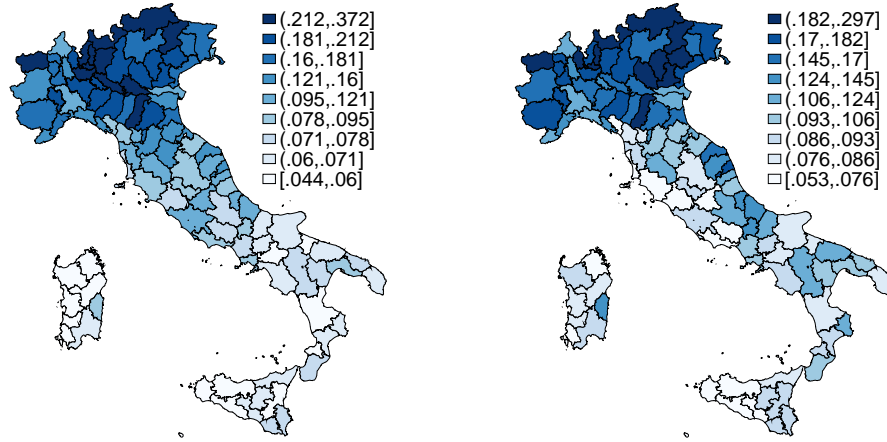


Figure 10: Heat map of Q1Q5 across provinces. Dark areas are more mobile. Left-panel (a): Q1Q5 computed on nominal income. Right-panel (b): Q1Q5 computed based on PPP-adjusted income (with regional price indexes)

emerge from this figure. First, there is substantial heterogeneity in upward mobility across provinces. The interquartile range of AUM across provinces is 0.12, almost twice 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 across the four macro areas (North-East, North-West, Center, South and Islands), accounts for 80% 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 Prato (Tuscany) at position 37/110. The lowest-ranked province for AUM not in the South and Islands macro-region is Latina (Lazio) at position 75/110.⁴⁸

Quantitatively these differences are meaningful. The province with the highest AUM is Bolzano (Trentino-Alto Adige), with a value of 0.63 and the one with the lowest AUM index is Ragusa (Sicily), with a value of 0.36. This expected rank differential corresponds to nearly 9,000 Euros of annual income in 2012, and hence, it translates into substantial gaps in children's lifetime incomes.

Panel (b) of Figure 9 reveals that, once adjusting for different cost-of-living levels across

⁴⁸Interestingly, Prato is the province with the largest Chinese community in Italy (15% of population) because of its vibrant textile sector. Latina, on the contrary, is a rural area with a predominantly agricultural economy.

regions, the predominance of the North-East macro area in terms of upward mobility is accentuated. However, now a few pockets of upward mobility emerge also in Southern Italy.

These differences in upward mobility are equally pronounced for the Q1-Q5, 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 10. The least mobile provinces have transition rates around 4-5% and the most mobile around 24-25%, i.e. larger by a factor of five. As evident from the heat maps, the correlation between AUM and Q1Q5 indicators is very strong (0.94).

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 12 percentiles of rank differential compared to someone born from parents at the top of the distribution. For children growing up in Bari (Puglia) —the one with the highest RRS among the largest provinces), it translates into a gap twice as big, i.e., 24 percentiles.

Figure 11 plots the full rank-rank relation in two of Italy's largest metropolitan areas, Milan and Bari. Milan (shown in blue) displays a rank-rank relationship that is both flatter and everywhere higher compared to Bari (shown in red). Through the lenses of a utilitarian planner, Milan dominates Bari: children who grow up in Milan 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 the same sharp increase in slope at the top of the distribution that emerges at the national level. Figures A7 and A8 in the Appendix display the rank-rank relationship for the top and bottom eight provinces.

We conclude that areas in Northern Italy (especially the regions in the 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 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 -0.78), 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. An important implication of this finding is that provinces with high (low) upward mobility also feature low (high) downward mobility.⁴⁹

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 and growth 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, when computing AUM using the national distribution, upward mobility in the South would be mechanically lower.⁵⁰ Moreover, in the absence of geographical migration among children, if the South grows less than the North, again, mechanically, upward mobility would be lower in the South.

In order to address this concern, we perform two exercises. In the first one, when we compute indexes of mobility of a province, we reweigh observations in each bin (e.g., 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. We find that the mobility estimates by province are barely affected. In our second exercise, instead of ranking parents and children with respect to their national income distributions, we rank them relative to the distribution of income in the region of residence of the father. Then, for each province, we estimate RRS, AUM and Q1Q5 with ranks computed from these regional distributions. Note that this approach implicitly assumes that each region is a separate entity from the others. Estimates of mobility computed this way are more affected, but the overall patterns are robust: the correlations between RRS, AUM and Q1Q5 across provinces when measuring mobility with respect to the national vs. regional income distributions are respectively 0.94, 0.43 and 0.52, and all highly significant.

⁴⁹For example, 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 -0.81.

⁵⁰As argued in Section 2, using the 25th percentile of the national parental distribution—which is what we do—already largely solves this problem.

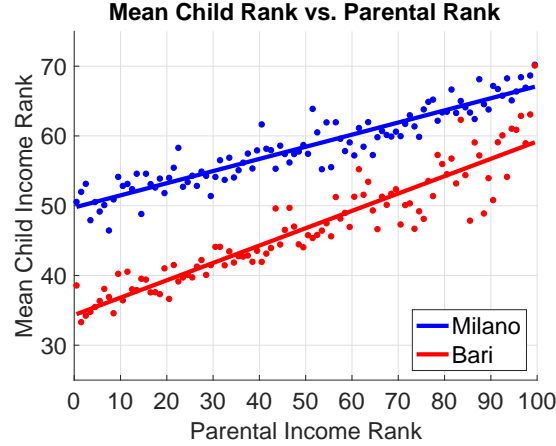


Figure 11: Child rank-parental rank relationship for children who grew up in Milano and Bari. The RRS are respectively, 0.18 and 0.24, and AUM indicators are 0.54 and 0.41.

7.1 Is there a Great Gatsby curve within Italy?

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

We report our results in Tables 15 and 16 for absolute and relative mobility, respectively. It is not surprising the relation exists: when inequality is higher, the rungs of the income ladder are further apart and it becomes more difficult to move. It is interesting to observe that, while the 90-10 income ratio is highly significant, the Gini coefficient is less so. The tables reveal that it is the Gini coefficient computed on the bottom 99% of the distribution that is strongly negatively related with mobility. Somewhat surprisingly, the income share of the top 1% is positively correlated with mobility. This result contrasts with the finding for the United States where the top income share is negatively correlated too (Table V in Chetty et al. (2014)).⁵¹ 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.

⁵¹The top income share instead is not significantly correlated with relative mobility in Chetty et al. (2014).

7.2 Comparison with Informational Content of Surname indexes

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 few years ago, 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 minimal data required for this methodology is a cross-section of individual records with information on income and on the surname of the individual. The data are then used to construct an indicator of the capacity of family names to capture the variance of the outcomes. 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 outcomes, such as ability and wealth, for example. The more these inherited characteristics matter for economic outcomes, the more information surnames contain on the values of outcomes (the higher is the ICS index), and the lower the degree of social mobility.

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.⁵² Here we have the opportunity, for the first time, to assess how well the ICS index correlates with more direct measures of intergenerational mobility based on rich administrative data. This validation exercise has never been performed before.

In Table 17 we report the cross-province correlation between the various ICS indexes calculated by Güell et al. (2018) and two of our measures of upward mobility, the AUM index and the Q1Q5 transition rates, also computed at the level of provinces. The Spearman rank correlation index (which takes values between -1 and 1) is negative, as expected, and highly statistically significant. However, Figure 12, which contains a scatterplot of standardized measures of ICS and AUM by province, shows that this statistical relationship is strong only for provinces with high values of the ICS index. The correlation between these two measures for areas with ICS

⁵²Güell 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 April 30th 2008. This act was supposed to be part of a general strategy to fight tax evasion through social stigma. The Italian Data Protection Authority quickly ordered the Ministry to take down the website, but at that point the data had already been downloaded and became publicly available.

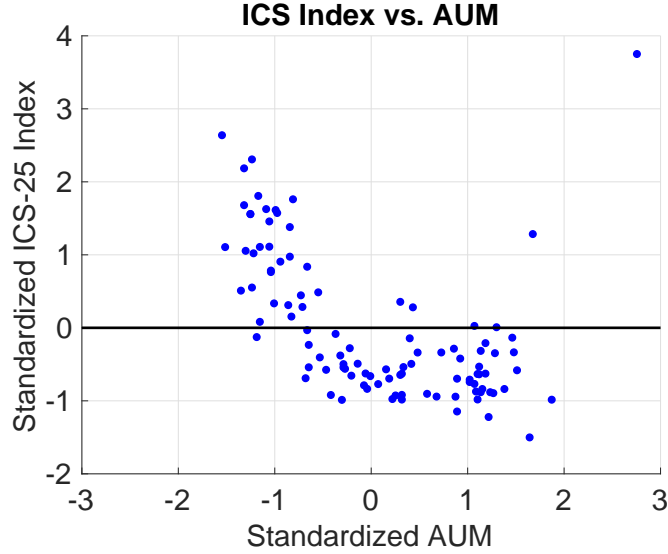


Figure 12: Relationship between the AUM index and the ICS-25 index across provinces.

below the mean is essentially zero.⁵³ In other words, the informational content of ICS about true mobility is very limited when the ICS is low: according to Figure 12, the ICS is a poor proxy of true upward mobility in at least half of Italian provinces. In light of this last finding, we conclude that researchers should be cautious in using this index.

8 What correlates with upward mobility?

The goal of this section is to make 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.⁵⁴

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)

⁵³The outlier in the top-right corner of the graph (with high ICS and AUM) is Bolzano, a largely German-speaking province.

⁵⁴Compared 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 very weakly correlated with true mobility for many provinces.

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, administrative efficiency, students per class), teachers' composition (e.g., teaching hours, teachers' age and gender, teachers on temporary contracts), students' test and examination scores.⁵⁵ A detailed description of our set of raw indicators is provided in Tables [A16](#) and [A17](#) in the Appendix.

We begin by documenting unconditional cross-province correlations between two measures of intergenerational mobility, AUM and Q1Q5, and these socio-economic markers. Tables [18](#), [19](#) and [20](#) summarize the results for a subset of these indicators. In general, most variables correlate with mobility indexes 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 female participation, youth unemployment and long-term unemployment) are 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. This 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. Crime statistics do not show a strong association (and when they do, the sign is generally unexpected), while the age structure and indexes of family instability

⁵⁵The 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.

do. The fraction of foreign-born and population inflows show a tight correlation with upward mobility. Finally, many of the measures of school quality —such as their structures, students per class, share of young teachers and test scores— are closely associated with upward mobility.⁵⁶ Adding macro-areas 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 captures 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).⁵⁷ We also include a measure of productivity. Finally, we include in the regression a measure of income inequality (the 90-10 ratio) that we found strongly significant in our analysis of Section 7.1.⁵⁸

Table 21 presents the results of the multivariate regression of Absolute Upward Mobility (AUM) on the factors we construct for each dimension of interest.⁵⁹ Overall, the included categories explain a very large portion of the variation of AUM (nearly 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

⁵⁶It 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. Indeed, our indicator of teachers' stability shows a strong positive correlation as well.

⁵⁷When constructing the disaggregated indexes 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 indexes of school quality by school level (pre-school, primary school, middle school, and high school).

⁵⁸Tables 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.

⁵⁹Variables are normalized to have mean 0 and standard deviation of 1, so regression coefficients can be interpreted as correlation coefficients.

quality. When disaggregating this factor into sub-categories, we find that the factor summarizing teachers' composition is the one which more strongly correlates with AUM. This factor loads positively on the average number of teaching hours available to each class, the share of male and young teachers, and teachers' tenure and stability in their position, and negatively on measures of teachers employed under temporary contracts. Also students' test scores is highly significant. When disaggregating the school quality index into school levels, the strongest effect is found for pre-school quality.

With respect to other correlates, family instability has a sizable correlation with AUM. We observe that the social capital index (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 relationship with AUM. This appears to be due to the strong correlation between the social capital and state of the labor market indexes. Excluding the measure of labor market conditions, social capital has a positive and statistically significant relationship with AUM. Finally, we find a positive correlation between value added per resident and upward mobility. The crime factor has no systematic relationship with upward mobility conditional on other observables.⁶⁰

8.1 An attempt at establishing causality

Our geographical analysis has documented a large geographical variation in the level of economic mobility. We have also uncovered that certain fixed characteristics of provinces (such as school quality, labor market conditions, etc.) can, statistically, explain much of this cross-province variation. However, so far we did not establish direct causality. For example, it is conceivable that certain provinces display higher social mobility not just because of their better intrinsic qualities, but also because they attract families that are more concerned with the economic outcomes of their children and invest more time and resources in their upbringing. How do we distinguish the role of the province from the unobserved attitudes of parents who reside in those provinces?

In the spirit of [Nakamura et al. \(2016\)](#), we make an attempt to establish causality in the role of provinces as determinants of social mobility by exploiting an exogenous shock that forced

⁶⁰Somewhat puzzling, life expectancy remains significant but not with the expected sign.

thousands of Italian families to migrate. The event is a violent earthquake (6.9 on the Richter scale) which took place in November 1980 and hit two regions of the South of Italy, Basilicata and Campania, causing nearly three thousand fatalities and enormous damages to hundreds of local villages. Over one hundred thousand families were forced to relocate because their houses were destroyed or severely damaged by the seism.

In our data, we can identify children who were born (all before 1980) in the three provinces that were hit hardest by the earthquake (Salerno, Potenza, and Avellino). Some of them have parents who still resided in those provinces in 1998 (stayers), while others have parents who, over that period, migrated to a different province (movers).⁶¹

We will work off two assumptions. First, these parents who migrated did so only because of the earthquake. Unfortunately, we cannot observe the precise year parents moved, so some of them may have migrated earlier or later, for different reasons. Support in favor of the hypothesis that many moves from those provinces over that period must have been induced by this unique event comes from comparing the migration rate of parents with children born in these three seism-hit provinces to the migration rate of parents with children born in all other provinces of the same two regions (Campania and Basilicata). We find that in the first group the migration rate is 0.074 and in the second group it is 0.028, and the difference is statistically significant at any level. The second assumption is that the selection between movers and stayers, within parental income rank, was random.

We compute that the expected rank of movers is 0.45. As a benchmark of comparison, we compute the expected rank of stayers conditional on the joint distribution of province of origin and rank of movers. This number captures the counterfactual expected child rank that would have realized if the parents of movers did not migrate. This counterfactual expected rank is 0.39. The difference between these two numbers should be interpreted as the differential effect of the provinces of destination, controlling for selection in the population of residents.⁶²

Since the expected rank in a province/percentile pair is the sum of the province effect and the selection effect coming from the composition of the resident population, intuitively, we should find that the expected rank of residents of the destination provinces, conditional on the

⁶¹The provinces with the largest share of movers coming from the ones hit by the earthquake are still in the South: Napoli, Benevento, Matera, Caserta and Cosenza.

⁶²The 99% confidence interval for the difference, obtained through bootstrapping with 100 repetitions, is [0.04, 0.08].

distribution of parental rank of movers, is certainly larger than 0.39 and larger than 0.45 only if this selection effect is strong. We estimate it to be 0.54.⁶³ Taken at face value, this result would suggest that around half of the province effect is due to selection of the resident population and the other half is the result of local socio-economic conditions.

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 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 bias due to the short panel dimension, misreporting of income due to tax evasion, and sample selection because of our ability to link generations only when children are claimed as dependent 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 concluded that our main results 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 mobility at the national level compared to many previous studies that tend to represent 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, they reveal 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,

⁶³The 99% confidence interval for the difference between the 0.54 and 0.45 point estimates, obtained through bootstrapping with 100 repetitions, is [0.08, 0.12].

but less everywhere else.

When we exploited this geographical variation to correlate a battery of socio-economic indicators with our measures of upward mobility, we uncovered that the quality of pre-school, children's test scores, indicators of family instability, and local labor market conditions have the strongest association to intergenerational mobility. An empirical exercise that exploits plausibly exogenous movers offers suggestive evidence that there is a causal effect of the province on upward mobility over and above the unobserved attributes of its resident population.

Another important results 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 comparisons as already done recently, for example, by [Alesina et al. \(2018\)](#) in their comparative study on 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. Atkison, 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.
- GUIO, 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.
- 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 Cyclicalities 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 Taxpayer 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.
- NAKAMURA, E., J. SIGURDSSON, AND J. STEINSSON (2016): "The Gift of Moving: Intergenerational Consequences of a Mobility Shock," Discussion paper, National Bureau of Economic Research.
- 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. (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	647,662	51
Mean mother's age	453,974	48
% Married father	640,734	92.7%
% Father residing in North	647,662	37.1%
% Father residing in Center	647,662	22.3%
% Father residing in South	647,662	40.6%
% Father residing in same region as were born	647,662	80.6%
% Father residing in same province as were born	647,662	72.0%
% Father born abroad	647,662	1.7%
% Families top-earner = father	647,662	87.5%
% Families with two positive incomes	647,662	68.3%
Corr. father-mother income (both positive)	453,974	0.2234
Median total parental income	647,662	27,032
Median total father's income	647,662	20,665
Median total mother's income	453,974	10,031
Children in 2012		
Mean age	647,662	37
% Married	647,662	46.4%
% Female	647,662	50.0%
% Residing in North	647,662	42.2%
% Residing in Center	647,662	24.1%
% Residing in South	647,662	33.7%
% Born abroad	647,662	1.8%
Median total income	647,662	19,174
Median total income males	323,856	21,445
Median total income females	323,806	16,894
% Individuals whose major income component is:		
Wage	647,662	71.9%
Entrepreneurship	647,662	12.1%
Other self-employment	647,662	9.2%
Capital	647,662	5.6%

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

Net Wage	Mean	SD	p99	p95	p90	p75	p50	p25	p10	p5	p1
SHIW	12,649	9,117	42,000	26,000	22,400	18,500	14,300	5,000	0	0	0
Tax Returns	13,691	10,229	44,023	30,824	25,721	20,046	14,331	3,100	1,650	0	0

Table 2: Comparison between after-tax wage income distributions in SHIW (household survey) and in the administrative tax return data, for individuals aged 35-40 in 2012 (our sample of children). Income is expressed in Euros.

	Core	Male	Female	Married Father	Div/Sep Father	2 Earners
Dependent Variable: Child Income Rank						
Parental income rank	0.246 (0.0012)	0.267 (0.0017)	0.233 (0.0016)	0.250 (0.0013)	0.182 (0.0061)	0.242 (0.0015)
Constant	0.377 (0.0007)	0.421 (0.0010)	0.330 (0.0010)	0.376 (0.0007)	0.403 (0.0038)	0.382 (0.0009)
AUM	0.438	0.487	0.388	0.438	0.449	0.442
Q1Q5	0.099	0.132	0.063	0.098	0.127	0.109
TMR	3.707	3.423	3.901	3.650	4.971	3.908
N	647,662	323,856	323,806	593,654	24,877	482,385

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	0.280 (0.0047)	0.268 (0.0017)	0.228 (0.0030)	0.209 (0.0065)	0.216 (0.0077)
Constant	0.359 (0.0022)	0.358 (0.0010)	0.382 (0.0013)	0.423 (0.0049)	0.390 (0.0023)
AUM	0.429	0.425	0.439	0.476	0.444
Q1Q5	0.055	0.075	0.100	0.158	0.118
TMR	3.060	2.889	5.379	6.043	3.873
N	54,361	375,529	102,461	21,494	21,797

Table 4: National indicators of intergenerational mobility for the core sample and various subgroups. 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 self-employment income, as opposed to 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	0.250 (0.0013)	0.220 (0.0035)	0.211 (0.0030)	0.159 (0.0034)	0.231 (0.0095)
Constant	0.376 (0.0007)	0.385 (0.0020)	0.413 (0.0019)	0.496 (0.0020)	0.396 (0.0056)
AUM	0.438	0.440	0.465	0.536	0.453
Q1Q5	0.097	0.113	0.132	0.169	0.129
TMR	3.523	5.324	4.042	1.693	1.939
N	557,324	90,338	114,790	78,206	10,789

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 2012 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	Mean Child Rank		95% CI for the Difference
	First Born	Later Born	
100	60.2	58.6	[0.3 , 2.9]
80	52.9	51.4	[0.1 , 2.8]
60	48.7	47.4	[0.0 , 2.6]
40	44.5	44.3	[-1.1, 1.5]
20	39.4	38.5	[-0.4, 2.0]

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

Parental Quintile	Child Quintile				
	1st	2nd	3rd	4th	5th
1st	28.66	25.40	20.66	15.36	9.92
2nd	22.02	22.73	22.37	19.33	13.56
3rd	18.81	20.20	21.51	21.82	17.66
4th	16.23	17.32	20.10	23.08	23.27
5th	14.27	14.34	15.38	20.41	35.60

Table 7: National quintile transition matrix (%)

Parental Decile	Child Decile									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
1st	16.13	15.12	13.66	11.92	10.60	9.33	7.51	6.70	5.26	3.77
2nd	13.04	13.02	13.09	12.13	11.20	10.18	8.72	7.80	6.21	4.60
3rd	11.45	11.72	12.00	11.56	11.68	10.90	9.57	8.61	7.25	5.26
4th	10.42	10.46	10.87	11.03	11.05	11.10	10.55	9.92	8.39	6.22
5th	9.80	10.03	10.35	10.56	10.78	10.81	10.82	10.56	9.29	7.01
6th	8.71	9.08	9.51	9.99	10.55	10.87	11.12	11.15	10.33	8.69
7th	8.03	8.44	8.62	9.33	9.90	10.88	11.42	11.58	11.54	10.26
8th	8.11	7.89	8.03	8.67	9.22	10.19	11.43	11.72	12.54	12.20
9th	7.55	7.38	7.51	7.97	8.29	8.81	10.51	11.74	14.08	16.17
10th	6.77	6.85	6.36	6.83	6.74	6.92	8.36	10.21	15.12	25.83

Table 8: National decile transition matrix (%)

	Real 1	Real 2	Real 6	Real 9	Real "Moretti"
Dependent Variable: Child Income Rank					
Parental Income Rank	0.239 (0.0012)	0.222 (0.0012)	0.215 (0.0012)	0.215 (0.0012)	0.225 (0.0012)
Constant	0.381 (0.0007)	0.389 (0.0007)	0.393 (0.0007)	0.392 (0.0007)	0.388 (0.0007)
AUM	0.440	0.445	0.446	0.446	0.444
Q1Q5	0.102	0.109	0.111	0.111	0.108
TMR	3.811	4.389	4.550	4.562	4.308
N	647,482	647,482	647,482	647,482	647,482

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 \(2013\)](#).

	$\sigma_{\kappa_0}^j$	ρ^j	σ_{η}^j	σ_{ϵ}^j
Fathers	0.320	0.998	0.107	0.123
Sons	0.471	0.989	0.164	0.294

Table 10: Parameter estimates of persistent-transitory process for log residual earnings used to assess attenuation bias.

	RRS	AUM	Q1Q5
3-year average	0.248	0.438	0.108
Lifetime income	0.297	0.426	0.094

Table 11: Average mobility statistics using 3-year average income or lifetime income based on simulations.

	Core	Dependency bias	With reweighting for missing 770	With adjustment for tax evasion	Combined
Dependent Variable: Child Income Rank					
Parental inc. rank	0.246 (0.0012)	0.262 (0.0005)	0.239 (0.0009)	0.286 (0.0012)	0.300 (0.0004)
Constant	0.377 (0.0007)	0.369 (0.0003)	0.385 (0.0005)	0.357 (0.0007)	0.346 (0.0002)
AUM	0.438	0.435	0.444	0.429	0.421
Q1Q5	0.099	0.098	0.102	0.090	0.085
N	647,662	3,226,186	1,098,452	647,662	5,892,577

Table 12: National indicators of intergenerational mobility. Correction for dependency bias, missing 770 forms, and tax evasion. Standard errors in parentheses. See the main text for details.

Province Name	Population in 2010	AUM	Q1Q5	RRS
Bolzano	507,657	0.627	0.372	0.098
Trento	529,457	0.561	0.161	0.124
Bergamo	1,098,740	0.549	0.215	0.171
Reggio-Emilia	530,343	0.548	0.177	0.153
Monza-Brianza	849,636	0.546	0.239	0.172
Modena	700,913	0.543	0.242	0.153
Milano	3,156,694	0.538	0.222	0.182
Treviso	888,249	0.536	0.210	0.128
Udine	541,522	0.534	0.176	0.135
Brescia	1,256,025	0.531	0.202	0.161

Table 13: Top 10 provinces by Absolute Upward Mobility.

Province Name	Population in 2010	AUM	Q1Q5	RRS
Cagliari	563,180	0.395	0.062	0.193
Reggio-Calabria	566,977	0.388	0.078	0.214
Agrigento	454,002	0.387	0.047	0.207
Catania	1,090,101	0.383	0.060	0.219
Trapani	436,624	0.382	0.053	0.239
Siracusa	404,271	0.382	0.073	0.204
Salerno	1,109,705	0.379	0.065	0.214
Palermo	1,249,577	0.378	0.056	0.215
Messina	653,737	0.378	0.062	0.227
Cosenza	734,656	0.366	0.059	0.230

Table 14: Bottom 10 provinces by Absolute Upward Mobility.

	AUM	AUM	AUM	AUM	AUM
Gini Coefficient	-0.199 (0.0943)				
Gini Bottom 99%		-0.416 (0.0875)			
Top 1% Income Share			0.448 (0.0860)		
90-10 Ratio				-0.565 (0.0794)	
SD log Income					-0.579 (0.0784)
R^2	0.040	0.173	0.201	0.319	0.335
N	110	110	110	110	110

Table 15: 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	0.192 (0.0944)				
Gini Bottom 99%		0.351 (0.0901)			
Top 1% Income Share			-0.314 (0.0914)		
90-10 Ratio				0.459 (0.0855)	
SD log Income					0.461 (0.0854)
R^2	0.037	0.123	0.099	0.210	0.213
N	110	110	110	110	110

Table 16: 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.

	ICS-15	ICS-20	ICS-25	ICS-30
AUM	-0.535 (0.0000)	-0.531 (0.0000)	-0.519 (0.0000)	-0.523 (0.0000)
Q1Q5	-0.429 (0.0000)	-0.423 (0.0000)	-0.411 (0.0000)	-0.414 (0.0000)

Table 17: Cross-province correlation of ICS indicators from Güell et al. (2018) with AUM and Q1Q5 estimated on our data.

	AUM Baseline	AUM Macroarea FE	Q1Q5 Baseline	Q1Q5 Macroarea FE
<i>Productivity</i>				
Value added per resident	0.691*** (0.0000)	0.192*** (0.0007)	0.624*** (0.0000)	0.168** (0.0178)
Manufacturing share of value added	0.659*** (0.0000)	0.222*** (0.0000)	0.601*** (0.0000)	0.193*** (0.0013)
Public works started	0.388*** (0.0000)	0.124*** (0.0074)	0.395*** (0.0000)	0.146** (0.0102)
Public works completed	0.261*** (0.0070)	0.101** (0.0207)	0.345*** (0.0003)	0.196*** (0.0002)
<i>Labor market</i>				
Unemployment rate	-0.812*** (0.0000)	-0.344*** (0.0000)	-0.723*** (0.0000)	-0.300*** (0.0024)
Youth unemployment rate (age 15-24)	-0.852*** (0.0000)	-0.454*** (0.0000)	-0.767*** (0.0000)	-0.426*** (0.0000)
Long-term unemployment rate	-0.763*** (0.0000)	-0.226*** (0.0030)	-0.686*** (0.0000)	-0.230** (0.0144)
Employment rate (college degree or higher)	0.769*** (0.0000)	0.223*** (0.0007)	0.715*** (0.0000)	0.194** (0.0184)
Labor force participation	0.771*** (0.0000)	0.230*** (0.0014)	0.710*** (0.0000)	0.266*** (0.0029)
Female labor force participation	0.745*** (0.0000)	0.194*** (0.0063)	0.669*** (0.0000)	0.192** (0.0293)
<i>Educational attainment</i>				
Share of illiterates	-0.426*** (0.0000)	0.006 (0.9006)	-0.373*** (0.0000)	0.003 (0.9494)
Education level achieved	0.519*** (0.0000)	0.102** (0.0265)	0.460*** (0.0000)	0.086* (0.0949)
School dropouts	-0.254*** (0.0087)	-0.006 (0.9071)	-0.203** (0.0372)	-0.005 (0.9361)
<i>Social capital</i>				
Blood bags collected per resident (GSZ 2004)	0.456*** (0.0000)	0.039 (0.4300)	0.454*** (0.0000)	0.085 (0.1586)
Self-efficacy measure (GSZ 2016)	0.657*** (0.0000)	0.209*** (0.0001)	0.700*** (0.0000)	0.358*** (0.0000)
Voter turnout, House of Representatives	0.689*** (0.0000)	0.181*** (0.0073)	0.583*** (0.0000)	0.128 (0.1288)
Voter turnout, European Parliament	0.425*** (0.0000)	0.039 (0.4587)	0.326*** (0.0007)	-0.013 (0.8482)
Recycling to total waste ratio	0.782*** (0.0000)	0.253*** (0.0003)	0.700*** (0.0000)	0.180** (0.0380)

Table 18: 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	0.264*** (0.0062)	-0.119** (0.0107)	0.282*** (0.0035)	-0.067 (0.2523)
Violent crimes	-0.119 (0.2257)	-0.054 (0.2160)	-0.090 (0.3581)	-0.043 (0.4228)
Distraints	-0.250*** (0.0097)	-0.056 (0.2098)	-0.264*** (0.0064)	-0.070 (0.2050)
Scam offenses	0.492*** (0.0000)	0.111** (0.0215)	0.480*** (0.0000)	0.122** (0.0398)
<i>Life expectancy</i>				
Female life expectancy at birth	0.431*** (0.0000)	0.121** (0.0176)	0.362*** (0.0001)	0.113* (0.0736)
Male life expectancy at birth	-0.167* (0.0886)	0.045 (0.3592)	-0.189* (0.0528)	0.054 (0.3753)
Old-age index (residents above age 65 / below age 15)	0.427*** (0.0000)	-0.193*** (0.0008)	0.321*** (0.0008)	-0.264*** (0.0002)
Number of suicides	0.431*** (0.0000)	-0.015 (0.7594)	0.435*** (0.0000)	0.035 (0.5730)
<i>Family instability</i>				
Divorce rate	0.623*** (0.0000)	-0.118* (0.0853)	0.570*** (0.0000)	-0.122 (0.1485)
Children in custody due to divorce	0.581*** (0.0000)	-0.087 (0.1701)	0.519*** (0.0000)	-0.109 (0.1616)
Separation rate	0.609*** (0.0000)	-0.016 (0.8139)	0.548*** (0.0000)	-0.018 (0.8292)
Children in custody due to separation	0.544*** (0.0000)	-0.052 (0.3943)	0.467*** (0.0000)	-0.094 (0.2087)
<i>Openness</i>				
Trade (exports+imports)	0.315*** (0.0009)	0.071 (0.1044)	0.323*** (0.0006)	0.093* (0.0833)
Net inter-province migration	0.667*** (0.0000)	0.127* (0.0748)	0.560*** (0.0000)	0.064 (0.4646)
Foreign-born residents	0.755*** (0.0000)	0.270*** (0.0002)	0.675*** (0.0000)	0.274*** (0.0024)
Inflow of graduates from other province or abroad	0.593*** (0.0000)	0.201*** (0.0001)	0.545*** (0.0000)	0.217*** (0.0005)

Table 19: 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	0.102 (0.2778)	0.009 (0.8148)	0.127 (0.1614)	0.034 (0.4397)
School assets	0.293*** (0.0015)	0.020 (0.6290)	0.232*** (0.0096)	-0.022 (0.6393)
Availability of teaching materials and technologies	0.267*** (0.0040)	0.053 (0.1965)	0.255*** (0.0043)	0.049 (0.2755)
Quality of school buildings	0.397*** (0.0000)	0.044 (0.3098)	0.381*** (0.0000)	0.064 (0.1814)
Students using canteen, school bus and other services	0.651*** (0.0000)	0.142** (0.0121)	0.576*** (0.0000)	0.132** (0.0369)
Students per class	-0.681*** (0.0000)	-0.194*** (0.0006)	-0.563*** (0.0000)	-0.096 (0.1348)
Extended teaching-time availability	-0.050 (0.5940)	-0.005 (0.8971)	-0.016 (0.8580)	0.013 (0.7655)
Processing time of teachers' rankings	0.167* (0.0749)	-0.075* (0.0831)	0.156* (0.0848)	-0.058 (0.2271)
Teaching hours per class	-0.235** (0.0117)	-0.114*** (0.0048)	-0.165* (0.0669)	-0.081* (0.0743)
Teachers below 40 years old: pre-school	0.646*** (0.0000)	0.138*** (0.0088)	0.601*** (0.0000)	0.121** (0.0422)
Teachers below 40 years old: primary school	0.469*** (0.0000)	0.096** (0.0312)	0.445*** (0.0000)	0.107** (0.0322)
Teachers below 40 years old: middle school	0.376*** (0.0000)	0.114*** (0.0075)	0.399*** (0.0000)	0.138*** (0.0040)
Teachers below 40 years old: high school	0.101 (0.2839)	0.071 (0.1029)	0.131 (0.1492)	0.058 (0.2294)
Teachers under temporary contracts: pre-school	0.652*** (0.0000)	0.212*** (0.0000)	0.579*** (0.0000)	0.192*** (0.0005)
Teachers under temporary contracts: primary school	0.728*** (0.0000)	0.183*** (0.0028)	0.688*** (0.0000)	0.246*** (0.0002)
Teachers under temporary contracts: middle school	0.461*** (0.0000)	0.088* (0.0502)	0.450*** (0.0000)	0.143*** (0.0038)
Teachers under temporary contracts: high school	0.425*** (0.0000)	0.024 (0.5855)	0.400*** (0.0000)	0.043 (0.3898)
Tenure and stability of teachers' position	0.280*** (0.0025)	0.122*** (0.0031)	0.272*** (0.0022)	0.091** (0.0494)
Students repeating school year	0.152 (0.1068)	0.067* (0.0886)	0.100 (0.2729)	0.021 (0.6274)
INVALSI test scores for primary and middle school	0.487*** (0.0000)	0.061 (0.1862)	0.424*** (0.0000)	0.034 (0.5064)
INVALSI test scores for high school	0.551*** (0.0000)	0.124*** (0.0081)	0.473*** (0.0000)	0.083 (0.1138)

Table 20: 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	0.092* (0.068)	0.133*** (0.009)	0.103** (0.047)	0.113*** (0.010)	0.125*** (0.004)	0.118*** (0.009)
Crime	0.013 (0.734)	0.009 (0.855)	0.033 (0.393)	0.015 (0.641)	-0.041 (0.314)	0.030 (0.374)
Educational attainment	0.001 (0.991)	0.118** (0.035)	0.020 (0.691)	-0.009 (0.808)	0.070 (0.148)	-0.006 (0.892)
Family instability	-0.077 (0.183)	-0.055 (0.332)	-0.053 (0.376)	-0.170*** (0.002)	-0.185*** (0.001)	-0.159*** (0.005)
Strong labor market	0.585*** (0.000)	0.482*** (0.000)	0.510*** (0.000)	0.386*** (0.000)	0.274*** (0.006)	0.353*** (0.000)
Life expectancy	-0.214*** (0.000)	-0.180*** (0.000)	-0.165*** (0.000)	-0.098** (0.018)	-0.067 (0.114)	-0.078* (0.068)
Economic openness	-0.021 (0.700)	-0.011 (0.847)	0.026 (0.649)	0.022 (0.636)	-0.017 (0.723)	0.043 (0.373)
Social capital	0.038 (0.582)	0.072 (0.320)	0.058 (0.419)	0.026 (0.660)	-0.002 (0.976)	0.047 (0.457)
90-10 ratio	-0.071 (0.148)	-0.086 (0.101)	-0.103** (0.050)	-0.109** (0.015)	-0.089* (0.052)	-0.120** (0.010)
School quality	0.312*** (0.000)			0.168** (0.011)		
Pre-school quality		0.288*** (0.002)			0.225*** (0.005)	
Primary school quality		0.067 (0.252)			0.091* (0.075)	
Middle school quality		0.163*** (0.001)			0.023 (0.612)	
High school quality		-0.017 (0.755)			-0.024 (0.605)	
School structures and resources			0.007 (0.874)			-0.007 (0.848)
School organization and services			0.135* (0.052)			0.098 (0.118)
Teachers' composition			0.143*** (0.004)			0.071 (0.111)
Students' grades and test scores			0.139*** (0.004)			0.088** (0.046)
Macroarea fixed effects				Yes	Yes	Yes
Observations	100	99	100	100	99	100
Adjusted R ²	0.87	0.88	0.87	0.91	0.91	0.91

Table 21: 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

Variable	N	Mean	SD	p99.9	p99	p95	p90	p75	p50	p25	p10	p5	p1
Parental income 1998	647,662	34,633	35,891	383,451	156,125	86,305	64,530	42,177	27,032	16,875	9,097	5,078	425
Top-earner's income 1998	647,662	27,520	30,967	337,215	132,042	69,614	50,697	30,625	21,362	14,811	7,951	4,548	400
Father's income 1998													
Total	647,662	26,473	30,859	331,376	130,102	68,646	49,749	29,942	20,665	13,581	6,242	2,488	37
Wage	647,662	16,551	21,708	173,191	88,743	52,642	37,804	24,562	14,697	0	0	0	0
Entrepreneurial	647,662	3,772	14,344	153,134	51,630	20,762	12,769	0	0	0	0	0	0
Professional	647,662	2,998	18,486	218,166	71,069	15,430	0	0	0	0	0	0	0
Capital	647,662	1,254	14,856	83,333	17,535	4,463	2,133	658	190	7	0	0	0
Pension	647,662	1,950	5,923	47,025	25,231	16,369	8,910	0	0	0	0	0	0
Mother's income 1998													
Total	453,974	11,642	13,717	129,383	55,549	29,290	23,299	18,312	10,031	982	186	78	0
Wage	453,974	7,543	10,046	65,302	37,134	23,200	20,665	16,112	0	0	0	0	0
Entrepreneurial	453,974	1,826	8,004	93,371	31,729	12,273	5,576	0	0	0	0	0	0
Professional	453,974	456	5,338	63,680	14,719	0	0	0	0	0	0	0	0
Capital	453,974	1,050	4,818	52,506	13,507	4,060	2,085	709	240	42	0	0	0
Pension	453,974	782	3,096	23,947	15,188	8,386	0	0	0	0	0	0	0
Child's income 2012													
Total	647,662	21,653	24,520	205,127	91,590	54,416	41,503	27,656	19,174	9,254	1,968	51	0
Wage	647,662	16,857	19,405	148,053	74,783	46,887	36,057	24,820	15,561	0	0	0	0
Entrepreneurial	647,662	2,138	12,538	109,171	38,157	15,243	5,923	0	0	0	0	0	-239
Professional	647,662	2,312	12,470	132,798	51,451	15,087	417	0	0	0	0	0	0
Capital	647,662	850	3,369	35,952	11,455	4,257	1,755	554	142	0	0	0	0
Pension	647,662	37	835	10,348	0	0	0	0	0	0	0	0	0

Table A1: Summary statistics on the income distribution and its components for parents and children. Income is in nominal terms and is measured 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	0.246 (0.0012)			0.267 (0.0017)		
Income rank	Top-earner income rank		0.233 (0.0012)			0.259 (0.0017)	
Income rank	Father income rank			0.220 (0.0012)			0.250 (0.0017)
Constant		0.377 (0.0007)	0.383 (0.0007)	0.390 (0.0007)	0.421 (0.0010)	0.424 (0.0010)	0.428 (0.0010)
AUM		0.438	0.442	0.445	0.487	0.488	0.490
Q1Q5		0.099	0.103	0.114	0.132	0.136	0.150
N		647,662	647,662	647,662	323,856	323,856	323,856
Child's Outcome	Parent's Income Definition	Female	Female	Female	Married Father	Married Father	Married Father
Income rank	Parental income rank	0.233 (0.0016)			0.250 (0.0013)		
Income rank	Top-earner income rank		0.211 (0.0016)			0.236 (0.0013)	
Income rank	Father income rank			0.192 (0.0016)			0.224 (0.0013)
Constant		0.330 (0.0010)	0.341 (0.0010)	0.351 (0.0010)	0.376 (0.0007)	0.383 (0.0007)	0.389 (0.0007)
AUM		0.388	0.394	0.399	0.438	0.442	0.445
Q1Q5		0.063	0.066	0.076	0.098	0.103	0.114
N		323,806	323,806	323,806	593,654	593,654	593,654
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	0.182 (0.0061)			0.242 (0.0015)		
Income rank	Top-earner income rank		0.177 (0.0061)			0.229 (0.0014)	
Income rank	Father income rank			0.166 (0.0062)			0.212 (0.0014)
Constant		0.403 (0.0038)	0.399 (0.0040)	0.406 (0.0040)	0.382 (0.0009)	0.396 (0.0009)	0.406 (0.0008)
AUM		0.449	0.444	0.447	0.442	0.453	0.459
Q1Q5		0.127	0.123	0.136	0.109	0.115	0.129
N		24,877	24,877	24,877	482,385	482,385	482,385

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	0.250 (0.0013)			0.220 (0.0035)		
Income rank	Top-earner income rank		0.237 (0.0013)			0.227 (0.0037)	
Income rank	Father income rank			0.245 (0.0013)			0.240 (0.0042)
Constant		0.376 (0.0007)	0.379 (0.0008)	0.368 (0.0008)	0.385 (0.0020)	0.404 (0.0018)	0.435 (0.0015)
AUM		0.438	0.438	0.430	0.440	0.460	0.494
Q1Q5		0.097	0.099	0.097	0.113	0.121	0.150
N		557,324	557,324	557,324	90,338	90,338	90,338
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	0.268 (0.0017)			0.228 (0.0030)		
Income rank	Top-earner income rank		0.258 (0.0018)			0.221 (0.0029)	
Income rank	Father income rank			0.236 (0.0017)			0.221 (0.0030)
Constant		0.358 (0.0010)	0.361 (0.0011)	0.373 (0.0011)	0.382 (0.0013)	0.385 (0.0013)	0.384 (0.0013)
AUM		0.425	0.425	0.432	0.439	0.441	0.439
Q1Q5		0.075	0.075	0.100	0.100	0.104	0.104
N		375,529	375,529	375,529	102,461	102,461	102,461
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	0.209 (0.0065)			0.216 (0.0077)		
Income rank	Top-earner income rank		0.202 (0.0065)			0.208 (0.0075)	
Income rank	Father income rank			0.195 (0.0066)			0.209 (0.0078)
Constant		0.423 (0.0049)	0.422 (0.0051)	0.427 (0.0051)	0.390 (0.0023)	0.392 (0.0023)	0.391 (0.0023)
AUM		0.476	0.472	0.476	0.444	0.444	0.444
Q1Q5		0.158	0.157	0.169	0.118	0.120	0.121
N		21,494	21,494	21,494	21,797	21,797	21,797

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 (0.0077)			0.172 (0.0029)		
Income rank	Top-earner income rank		0.206 (0.0076)			0.161 (0.0029)	
Income rank	Father income rank			0.183 (0.0071)			0.148 (0.0029)
Constant		0.389 (0.0058)	0.396 (0.0057)	0.416 (0.0052)	0.482 (0.0017)	0.488 (0.0017)	0.494 (0.0017)
AUM		0.443	0.447	0.462	0.525	0.528	0.531
Q1Q5		0.150	0.144	0.158	0.169	0.172	0.190
N		20,886	20,886	20,886	115,974	115,974	115,974

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 2012 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	0.257 (0.0017)	0.261 (0.0018)	0.195 (0.0090)	0.181 (0.0072)	0.163 (0.0080)	0.259 (0.0132)
Constant	0.371 (0.0010)	0.370 (0.0010)	0.389 (0.0059)	0.432 (0.0045)	0.499 (0.0046)	0.383 (0.0078)
AUM	0.436	0.435	0.438	0.477	0.539	0.448
Q1Q5	0.105	0.104	0.140	0.147	0.133	0.119
TMR	4.460	4.439	5.830	4.655	-2.344	1.205
N	323,856	297,154	12,018	24,916	16,440	5,415

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	0.281 (0.0026)	0.280 (0.0025)	0.263 (0.0041)	0.236 (0.0073)	0.224 (0.0104)
Constant	0.348 (0.0016)	0.349 (0.0015)	0.370 (0.0018)	0.410 (0.0053)	0.389 (0.0027)
AUM	0.418	0.419	0.436	0.469	0.445
Q1Q5	0.071	0.074	0.103	0.162	0.126
TMR	3.480	3.569	4.636	5.567	7.863
N	178,624	191,514	60,352	15,300	16,249

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	29.43	26.14	19.25	14.72	10.46
2nd	21.61	24.11	22.25	18.55	13.47
3rd	17.89	19.80	23.22	22.10	17.00
4th	16.61	17.11	20.67	23.51	22.09
5th	14.46	12.83	14.61	21.13	36.98

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	16.80	15.26	13.90	11.87	10.25	8.35	7.33	6.59	5.22	4.42
2nd	12.54	14.27	14.18	12.34	10.87	9.02	8.38	7.14	6.26	5.02
3rd	10.81	12.05	12.30	12.78	11.45	10.23	9.14	8.36	7.16	5.72
4th	9.81	10.56	11.11	12.03	11.78	11.04	10.35	9.25	7.80	6.27
5th	8.97	9.23	9.83	10.82	11.66	11.74	11.04	10.14	9.11	7.47
6th	8.91	8.67	9.12	9.83	11.11	11.94	11.94	11.07	9.61	7.80
7th	8.69	8.40	8.65	9.10	10.39	11.50	11.87	11.42	10.92	9.06
8th	8.20	7.93	8.14	8.34	9.04	10.41	11.70	12.03	12.66	11.54
9th	7.82	7.19	6.96	7.29	7.90	9.19	10.40	12.51	15.05	15.71
10th	7.46	6.45	5.81	5.59	5.55	6.59	7.84	11.51	16.20	27.00

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 and consider first fathers. For each father in the sample, we have at most three consecutive observations at dates $t, t-1, t-2$. The cross-sectional moments we use for estimation are:

$$var(y_t) = var(\kappa_t) + \sigma_\varepsilon$$

and

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

where

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

Identification can be achieved in different ways. For example:

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

For sons, the logic of the identification is the same with the caveat that observations are at dates $t, t-2, t-3$. 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 2012 because their income fell below this threshold level, we can gauge the effect of omitting poor children from our sample.

We have 884,808 parents with missing information on children. Applying the same age selection criteria as in the core sample, we are left with 603,426 records for parents. In the core sample, we retain only records where children are at least 35 years old in 1998 (29% of the total). However, we do not have information on children age for non-filers, thus we keep a random sample of 29% of our 603,426 records. For each of these records, we construct an artificial observation of a child with zero income, add them to the core sample and re-estimate the national mobility measures. The impact of this correction is very small: the RRS increases from 0.246 to 0.250, AUM decreases from 0.438 to 0.437 and the Q1Q5 decreases from 0.099 to 0.092.

Correction for missing capital income

We have also implemented a correction for missing capital income and concluded that this potential source of bias is insignificant. 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.



Figure A2: Map of the Italian provinces

Province	Population in 2010	Region	Macro area
Genova	882,718	Liguria	North-West
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	
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)

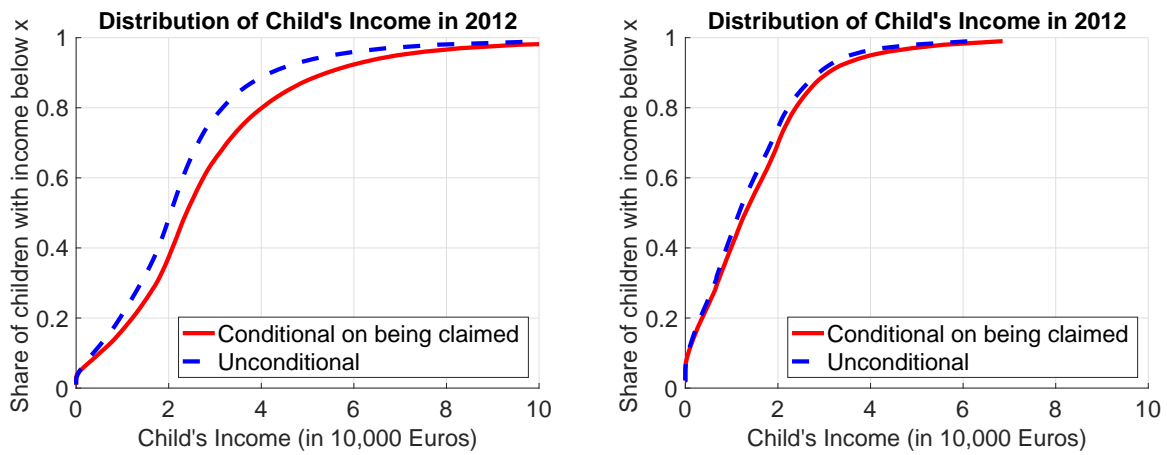


Figure A3: Unconditional and conditional distributions of child income in Lombardia (left panel) and Sicilia (right panel).

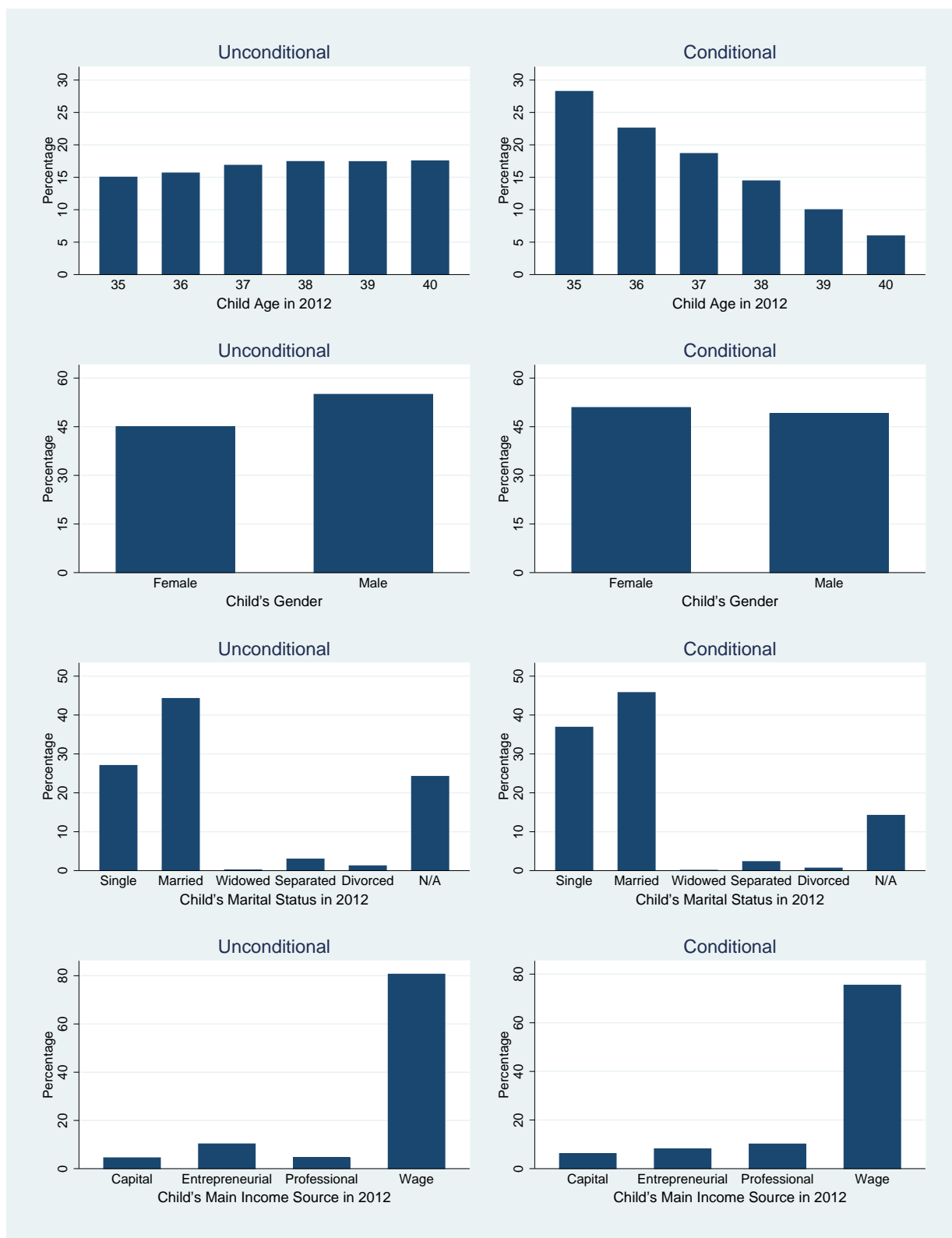


Figure A4: Comparison of the unconditional and conditional (on being claimed as dependent) distributions of various characteristics for children who reside in Lombardia in 2012

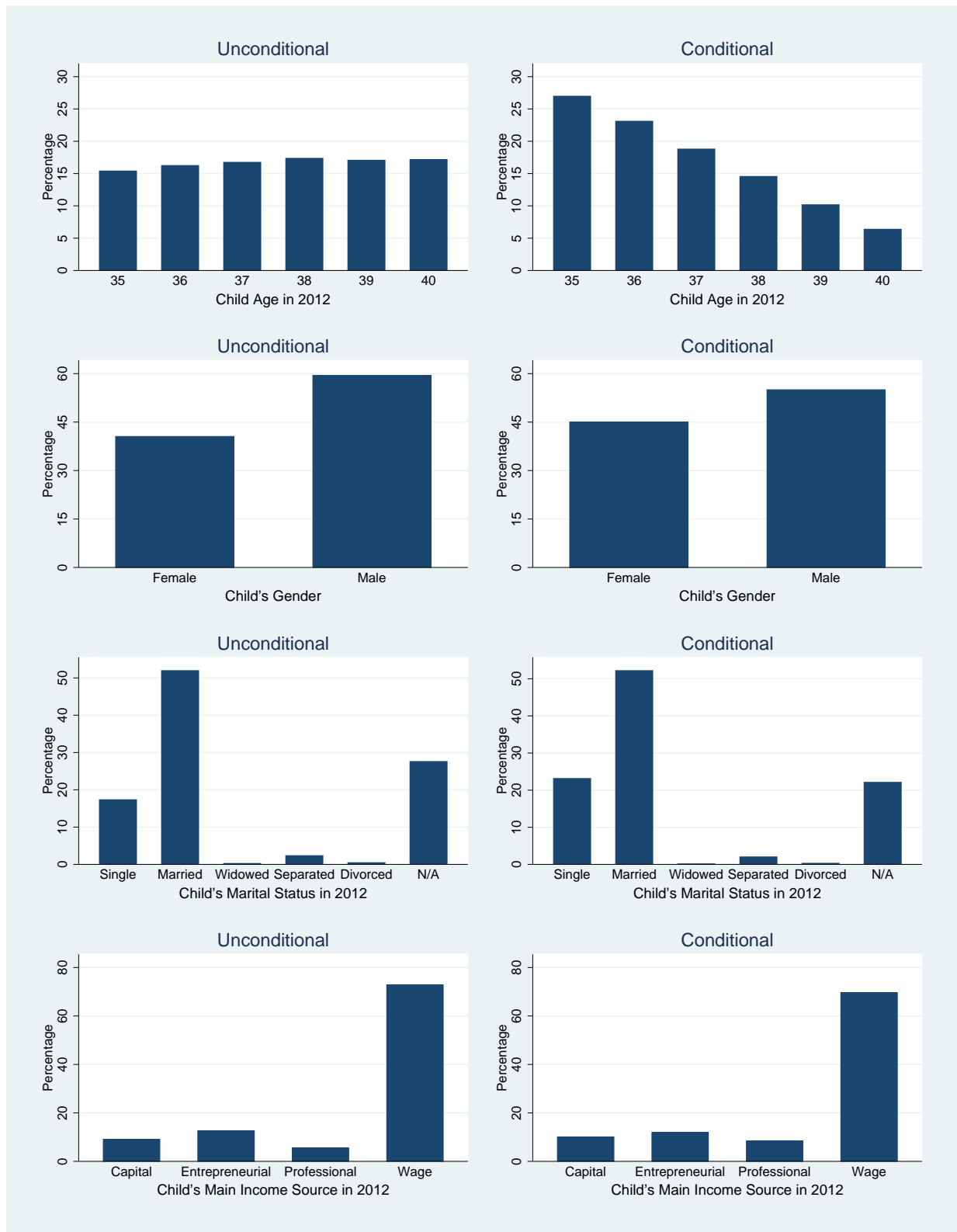


Figure A5: Comparison of the unconditional and conditional (on being claimed as dependent) distributions of various characteristics for children who reside in Sicilia in 2012

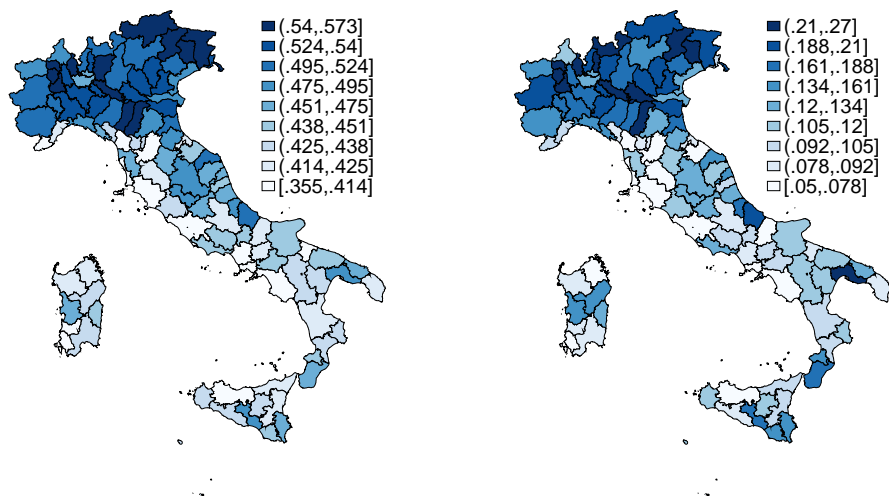


Figure A6: Heat map of Absolute Upward Mobility and Q1Q5 based on PPP-adjusted income (with province-level price indexes). Dark areas are more mobile. Left-panel (a): AUM. Right-panel (b): Q1Q5

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	N
1	Bolzano	507,657	0.627	0.372	0.098	2,080
2	Lecco	340,167	0.573	0.260	0.129	2,651
3	Trento	529,457	0.561	0.161	0.124	3,736
4	Aosta	128,230	0.559	0.267	0.101	974
5	Cremona	363,606	0.551	0.224	0.137	3,029
6	Bergamo	1,098,740	0.549	0.215	0.171	7,046
7	Reggio-Emilia	530,343	0.548	0.177	0.153	3,814
8	Monza-Brianza	849,636	0.546	0.239	0.172	8,269
9	Modena	700,913	0.543	0.242	0.153	5,658
10	Milano	3,156,694	0.538	0.222	0.182	30,541
11	Lodi	227,655	0.537	0.192	0.165	1,872
12	Treviso	888,249	0.536	0.210	0.128	7,239
13	Udine	541,522	0.534	0.176	0.135	4,987
14	Pordenone	315,323	0.533	0.181	0.155	3,062
15	Brescia	1,256,025	0.531	0.202	0.161	7,819
16	Vicenza	870,740	0.531	0.203	0.154	6,143
17	Belluno	213,474	0.529	0.247	0.134	1,688
18	Biella	185,768	0.528	0.170	0.137	1,464
19	Parma	442,120	0.528	0.183	0.191	3,723
20	Como	594,988	0.527	0.221	0.152	4,433
21	Padova	934,216	0.527	0.194	0.130	8,509
22	Cuneo	592,303	0.526	0.176	0.169	4,161
23	Venezia	863,133	0.526	0.162	0.134	8,498
24	Bologna	991,924	0.525	0.211	0.163	7,664
25	Pavia	548,307	0.524	0.192	0.188	4,669
26	Verona	920,158	0.524	0.197	0.156	7,444
27	Mantova	415,442	0.521	0.222	0.145	3,531
28	Varese	883,285	0.521	0.200	0.175	7,521
29	Piacenza	289,875	0.515	0.212	0.187	2,314
30	Sondrio	183,169	0.513	0.214	0.177	1,497
31	Vercelli	179,562	0.513	0.170	0.168	1,553
32	Novara	371,802	0.512	0.150	0.184	3,417
33	Trieste	236,556	0.511	0.162	0.138	2,201
34	Ravenna	392,458	0.503	0.181	0.157	3,632
35	Alessandria	440,613	0.500	0.121	0.192	3,834
36	Genova	882,718	0.494	0.160	0.184	9,827
37	Prato	249,775	0.488	0.163	0.127	2,207
38	Gorizia	142,407	0.485	0.113	0.178	1,248
39	Firenze	998,098	0.484	0.135	0.151	9,785
40	Savona	287,906	0.483	0.163	0.221	2,568

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

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	N
41	Verbania	163,247	0.479	0.117	0.174	1,271
42	Forli-Cesena	395,489	0.478	0.142	0.185	4,113
43	Rovigo	247,884	0.478	0.112	0.199	2,759
44	Torino	2,302,353	0.478	0.158	0.208	23,981
45	Ferrara	359,994	0.477	0.145	0.195	4,019
46	Siena	272,638	0.477	0.113	0.180	2,983
47	Asti	221,687	0.474	0.184	0.233	1,770
48	Ancona	481,028	0.472	0.132	0.174	5,595
49	La Spezia	223,516	0.470	0.159	0.172	2,641
50	Arezzo	349,651	0.468	0.135	0.177	3,573
51	Pesaro-Urbino	366,963	0.463	0.088	0.156	4,227
52	Fermo	177,914	0.458	0.156	0.108	1,722
53	Pisa	417,782	0.458	0.116	0.167	5,327
54	Macerata	325,362	0.456	0.100	0.126	3,252
55	Pistoia	293,061	0.455	0.103	0.194	2,798
56	Rimini	329,302	0.454	0.128	0.144	3,721
57	Roma	4,194,068	0.450	0.115	0.210	52,884
58	Terni	234,665	0.446	0.089	0.186	3,182
59	Rieti	160,467	0.445	0.117	0.191	2,155
60	Perugia	671,821	0.442	0.081	0.173	7,962
61	Imperia	222,648	0.441	0.134	0.175	1,688
62	Lucca	393,795	0.441	0.094	0.183	4,202
63	Ascoli Piceno	214,068	0.440	0.103	0.188	2,419
64	Chieti	397,123	0.439	0.100	0.181	5,759
65	Livorno	342,955	0.436	0.077	0.192	5,005
66	Massa-Carrara	203,901	0.433	0.095	0.173	2,760
67	Pescara	323,184	0.430	0.103	0.155	4,345
68	Grosseto	228,157	0.426	0.094	0.217	2,969
69	Isernia	88,694	0.425	0.095	0.208	1,254
70	L'Aquila	309,820	0.419	0.078	0.226	5,259
71	Viterbo	320,294	0.419	0.075	0.232	3,920
72	Frosinone	498,167	0.418	0.072	0.196	7,875
73	Taranto	580,028	0.418	0.080	0.216	10,284
74	Teramo	312,239	0.417	0.088	0.194	3,618
75	Latina	555,692	0.415	0.090	0.204	7,450
76	Brindisi	403,229	0.414	0.074	0.205	5,961
77	Caltanissetta	271,729	0.409	0.070	0.232	3,502
78	Potenza	383,791	0.408	0.077	0.193	5,361
79	Bari	1,258,706	0.407	0.071	0.242	17,223
80	Napoli	3,080,873	0.407	0.084	0.225	36,142

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

Upward Mob. Rank	Province Name	Population in 2010	AUM	Q1Q5	RRS	N
81	Campobasso	231,086	0.406	0.068	0.190	3,552
82	Caserta	916,467	0.401	0.073	0.226	9,538
83	Foggia	640,836	0.399	0.063	0.231	9,927
84	Lecce	815,597	0.398	0.066	0.217	11,578
85	Avellino	439,137	0.397	0.060	0.184	5,291
86	Cagliari	563,180	0.395	0.062	0.193	7,761
87	Matera	203,726	0.395	0.076	0.225	3,254
88	Benevento	287,874	0.394	0.071	0.195	3,744
89	Catanzaro	368,597	0.394	0.069	0.220	4,966
90	Crotone	174,605	0.392	0.064	0.240	2,038
91	Ogliastro	57,965	0.390	0.079	0.158	622
92	Olbia - Tempio	157,859	0.389	0.051	0.169	1,337
93	Reggio-Calabria	566,977	0.388	0.078	0.214	5,689
94	Vibo Valentia	166,560	0.388	0.055	0.207	2,002
95	Agrigento	454,002	0.387	0.047	0.207	4,876
96	Carbonia - Iglesias	129,840	0.387	0.044	0.257	2,207
97	Sassari	337,237	0.386	0.059	0.205	4,735
98	Enna	172,485	0.384	0.070	0.197	1,933
99	Catania	1,090,101	0.383	0.060	0.219	10,655
100	Oristano	166,244	0.383	0.050	0.174	1,943
101	Siracusa	404,271	0.382	0.073	0.204	5,687
102	Trapani	436,624	0.382	0.053	0.239	4,516
103	Barletta-Trani-Andria	392,863	0.380	0.060	0.241	5,194
104	Medio Campidano	102,409	0.379	0.051	0.180	1,456
105	Salerno	1,109,705	0.379	0.065	0.214	13,349
106	Messina	653,737	0.378	0.062	0.227	7,771
107	Palermo	1,249,577	0.378	0.056	0.215	13,622
108	Nuoro	160,677	0.376	0.049	0.194	1,744
109	Cosenza	734,656	0.366	0.059	0.230	9,836
110	Ragusa	318,549	0.364	0.075	0.214	3,457

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

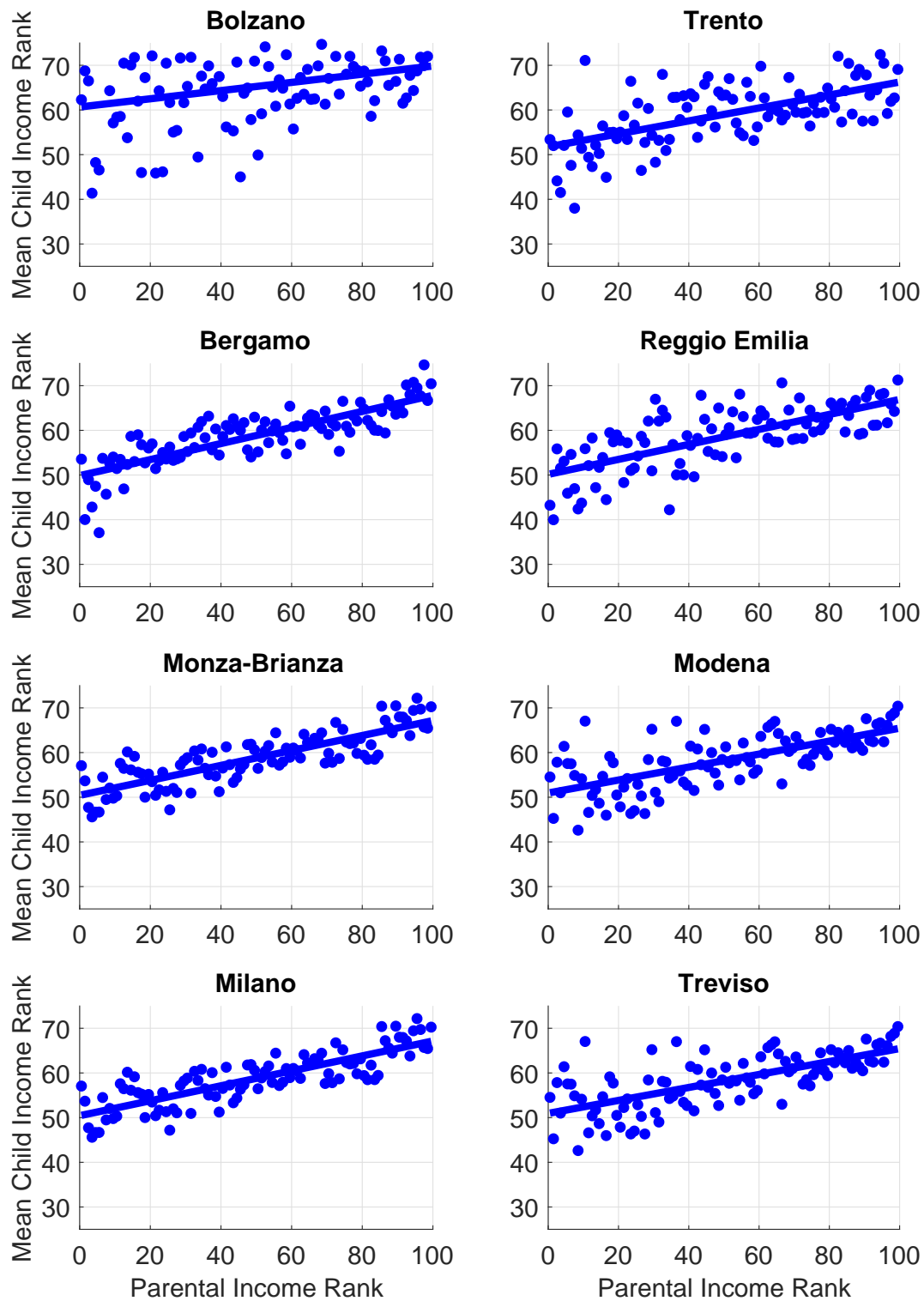


Figure A7: Rank-rank relationship for children who grew up in top provinces for AUM.

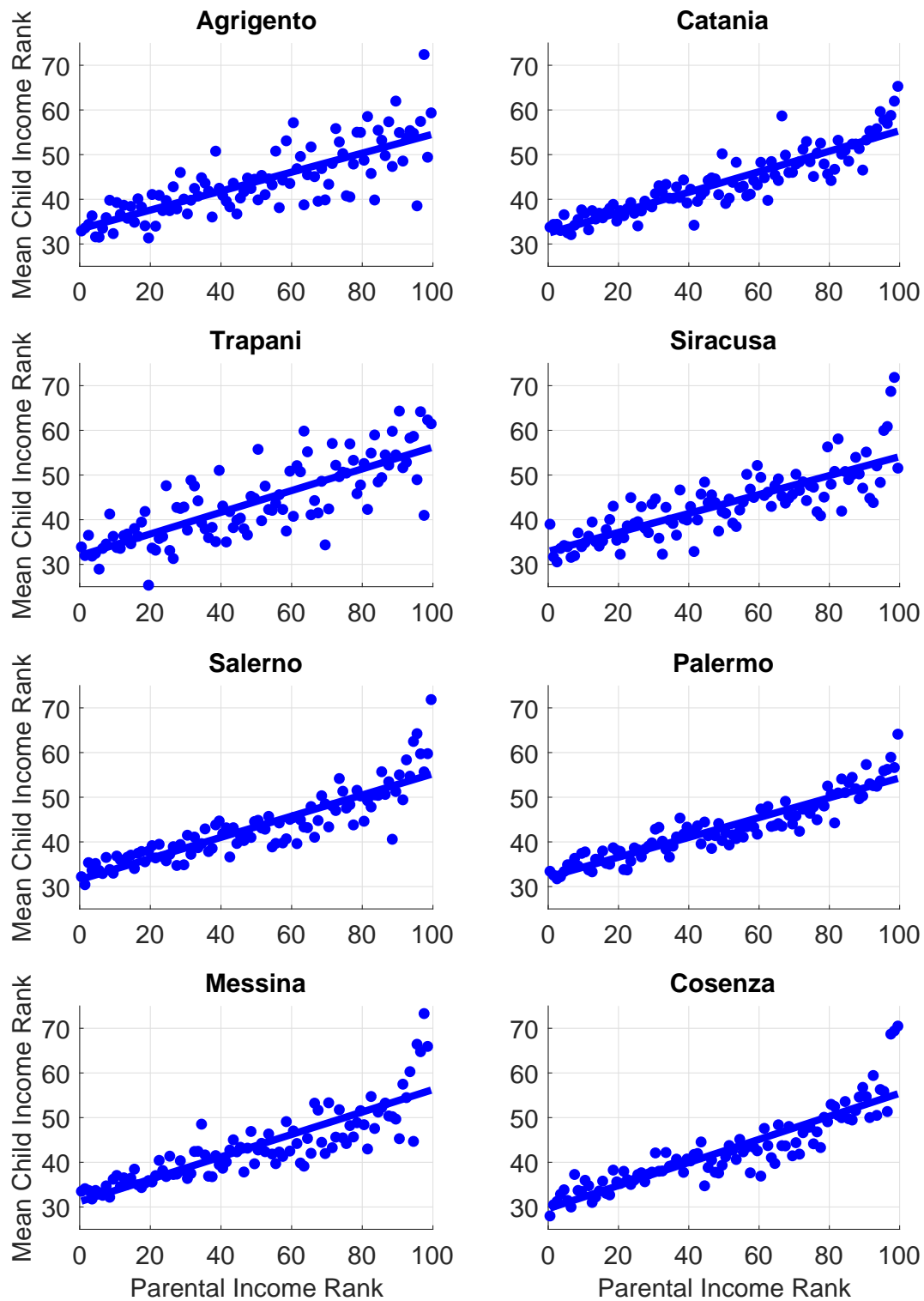


Figure A8: Rank-rank relationship for children who grew up in bottom provinces for AUM.

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
90-10 ratio	Our estimation
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)