

# Intergenerational Earnings Persistence in Italy along the Lifecycle\*

Francesco Bloise<sup>†</sup>, Michele Raitano<sup>‡</sup>

September 12, 2018

## Abstract

This study provides new estimates of the degree of intergenerational earnings persistence in Italy being able to observe for the first time actual fathers-sons pairs. Using high-quality administrative data merged with former Italian waves of the European Union Statistics on Income and Living Conditions, we exploit the longitudinal dimension of the dataset to observe the fathers over a 14 years period and the sons yearly over the first 6 years after they left education. Our preferred results show an estimated intergenerational elasticity of 0.392 and a rank-rank slope of 0.216 only 6 years after the sons left education. According to subsequent empirical tests, the selection rules adopted prevent our estimated elasticity to be strongly affected by the lifecycle bias unlike the estimated rank-rank slope. Therefore, after re-estimating the two measures around 10.5 years after a subsample of sons left education and correcting for the residual estimated lifecycle bias, we report high background-related earnings advantages finding the elasticity and rank-rank slope to be around 0.44 and 0.35 respectively.

***JEL Classification:*** J62, D31, D63. ***Keywords:*** intergenerational earnings persistence; earnings inequality; Italy.

---

\*This is a preliminary draft. Please do not cite or distribute without permission of the authors.

<sup>†</sup>University of Urbino Carlo Bo: francesco.bloise@uniurb.it

<sup>‡</sup>Sapienza University of Rome: michele.raitano@uniroma1.it

# 1 Introduction

Previous studies estimating the intergenerational earnings persistence in Italy show a high intergenerational earnings elasticity (IGE) compared to most of other developed countries (Piraino, 2007; Mocetti (2007); Barbieri et al., 2018) suggesting Italy to be a very low mobility society. However, due to the lack of data following two generations during their working careers, these studies estimate the IGE by imputing fathers' lifetime earnings through the two-sample two-stage least squares (TSTSLS) method, which is similar to the two-sample instrumental variable (TSIV) approach originally proposed by Angrist and Krueger (1992) and Arellano and Meghir (1992), and firstly used in the empirical literature on intergenerational mobility by Björklund and Jäntti (1997) to compare the IGE between Sweden and the US. Though the TSTSLS method is essential to estimate the IGE in countries for which it is not possible to directly link earnings of sons to those of their fathers, it may produce coefficients which are not perfectly comparable to the ones obtained from an OLS regression of the logarithm of sons' earnings on their actual fathers' given that, when parental income is imputed, the estimated IGE is assumed to be upward biased (Blanden, 2013).

To overcome this potential problem, this study provides brand new estimates of the IGE and rank-rank slopes by directly linking earnings of sons to those of their actual fathers. In particular, using administrative archives managed by the Italian Social Security Institute (INPS) merged with the 2004 to 2008 Italian waves of the EU-SILC, we take a sample of co-resident sons from 1 to 2 years after they left education together with information about their actual fathers. Since Italy is characterized by a high share of individuals co-residing with their parents before and right after leaving education, our sample is representative of more than 90% of Italian males taken at that specific point of the life cycle. Then, we exploit the large longitudinal dimension of INPS archives to measure averaged earnings of fathers when sons were aged 1 to 14 and earnings of sons from 1 to 6 years after they left education, until 2014.

According to this selection rules, we find important results showing that the intergenerational earnings persistence in Italy measured by the IGE is high and increases over the sons' career with an estimated coefficient 0.392 only 6 years after their educational achievements. On the contrary, positional persistence in the earnings distribution (i.e. persistence related to the copula of the distribution) measured by the rank-rank slope appears to be low in cross-county comparison and equal to 0.216.

Though these initial results may suggest that most of the intergenerational earnings

persistence derives from an increase in inequality occurred across the two generations rather than on positional persistence, they can also be related to the lifecycle bias affecting our estimated coefficients, as sons are observed at the beginning of their careers. This the reason why we adapt the error-in-variables model proposed by [Haider and Solon \(2006\)](#) to assess left-hand side measurement errors in Italy by evaluating the career-earnings profile of Italian workers through the so called “forward regression” of yearly earnings on a proxy of lifetime earnings. To do that we follow a representative sample of Italian workers that left education between 1979 and 1984 for 30 years and we show that while the estimated IGE at the 6th year after sons left education is probably not highly affected by the lifecycle bias, the estimated RR at the same point of the career is likely to be downward biased. Then, we re-estimate the IGE and rank-rank slope observing a subsample of sons that left education from 2002 to 2004 about 10.5 years after and we actually find higher estimated coefficients of 0.414 for the IGE and 0.265 for the rank-rank slope. Given that the comparison between the latter estimated coefficients and the ones obtained 6 years after the full sample of sons left education are highly consistent with the career-earnings profile estimated from the “forward regression”, we are able to correct our estimated measures for the residual lifecycle bias reporting the IGE and rank-rank slope to be around 0.44 and 0.35 in Italy.

According to these final results, we will show that the estimated IGE is not so far from the ones obtained by [Piraino \(2007\)](#), [Mocetti \(2007\)](#) and [Barbieri et al. \(2018\)](#) exploiting the TSTSLS method and two generations of pseudo-fathers and sons taken around 40 years old. Therefore our results suggest that previous estimates for Italy are likely not to be significantly upward biased, despite the use of imputed fathers’ earnings and the TSTSLS methodology. Moreover, both our final IGE and rank-rank slope suggest Italy to be a low mobility country likewise the US where previous evidence show the IGE to be around 0.4/0.5<sup>1</sup> ([Solon, 1992](#); [Zimmerman, 1992](#)) and the rank-rank slope between 0.32 and 0.50 ([Chetty et al., 2014a](#); [Bratberg et al., 2017](#)). This means that only a small fraction of the earnings persistence in Italy is related to changes in inequality occurred across generations whereas positional persistence in the earnings distribution seems to be very high in Italy.

The structure of the work is the following. Section 2 describes the empirical framework associated to the intergenerational transmission of inequality and previous evidence for Italy in cross-country comparison. Section 3 presents the data and the

---

<sup>1</sup>These results are obtained by averaging fathers’ earnings over a 4/5 years period. Alternative estimates averaging yearly earnings of fathers over many years report higher IGEs ([Mazumder, 2005](#))

selection of sons and fathers into the final samples. Section 4 discusses the results obtained in terms of both IGEs and rank-rank slopes following sons from 1 to 6 years after they left education. Section 4 presents the results of the “forward regression” of yearly earnings on lifetime earnings for a representative sample of Italian workers observed either at any given age or year of distance from when they left education. Section 6 shows final results obtained after observing a subsample of sons about 10.5 years after they left education and correcting for the residual estimated lifecycle bias. Section 7 concludes.

## 2 Empirical framework and previous literature

In the last few decades, many empirical studies carried out by both economists and social scientists analyze to what extent economic advantages are transmitted from one generation to the next <sup>2</sup>. In this literature, many indicators have been used to summarize the degree of intergenerational mobility or persistence. The most common indicator used by economists to measure degree of intergenerational earnings persistence is the IGE which can be estimated by regressing the logarithm of children’s lifetime earnings on the logarithm of parents’ as in the following equation:

$$\ln y_i^c = \alpha + \beta \ln y_i^p + \epsilon_i \tag{1}$$

According to this specific measure of persistence, a country is completely mobile when the estimated  $\beta$  equals 0 and the higher the estimated  $\beta$ , the higher the degree of intergenerational economic persistence.

Since background-related earnings advantages last over the whole working career of individuals, empirical studies analyzing the persistence of earnings aim to estimate the IGE by considering lifetime rather than yearly earnings of the two generations. However, despite the seemingly straightforward empirical framework previously described, lifetime earnings of children and their parents are usually not available because of the lack of panel data following two subsequent generations during their entire working career. This is the reason why researchers have to face several measurement issues when estimating the IGE causing an underestimation of the true intergenerational earnings persistence. For instance, earlier studies estimating the intergenerational earnings per-

---

<sup>2</sup>For more details about the approaches used by empirical researchers to estimate mobility across generations, see [Black and Devereux \(2011\)](#)

sistence in the US, report very low IGEs around 0.2 describing the US as a very mobile society according to the ideal of “American Dream” (Becker and Tomes, 1986; Behrman and Taubman, 1986). Subsequent works demonstrated that previous estimates were substantially downward biased due to the fact that they use yearly rather than lifetime earnings of fathers causing an attenuation a bias related to right-hand side measurement errors. For instance, Solon (1992) and Zimmerman (1992), based on larger and more representative data estimate the IGE using fathers’ earnings averaged over four or five years and obtaining values around 0.4/0.5, consistently higher than the ones estimated until the late 80s. Subsequently, Mazumder (2005) and Chen et al. (2017) in two studies respectively on the US and Canada show that even using 5-years averaged earnings may lead to an underestimation of the IGE since transitory shocks are likely to be extremely persistent. In any case, almost all of studies which estimate the  $\beta$  still use 4/5 years averaged father’s earnings as the baseline method to minimize the attenuation bias due to right-hand side measurement errors, since suitable data that follow parents over more than 5 years are not available in most countries.

Despite left-hand side measurement errors are often not assumed to cause any bias in an OLS regression, according to many empirical studies children’s earnings measured with error may cause the so-called lifecycle bias lowering the estimated  $\beta$  if children are taken when they are too young. This is because age-earnings profile is steeper for individuals with higher expected future income as the earnings growth rate of high-skilled workers is often higher than that of other individuals. Moreover, at a given age or for each given age range, individuals have different years of work experience according to their education, since tertiary graduated individuals usually enter the labor market several years after low-skilled individuals. For the two reasons just described, the earnings dispersion of young workers generically selected by age is likely to be consistently lower than the dispersion of their lifetime earnings. Therefore, since the estimated  $\beta$  from the equation 1 is equal to the following expression:

$$p \lim \hat{\beta} = \frac{Cov(y_i^c, y_i^p)}{Var(y_i^p)} = \rho_{cp} \frac{\sigma_c}{\sigma_p} \quad (2)$$

where  $\sigma_{cp}$  is the correlation between children’s lifetime earnings and parents’,  $\sigma_c$  is the standard deviation of children’s lifetime earnings and  $\sigma_p$  is the standard deviation of parents’, the estimated  $\beta$  is downward biased if we include  $y_i^{yc}$  rather than  $y_i^c$  in the right-hand side of equation 1, where  $y_i^{cy}$  are yearly earnings of young children, with  $\sigma_{yc} < \sigma_c$ .

Nevertheless, two empirical estimations of the lifecycle earnings variation made by Haider and Solon (2006) and Böhlmark and Lindquist (2006) for the US and Sweden respectively, suggest that this specific source of downward bias can be greatly reduced by selecting the second generation around median ages since the difference between earnings of individuals at mid-careers should be the closest to the difference between their lifetime earnings. Subsequently Nybom and Stuhler (2016), while confirming that the best way to minimize the lifecycle bias is to take individuals at median ages, warn that it can be very difficult to identify the specific age at which children’s lifetime earnings are perfectly approximated for all individuals as the age-earnings profile may be worker specific.

## 2.1 Previous evidence for Italy in a cross-country comparison

Previous empirical studies for Italy carried out by Mocetti (2007), Piraino (2007), and Barbieri et al. (2018) estimate the IGE by exploiting the TSTOLS methodology due to the impossibility to directly link children observed at their mid-careers to their actual parents. Following the TSTOLS method and according to microdata at their disposal, these scholars exploit retrospective time-invariant socioeconomic information of Italian fathers (e.g. educational achievements, occupational status, region of residence, sector of activity) recalled by their sons and an auxiliary sample of pseudo-fathers to get a prediction of lifetime fathers’ earnings and estimate the IGE<sup>3</sup>. Using this estimation methodology they obtain an IGE between 0.44 and 0.50 depending on the number of predictors used to impute fathers’ lifetime earnings, the number of years the two generations are observed, the data used and the income definition adopted (i.e. net or gross of taxes and deductions). According to these results, Italy is reported to have a higher IGE compared to most of other developed countries such as Norway, Sweden, Canada, France and Germany and close to the ones estimated for Spain, the US and the UK.

As it is clear from the table 1 , the TSTOLS has been often used to estimate IGEs in countries such as France, Spain and Italy where it is not possible to directly link earnings of sons to those of their fathers<sup>4</sup>. However this estimation methodology may produce coefficients which are not perfectly comparable to the ones obtained from an OLS

---

<sup>3</sup>As in most of empirical studies estimating the IGE, they focus on father-son pairs to avoid the selection bias related to the low labor force participation of women.

<sup>4</sup>The TSTOLS methodology has been widely used also to obtain estimates for many less developed countries.

Table 1: Intergenerational earnings elasticity: cross-country comparison

Country	Source	Empirical approach	IGE
US	Zimmerman (1992)	OLS	0.54
Italy	Barbieri et al. (2018)	TSTOLS	0.44-0.50
Italy	Mocetti (2007)	TSTOLS	0.50
Italy	Piraino (2007)	TSTOLS	0.44
UK	Gregg et al. (2017)	OLS	0.43
US	Solon (1992)	OLS	0.42
Spain	Cervini-Plá (2015)	TSTOLS	0.42
France	Lefranc and Trannoy (2005)	TSTOLS	0.40
Germany	Schnitzlein (2016)	OLS	0.39
Canada	Chen et al. (2017)	OLS	0.32
Sweden	Björklund and Chadwick (2003)	OLS	0.24
Norway	Bratberg et al. (2005)	OLS	0.13

**Notes:** To maximize the degree of comparability, almost all of reported IGEs which use the OLS estimator have been obtained by averaging fathers' earning using 4/5 yearly observations. Three exceptions are the studies by [Chen et al. \(2017\)](#), that take fathers with positive earnings in at least 10 years, [Schnitzlein \(2016\)](#) who observes fathers for ten years excluding those with less than 5 positive earnings observations and [Gregg et al. \(2017\)](#) that take two parental income observations. The two estimates by [Barbieri et al. \(2018\)](#) differ from the number of years used to measure sons' earnings (i.e. single year or 5-years average).

regression of the logarithm of sons' earnings on their actual fathers'. More specifically, the TSTOLS estimator produces estimated IGEs which are usually considered biased as the standard deviation of imputed fathers earnings is by construction lower than the standard deviation of actual fathers earnings and may not predict all components of earnings which are correlated across generations. ([Olivetti and Paserman, 2015](#); [Barbieri et al., 2018](#)). For instance [Blanden \(2013\)](#), suggests to re-scale all estimated IGEs obtained by using the TSTOLS method by a factor of 0.75 according to the results reported by [Björklund and Jäntti \(1997\)](#) that estimate the IGE in the US by using both the OLS and the TSIV estimators<sup>5</sup>.

<sup>5</sup>However it is well acknowledged by [Blanden \(2013\)](#) that it is a strong assumption to generalize the bias found by [Björklund and Jäntti \(1997\)](#) to other countries and studies which use different combinations of predictors to impute parental earnings.

## 2.2 An alternative measure of intergenerational earnings persistence: the rank-rank slope

Though the IGE is the most commonly estimated index to summarize the degree of intergenerational earnings persistence in a given country, another summary measure of intergenerational earnings persistence called rank-rank slope (RR) is nowadays very often used beside the IGE ever since [Dahl and DeLeire \(2008\)](#) introduced it and [Chetty et al.](#) made it famous in several studies focusing on the US ([Chetty et al., 2014a,b](#); [Chetty and Hendren, 2018a,b](#)). The RR can be simply obtained by estimating the following equation:

$$\text{rank}(y_i^c) = k + \varphi \text{rank}(y_i^p) + u_i \quad (3)$$

where  $\text{rank}(y_i^c)$  is the child's percentile in the lifetime earnings distribution of the second generation and  $\text{rank}(y_i^p)$  is the parent's percentile in the lifetime earnings distribution of the first generation.

One important advantage of this alternative measure of intergenerational persistence is that unlike the IGE which is affected by any change in inequality that occurred between the two generations considered, it depends only on positional persistence. More in detail, according to the equation 2, the estimated  $\beta$  depends on the correlation between parents' and children's earnings and on the ration between the two standard deviations. Thus, if the earnings dispersion increases (decreases) over time, the IGE will automatically becomes higher for a given correlation coefficient between parents' and children's earnings. On the contrary, as described in detail by [Chetty et al. \(2014a\)](#), the RR can depends only on the so called copula of the earnings distribution and it is a scale invariant measure independent from any change in inequality occurred across generations.

For the reason just described, in this work we want to present both estimates of the intergenerational earnings persistence to get a broader picture of the relationship between children's economic opportunities and parental background in Italy. It may be important to estimate this specific measure of mobility also because, apart for the recent study by [Barbieri et al. \(2018\)](#) which estimate the RR by imputing the percentile of the first generation within the TSTSLS approach, there is a lack of empirical works that provide comparable estimates of this specific measure of intergenerational earnings persistence for Italy or any other Mediterranean country. This lack of esti-

mates derives from the fact that, at the best of our knowledge, it was not possible to directly link earnings of children to those of their parents so far in Italy, Portugal or Spain so far because of the absence of suitable longitudinal data. Moreover, though RRs have been frequently estimated for the US, both at the national level and in its subareas [Chetty et al. \(2014a\)](#); [Chetty and Hendren \(2018b\)](#), there is a few evidence also for non Mediterranean countries. An important exception is the study by [Bratberg et al. \(2017\)](#) which compares estimated RRs obtained by considering children and their parents ranks in 4 developed countries including Germany, Norway, Sweden and the US. Their estimated RRs confirm the US to be the country with the highest level of intergenerational earnings persistence among the 4 analyzed with a coefficient of about 0.395. On the contrary their estimated RRs for Germany, Norway and Sweden are consistently lower and around 0.245, 0.223 and 0.215 respectively. [Chetty et al. \(2014a\)](#) found a slightly lower coefficient of 0.317 when considering only sons and their parents in the US.

Generally speaking, RRs are resulted to be less sensitive to the age at which children’s earnings are observed as permanent earnings of the second generation are commonly not available due to data limitations. This is means that most of studies simply compute RRs by selecting the second generation at a median age following the advises by [Haider and Solon \(2006\)](#) to minimize the amount of the lifecycle bias in the case of IGE without considering that the age-rank profile may differs from the age-earnings profile. Accordingly, a recent work by [Nybom and Stuhler \(2017\)](#) adapts the approach originally proposed by [Haider and Solon \(2006\)](#) to analyze biases affecting various measures of the intergenerational earnings persistence usually adopted by scholars beside the IGE. For instance, in the case of rank measures, they show analytically that rank measurement errors are not classical since ranks are uniformly distributed and observed and true ranks have by construction the same variance. This means that as the estimated  $\varphi$  in the equation is equal to the following expression:

$$p \lim \hat{\varphi} = \frac{Cov(rank^c, rank^p)}{Var(rank^p)} \quad (4)$$

the only source of bias related to left-hand side measurement errors in equation 3 depends on the term  $Cov(rank^c, rank^p)$  which is showed to be lower than the true value by [Nybom and Stuhler \(2017\)](#) at any given age when yearly rather than permanent earnings of children are used. This derives from the fact that the covariance between permanent and yearly ranks is always lower than one. Moreover, unlike the case of the

lifecycle bias in the IGE, Nybom and Stuhler (2017) show that there is no a specific age at which the lifecycle bias affecting rank measures of intergenerational earnings persistence equals 0.

### 3 Data and sample selection

As in Barbieri et al. (2018), we estimate the intergenerational earnings persistence in Italy by exploiting the so-called AD-SILC dataset, built merging the 2004 to 2008 waves of the IT-SILC, which is a specific version of the Italian sample of the European Survey on Income and Living Conditions (EU-SILC) including some additional country-specific variables, with high-quality administrative data managed by the Italian Social Security Institute (INPS). The latter records employment characteristics and gross earnings of Italian workers (including personal taxes and pension contributions) from the moment they entered the labor market until 2014, together with other demographic characteristics (e.g. gender, year of birth, region of residence) and detailed information on every job relationship that individuals experience in a specific year (e.g. duration, fund where workers pay contributions). Whereas the IT-SILC waves include other important information at both the household and personal level which is absent in administrative archives and is fundamental for our empirical purposes. More in detail, it provides information about the specific relation between the respondent and the person of reference of the household and the year when respondents obtain their highest educational attainment. Additionally, all respondents are asked if they are still in education at the time of the interview. Therefore, by combining the latter two information reported in the IT-SILC it is possible to identify the specific year when all individuals left education. On the contrary, the former variable can be used to link children to their actual parents.

Using the AD-SILC dataset, we select two generations according to some specific rules. Firstly, we focus on father-son pairs, consistently with most studies estimating the IGE to avoid the potential selection bias arising from the low labor force participation of women. Secondly, we consider sons aged 16 to 30 in the 2004 to 2008 IT-SILC waves conditionally to have left education no later than 2 years before the year of the interview. This means that selected sons interviewed in the IT-SILC wave of 2004 left education from 2002 to 2004; selected respondents interviewed in the wave of 2005 from 2003 to 2005 and so on for the subsequent IT-SILC waves until that of 2008. Thus, since the

INPS administrative archives record individuals earnings until the end of 2014, we are able to observe all selected sons' with positive earnings from 1 to 6 years after they left education despite of their age. Obviously, at any given distance from the date when they left education, age differences will simply depend on educational differences as more educated sons generally leave education at a higher age. This selection approach is directly inspired to the one proposed by [Raitano and Vona \(2018\)](#) to analyze the association between parental background, proxied by fathers' years of education, and sons' earnings for a representative sample of Italian workers followed over their working career.

For each selected son we can exploit information provided by the IT-SILC on his parental relationships with members of the household to identify his coresident father. Given that Italy is a country where almost all young individuals live with their parents at least until they leave education, this selection rule allow us to link the 90.2% of all respondents that left education no later than 2 years before the year of the interview. Moreover, according to information provided by the 2004 to 2008 waves of the IT-SILC, about 1.5% of respondents does not live with their parents at 15 years old. Therefore, the percentage of sons that we are able to link to their actual fathers is considerably high.

Tough we can directly link the vast majority of sons that left education no later than 2 years before the interview to their actual fathers, it may be important to compare some summary statistics regarding the characteristics of sons in the selected sample to the same set of characteristics observed in the full sample of individuals at the same point of the lifecycle. For instance, the figure [A.1](#) in the appendix [A](#) shows the percentage of individuals with zero earnings in the two samples from 1 until 6 years after they left education. As we could expect, in both the two samples the percentage of sons with zero earnings decreases as the distance from the year in which they obtain their highest educational attainment becomes higher starting from the 35.2%, 1 year after they left education until the 13.3%, 6 years after.

Individuals in the two samples are highly comparable also in terms of gross earnings as it is shown in the figure [A.2](#) in the appendix [A](#). More in detail, on average sons' earnings increases quickly in the first 4 years after they left education to stabilize from the fifth year. Similarly, the earnings dispersion measured by the standard deviation seems to increase until the fifth year in both the selected sample and the full sample.

Eventually, the table [A.1](#) summarize many other socio-economic characteristics of the second generation comparing again our selected sample to the full sample of male

workers interviewed no later than 2 years after they left education whose earnings profile is observed until 6 years after their highest educational attainment. As it is clear from the table [A.1](#), the observed distribution of characteristics is similar in the two samples since selected sons are highly comparable to the full sample of sons in terms of age when leaving education (around 21 years old), weeks of work experience when leaving education (32.01 vs 35.87), years of education (around 11) and fund where workers pay contributions.

As regard the first generation, fathers are observed when their sons were from 1 to 14 years old. In order to decrease the incidence of right-hand measurement errors causing a downward bias in the estimated  $\beta$  from the equation [1](#), we include in our baseline model only fathers with at least 3 positive earnings observations. We thus compute an average according to the number of positive observations available in the 14 years period. Given that many fathers have positive earnings observations in most of the 14 years considered with an average of about 11 observations, our measure of fathers' earnings appears to be a good proxy of fathers lifetime earnings. In any case, in the section [4.1](#) additional estimates of the IGE are provided by considering a different number of observations to assess the robustness of our baseline estimates to right-hand measurement errors.

Table [2](#) reports summary statistics of our final selected sample for both sons and their fathers. Our main measure of economic outcome of the two generations is computed as the sum of all CPI adjusted earnings gross of personal taxes and pension contributions received by employees and self-employed workers. In the first years after the sons obtain their highest educational attainment, their earnings are obviously lower and less dispersed than fathers' since they are measured when the latter were around 38 years old on average. In any case, as we could expect, differences in terms of earnings level and dispersion become lower as we move along the sons' working career.

Since our selected sons are only 21.49 years old on average 1 year after they left education and 26.13 years old 6 years after, it appears clear from the table [2](#) that we are not able to precisely follow the selection rules suggested by [Haider and Solon \(2006\)](#) to minimize the lifecycle bias arising when lifetime earnings of the second generation are not available. Nevertheless the specific selection rules we adopt in this work allow us to reduce one of the two aspects described in the section [2](#) related to the lifecycle bias. More in detail, despite of sons' age, our estimated IGEs and RRs are obtained by taking all workers of the second generation approximately at the same point of their career since the distance from the year in which they left education is the same for all sons

Table 2: Final samples: summary statistics

Yrs after leaving educ.	Stat.	Earnings		Age	
		Sons	Fat.	Sons	Fat.
1	Mean	12385.37	24854.61	21.49	38.91
	Sd	8378.74	12354.09	3.26	5.12
	Obs.	574	574	574	574
2	Mean	15451.00	24391.42	22.31	38.80
	Sd	8673.39	12326.44	3.25	5.17
	Obs.	651	651	651	651
3	Mean	16923.08	24444.40	23.28	38.79
	Sd	9352.75	12506.84	3.29	5.08
	Obs.	713	713	713	713
4	Mean	18511.61	24186.50	24.18	38.86
	Sd	9989.01	12147.38	3.32	5.25
	Obs.	745	745	745	745
5	Mean	19924.23	24182.04	25.24	38.90
	Sd	10921.11	12156.76	3.32	5.21
	Obs.	749	749	749	749
6	Mean	20378.45	24104.99	26.13	38.91
	Sd	11010.41	11991.19	3.29	5.18
	Obs.	758	758	758	758

**Notes:** All earnings are CPI adjusted in 2012 Euro. At any given distance from the year in which sons left education, only workers with positive earnings and their fathers with at least 3 positive earnings observations recorded when the sons were 1 to 14 years old are considered. **Source:** Authors' elaborations on the AD-SILC dataset.

taken. This means that we are not estimating the intergenerational earnings persistence by generically selecting young sons within a given age range and thus at different points of their working career. The only source of lifecycle bias in our estimated coefficients is thus related to existing differences in the earnings growth rate over the working career which are likely to persist even later than the sixth year after the sons left education. In any case, in the section 5, we present an estimation of the lifecycle bias associated to selecting individuals by years of distance from their highest educational attainments and we compare it to the potential bias arising when individuals are generically selected by age. Therefore, using administrative records of Italian workers observed from 1980 to 2014, we adapt to Italy and to the selection rules adopted in this work the approach firstly proposed by [Haider and Solon \(2006\)](#) for the US.

## 4 New estimates of the intergenerational earnings persistence in Italy

### 4.1 Estimated IGEs

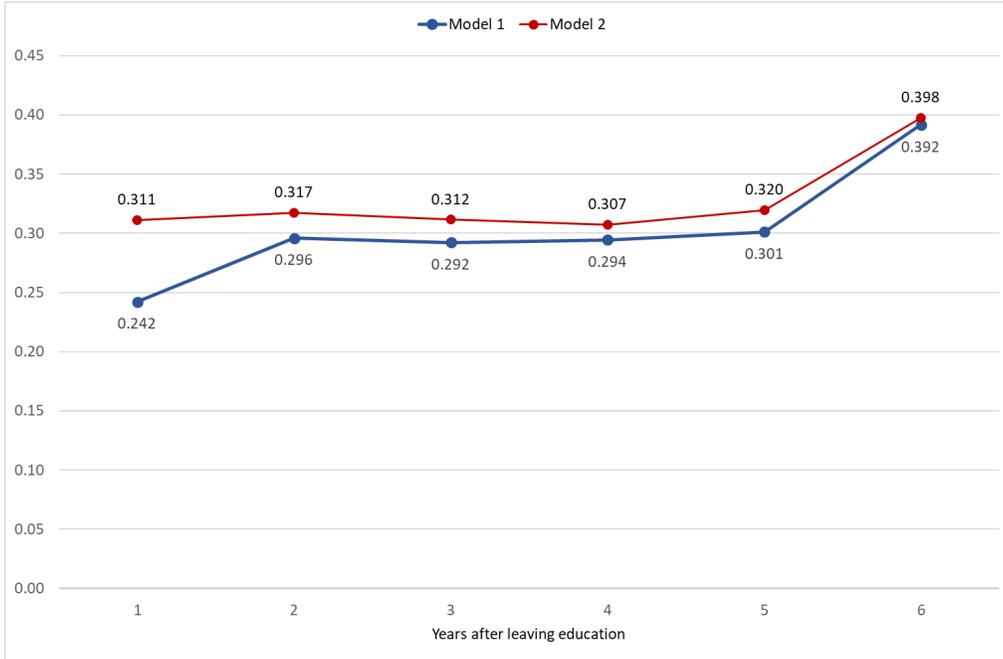
In this section we present our estimated IGEs for Italy, according to the selection rules previously described. In order to obtain our main measure of the degree of the intergenerational earnings persistence in Italy, we estimate an equation similar to equation 1 presented in the section 2 by considering only sons and their fathers. As a proxy of fathers' lifetime economic outcomes we use averaged earnings calculated by considering only those with at least 3 positive observations over a 14 years period. On the contrary, sons' earnings are selected by distance from when they left education in order to obtain different estimated IGEs at different points of their careers. As a general control variable, we include in our model the year dummies which indicate when sons' earnings are observed. These time dummies are necessary since, at any given distance, sons are observed in different years according to when they left education (model 1 henceforth). An additional model is estimated by controlling also for the number of sons' weeks of work experience gained before they left education (model 2 henceforth). However we do not expect very large differences in the results obtained from the 2 models because, as we know from the table A.1 in the Appendix A, sons had on average only about only about 32 weeks of work experience when leaving education<sup>6</sup>.

The estimated IGEs are plotted in the figure 1 for both the two models. The estimated IGE using the model 1, is around 0.24 taking sons 1 year after they left education and, as we could expect, it increases as we move over the sons' career reaching the value of 0.392 at the last distance considered (blue line in the figure 1). Therefore, according to these results the intergenerational earnings persistence appears to be considerably high in Italy though we are considering sons' at early stages of their career. However, if we compare our maximum estimated IGE at the 6th year to previous evidence for Italy obtained by using the TSTSLS method we can see that the degree of intergenerational mobility in Italy could appear to be slightly higher than previously suggested. However, it is not possible to say in this section whether these differences derive from the fact that our estimated IGE is downward biased because of left-hand side measurement errors

---

<sup>6</sup>Observe that we do not control for a polynomial of sons' age as it is often done in the literature that selects individual by age, since at any given distance sons' age is related to their human capital, an important channel through which fathers' economic status is transmitted to their sons (Becker and Tomes, 1979, 1986).

Figure 1: Estimated elasticity of sons' yearly earnings with respect to fathers'



**Notes:** At any given distance from the year in which the sons left education, those with positive earnings and their fathers with at least 3 positive earnings observations recorded when the sons were 1 to 14 years old are considered. In the model 1 we estimate the IGE by controlling for the year dummies. In the model 2 we control for year dummies and the sons' work experience gained before leaving education. **Source:** Authors' elaborations on the AD-SILC dataset.

or if previous studies report upward biased estimates of the IGE due to the TSTSLS methodology adopted. In any case, we will try to answer to this important question in the subsection 6.

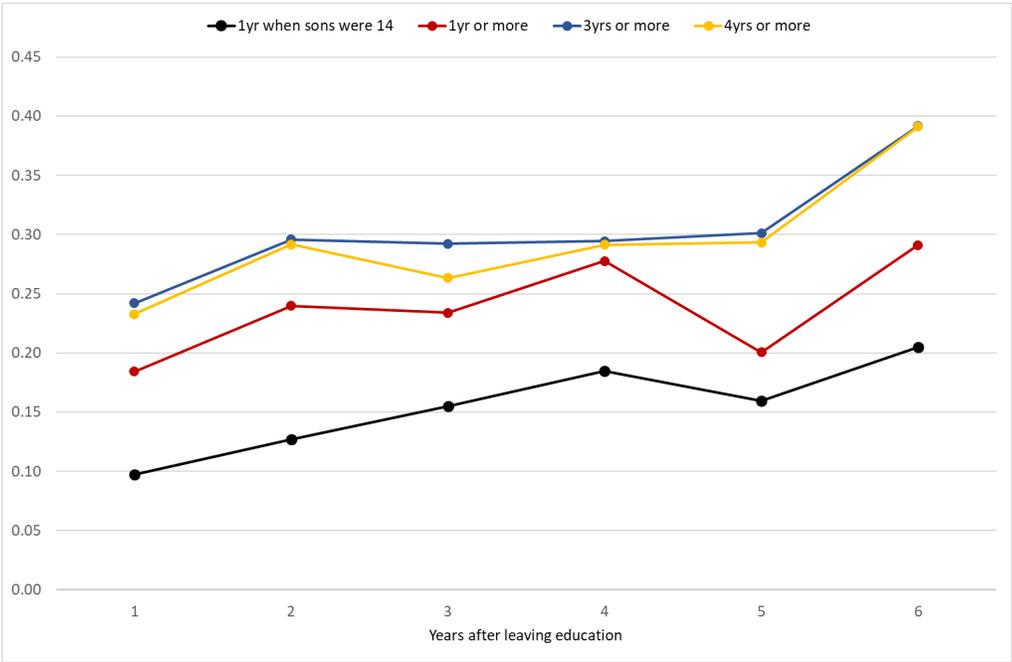
When we control also for the sons' weeks of work experience gained before leaving education, the estimated IGEs are slightly higher than those obtained from the baseline model 1 (red line in the figure 1). This estimated difference is more relevant considering the first years after sons' educational achievements and becomes lower the closer we get to the sixth year. This result suggests that apart for the first year when the difference is not negligible, we are basically selecting sons by their work experience at the sixth year after they left education. Therefore we will consider the model 1 as our baseline model from now on.

To test the sensitivity of our estimated IGEs to right-hand side measurement errors, we compare different estimated elasticities by varying the number of years used to average fathers' earnings in order to verify whether and to what extent the IGE is

influenced by this kind of methodological choices. More in detail, our baseline estimates obtained by excluding fathers with less than 3 positive earnings observations in the 14 years period considered is compared with alternative estimates produced: 1. by observing fathers in a unique year when sons were 14 years old; 2. by averaging fathers' earnings without excluding those with less than 3 positive earnings observations; 3. by excluding all those fathers with less than 4 positive earnings observations in the period when sons were 1 to 14 years old.

Figure 2 shows that our baseline estimates appear to be highly comparable to the

Figure 2: Estimated elasticity of sons' yearly earnings with respect to fathers': sensitivity to attenuation bias



**Notes:** Alternative estimated IGEs are provided by averaging fathers' earnings using a different number of positive earnings observations: 1 positive earnings observation when sons were 14 years old (black line); 1 or more positive earnings observations when sons were 1 to 14 years old (red line); 3 or more positive earnings observations when sons were 1 to 14 years old (blue line); 4 or more positive earnings observations when sons were 1 to 14 years old (yellow line). In all the estimates we control for the year dummies. **Source:** Authors' elaborations on the AD-SILC dataset.

ones obtained by considering fathers with at least 4 positive observations. This result suggests that our measure of fathers' earnings is basically robust to right-hand side measurement errors probably because, as we already stated in the section 3, the selected

fathers have positive earnings in most of the 14 years period during when they are observed<sup>7</sup>. On the contrary, estimates produced by measuring fathers' earnings in an unique year when sons were 14 years old or without excluding those fathers with 1 or 2 positive observations are considerably lower than our baseline. The attenuation bias is particularly relevant in the former case when the estimated IGE is about 50% lower than our baseline at the sixth year considered. Therefore, confirming previous evidence for the US (Solon, 1992; Zimmerman, 1992; Mazumder, 2005), it is extremely important to take into account right-hand measurement errors in order to avoid an overestimation of the degree of intergenerational mobility characterizing modern societies.

## 4.2 Estimated RRs

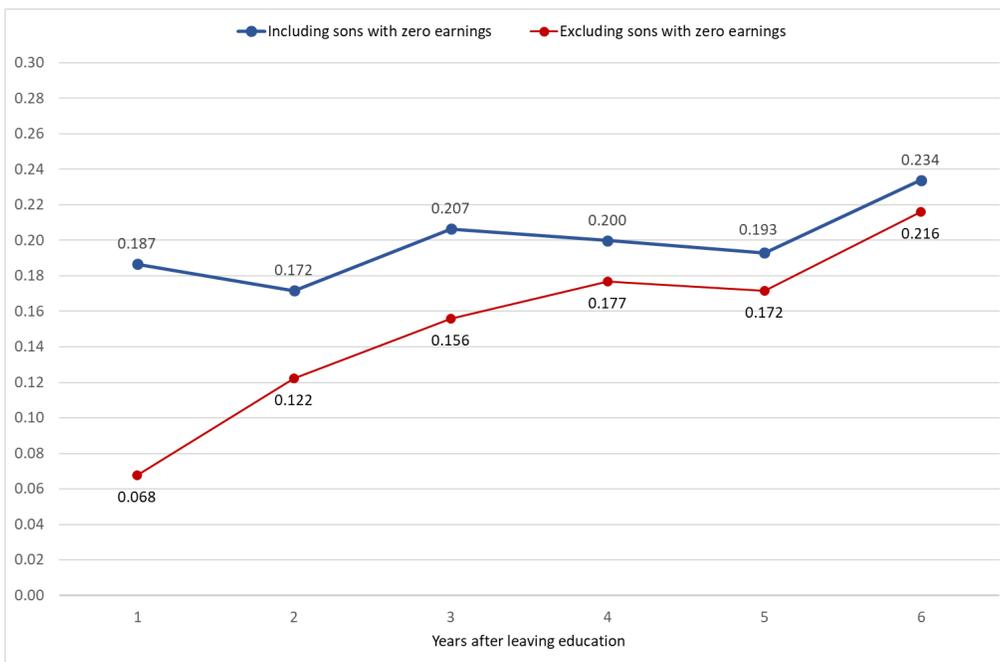
In this subsection we present our estimated RR by considering two alternative samples of sons. In the first one, we consider only sons with positive earnings observed from 1 to 6 years after they left education, as in the case of the estimated IGEs. In the second, we do not exclude zero earnings since, unlike the case in which the logarithm transformation is performed to estimate the IGE, ranking individuals by the percentile to which they belong in the earnings distribution do not force researchers to exclude those who are not working. Differently to what is commonly done in the empirical literature when the sample of sons is selected by age, we do not rank individuals of the two generations by birth cohort. This choice derives from the fact that, at any given distance from educational achievements, sons' age is strongly related to their educational level. Therefore, ranking individuals by birth cohort would have reduced a large fraction of the intergenerational transmission of economic status related to the sons' background-related human capital accumulation. Alternatively, as in the case of the estimated IGEs we control in all regressions for the year dummies to take into account that at any given distance, sons' economic outcomes are observed at different points in time.

Figure 3 plots all estimated RRs with (red line) or without (blue line) considering zero-earnings sons. In the former case, the estimated distance-rank-rank slope profile is extremely steep starting from a coefficient of 0.068 1 year after the sons left education to reach the value of 0.216 at the sixth year. On the contrary, when we include also zero-earnings sons, the estimated coefficients are higher, particularly by looking at the

---

<sup>7</sup>We also try to estimate the IGE by excluding those fathers with less than 5 or 6 positive earnings observations with no significant differences from our baseline estimate.

Figure 3: Estimated rank-rank slope



**Notes:** At any given distance from the year in which sons left education, individuals with either positive or positive and zero earnings and their fathers with at least 3 positive earnings observations recorded when the sons were 1 to 14 years old are considered. Sons' earnings are percentile ranked within each distance. In all the estimates we control for the year dummies. **Source:** Authors' elaborations on the AD-SILC dataset.

first year after educational achievements. This result probably derives from the fact that sons coming from poor families take a longer time to enter the labor market or to get a stable job than other sons the same distance from the year when leaving education. In any case, we decide to select the model in which consider only sons with positive earnings as our baseline to consider exactly the same samples of individuals in both estimates of the IGE and of the RR.

Apparently, our estimated RRs are very low when compared to the value of 0.317 obtained by [Chetty et al. \(2014a\)](#) for the US when they consider sons' and their parents' ranks. This results seems to be in contrast to what suggested by our estimated IGE which is only slightly lower than the ones previously estimated for the US ([Solon, 1992](#); [Zimmerman, 1992](#)). On the contrary, our estimated RR is similar to the ones estimated by [Bratberg et al. \(2017\)](#) for Germany, Sweden and Norway suggesting that the positional intergenerational persistence may not be so high in Italy compared to other developed countries. Thus, according to our results, a large fraction of the IGE

would be related to distributional change occurring across generations rather than to positional persistence along the earnings distribution.

However, even in the case of the RR, our highest estimated coefficient is obtained by observing sons' earnings only 6 years after they left education unlike previous evidence by [Bratberg et al. \(2017\)](#) where children are observed at a median age. Thus, as we are not able to observe sons' earnings over their entire working career, we have to take into account that our highest estimated RR is likely to be downward bias according to the only evidence on the lifecycle bias affecting RR measures of intergenerational persistence proposed by [Nybom and Stuhler \(2017\)](#). However, we need to evaluate the potential amount of downward bias in estimated RRs in the case of Italy as results presented by [Nybom and Stuhler \(2017\)](#) are not perfectly applicable to our case since they only focus on Swedish data. This the reason why we present an empirical estimation of the lifecycle bias affecting rank measures in Italy in the subsection [5.2](#).

## 5 Empirical estimates of the lifecycle bias in Italy

### 5.1 Lifecycle bias and estimated IGEs

In this section we adapt the textbook error-in-variables following the approach firstly proposed by [Haider and Solon \(2006\)](#) in their study on the US and subsequently applied by [Böhlmark and Lindquist \(2006\)](#) and [Chen et al. \(2017\)](#) to Swedish and Canadian data respectively. Given that we are only interested in the bias arising from measurement-errors in lifetime sons' earnings, we will from now use the simplify assumption of no right-hand side measurement errors , even though we are aware that even if we had properly averaged fathers' earnings over many yearly observations according to the approach commonly used in previous studies, we are not able to observe "true" lifetime earnings of fathers.

Beside giving information on the specific amount of bias related at any given age or distance to the use of yearly earnings as a proxy or lifetime earnings, this method allows us to evaluate the potential downward bias affecting our highest estimated elasticity obtained 6 years after sons left education. The method proposed by [Haider and Solon \(2006\)](#) consists in evaluating the bias associated to the use of yearly instead of lifetime earnings by regressing the former on the latter having at disposal a proper dataset which follows individuals approximately over their entire working career. Ideally, we

would like to estimate the IGE by means of OLS:

$$y_i^s = \alpha + \beta y_i^f + \epsilon_i \quad (5)$$

where  $y_i^s$  and  $y_i^f$  are the logarithm of sons' and fathers' lifetime earnings respectively,  $\epsilon_i$  is the classical disturbance and the estimated coefficient  $\beta$  is the IGE. Given that we can not directly observe lifetime earnings of sons, we are likely to obtain biased estimates of the IGE due to left-hand side measurement errors. However, we can directly measure a proxy of lifetime earnings for a representative sample of Italian workers thanks to the large longitudinal dimension of the AD-SILC dataset. We can thus regress their yearly earnings on their lifetime earnings according to the so-called “forward regression” of  $y_{it}$  on  $y_i$ :

$$y_{it} = \theta_t y_i + \omega_{it} \quad (6)$$

where  $y_{it}$  are yearly earnings which can be observed either at a given age or distance from the year when they left education. Therefore, by assuming that the estimated relation between yearly and lifetime earnings is the same among our selected sons and substituting the 6 into the 5 we obtain:

$$y_{it} = \theta_t \alpha + \theta_t \beta y_i^f + (\omega_{it} + \theta_t \epsilon_i) \quad (7)$$

The equation 7 is very useful to give us properly information on the amount of bias affecting the estimated IGE in Italy when lifetime earnings of sons are not available. In particular, by assuming that  $Cov(\omega_{it}, y_i^f) = 0$ , the lifecycle bias equals  $\theta_t$  and it disappears only when this estimated coefficient equals 1. On the contrary, the estimated  $\beta$  is likely to be downward biased when sons are observed when they are too young ( $\theta_t < 1$ ) and upward biased after a given median age ( $\theta_t > 1$ ). Therefore, according to this empirical approach, it could be theoretically possible to correct a biased estimated  $\beta$  by simply exploit the estimated  $\theta_t$  corresponding to the specific age at which sons are observed. In any case Haider and Solon (2006) suggest that the corrected  $\beta$  can be biased if the sample of individuals taken to estimate the “forward regression” 6 is taken from another country as the age-income profile may differ from a specific labor market to another or if the age-income profile changes within the same country from one cohort to another. Moreover, Nybom and Stuhler (2016) and Chen et al. (2017) point out that another source of bias may arise if  $Cov(\omega_{it}, y_i^f) \neq 0$ . Nevertheless, though all

previous studies are perfectly aware of all possible biases related to the estimated  $\theta_t$ , they still acknowledged that this is the best way available so far to choose the age at which sons should be taken when it is not possible to measure their lifetime earnings and to evaluate the potential bias related to any given age.

In this work, we estimate the “forward regression” for Italian workers by selecting a representative sample of male Italian workers that left education from 1979 to 1984 and followed from the subsequent 30 years (i.e. from 1980 to 2009 those that left education in 1979; from 1981 to 2010 those that left education in 1980 and so on) for a final sample of 4520 individuals. Following [Haider and Solon \(2006\)](#) and [Chen et al. \(2017\)](#) we obtain our proxy of lifetime earnings by averaging earnings of workers with at least 10 years positive observations in the period considered. However, given that we are not able to completely follow individuals exactly for their entire career as we cannot observe them after the 30th year after they left education, the 10 or more positive observations are taken from the 6th to 30th year of distance from their highest educational achievement<sup>8</sup>. This means that we are not able to measure their “true” lifetime earnings as we are not considering some potential earnings obtained at the beginning of the career or in the last years before retirement, as in [Chen et al. \(2017\)](#). In any case, since the selected male workers are observed for 21.84 years on average, we can consider our measure of lifetime earnings as a good proxy of lifetime earnings.

The results from our estimated “forward regression” is presented in the figure 4 for any given age from 20 to 55<sup>9</sup>. As in previous studies, we can see that the estimated  $\theta_t$  is lower than 1 when individuals are very young, it becomes greater than 1 later during the lifecycle and it decreases again in the later stages of the individuals working career. The age at which the estimated parameter is the closest to one is around 31 years old which is slightly lower than the one found in [Haider and Solon \(2006\)](#) for the US (i.e. around 32 years old). This result is related to the fact that Italian male workers have a lower level of education compared to other developed countries so that their age-earnings profile is slighter than in the US or Sweden.

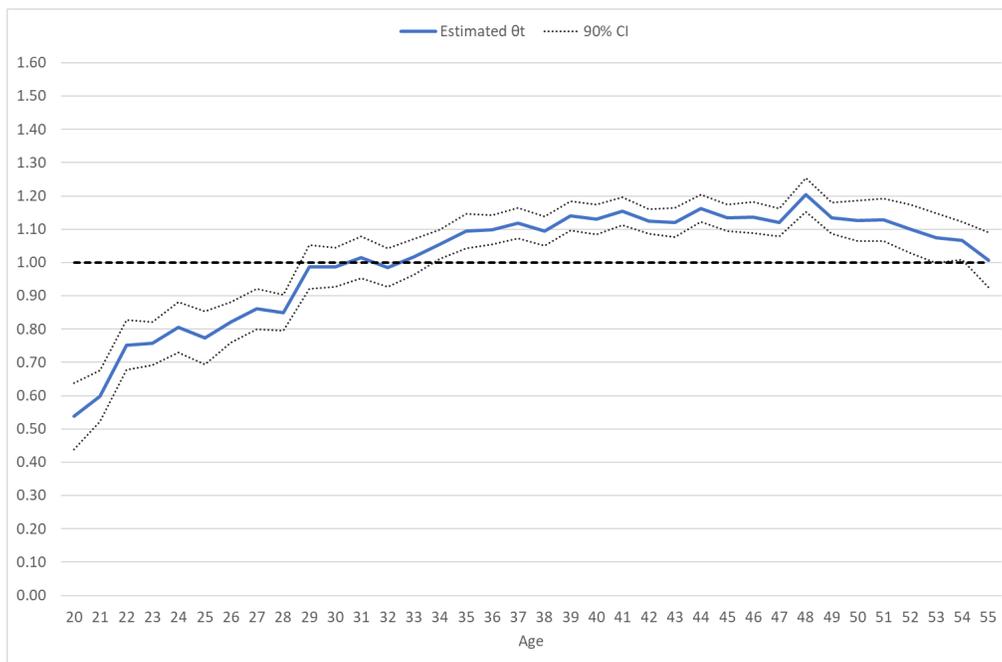
Additionally, we re-estimate the “forward regression” by observing yearly earnings of male individuals by years of distance from their highest educational achievement rather than by age. The estimated  $\theta_t$  are plotted in the figure 5 which shows that the lifecycle

---

<sup>8</sup>Alternative results are presented in the figure 3 in the appendix A by estimating the “forward regression” after computing the proxy of lifetime earnings considering individuals with at least either 8 or 15 positive earnings observations with no significant differences with respect to the baseline estimate.

<sup>9</sup>Even if results are not presented in the figure 3, we observe individuals also before 20 years old and after 60 years old if they are within the distance-range adopted.

Figure 4: Forward regression of log annual earnings by age on log lifetime earnings.

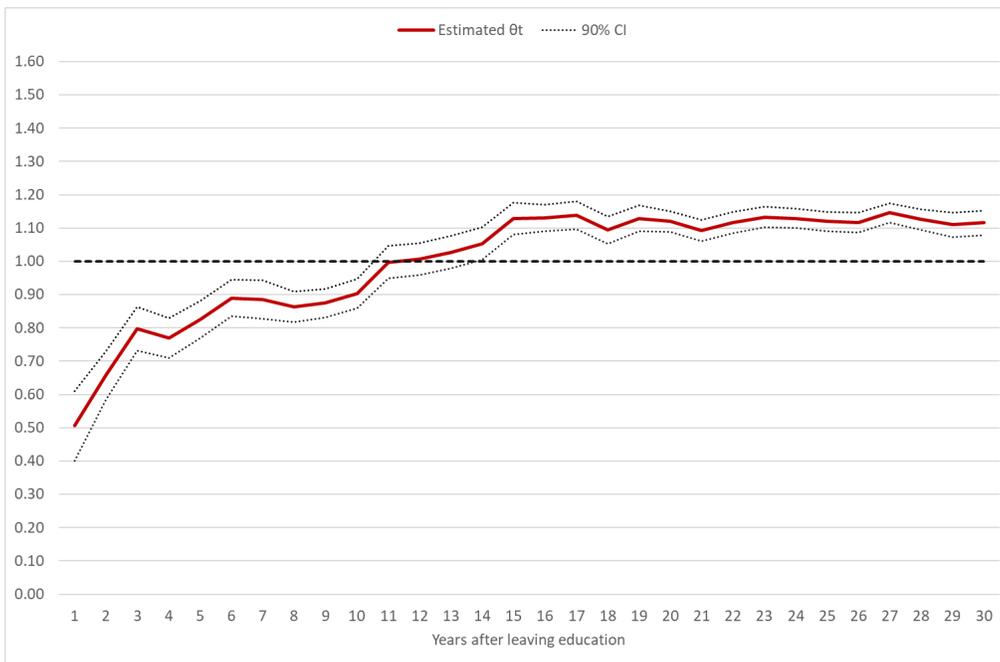


**Notes:** Workers that left education between 1979 and 1984 are followed for the 30 subsequent years. Their proxy of lifetime earnings has been calculated by averaging at least 10 positive observations in a 25 years period which begins at the 6th year after they left education and ends with the 30th. In all the estimates we control for the year dummies. **Source:** Authors' elaborations on the AD-SILC dataset.

bias is likely to be minimized around 11 years after workers left education, despite of their age. Moreover, at 6 years of distance the estimated downward biased results to be only around 10%. This means that our highest estimate elasticity presented in the section 4 is likely not to be so affected by the lifecycle bias even though our selected sample is not observed at a median age.

Another interesting result derives from the fact that our choice of selecting individuals by distance from the year in which they left education may reduce, at least partially, the lifecycle bias due to the young age at which sons are taken. For instance, given than our selected sons are on average 26 years old 6 years after they left education, the downward bias would have been greater and around 20% if sons were selected by age as the commonly adopted selection method of taking sons within a given age range, does not take into account that individuals are at a different stages of their working career according to their educational level. Therefore, this additionally result may suggest that a large fraction of the commonly estimated lifecycle bias is not directly related to

Figure 5: Forward regression of log annual earnings by years after leaving education on log lifetime earnings



**Notes:** Workers that left education between 1979 and 1984 are followed for the 30 subsequent years. Their proxy of lifetime earnings has been calculated by averaging at least 10 positive observations in a 25 years period which begins at the 6th year after they left education and ends with the 30th. In all the estimates we control for the year dummies. **Source:** Authors’ elaborations on the AD-SILC dataset.

the age-income profile which is steeper for higher educated individuals but to existing differences in the years of work experience gained by individuals at any given age.

## 5.2 Lifecycle bias and estimated RRs

In this subsection we now analyze left-hand side measurement errors affecting RR measures of the intergenerational earnings persistence following the method recently proposed by Nybom and Stuhler (2017) to adapt the one implemented by Haider and Solon (2006). In this case, we want to estimate the RR according to the following equation:

$$rank(y_i^s) = k + \varphi rank(y_i^f) + u_i \quad (8)$$

and given that again we are not able to follow sons over their entire career, we need to estimate the “forward regression” using ranks obtained using yearly earnings and ranks

obtained exploiting a proxy of lifetime earnings measured using the AD-SILC dataset and the representative sample of male workers that left education from 1979 to 1984, according to the following expression:

$$\text{rank}(y_{it}) = \lambda_t \text{rank}(y_i) + v_{it} \quad (9)$$

As in the case of the elasticity, by substituting the 9 into the 8, we can obtain the following equation:

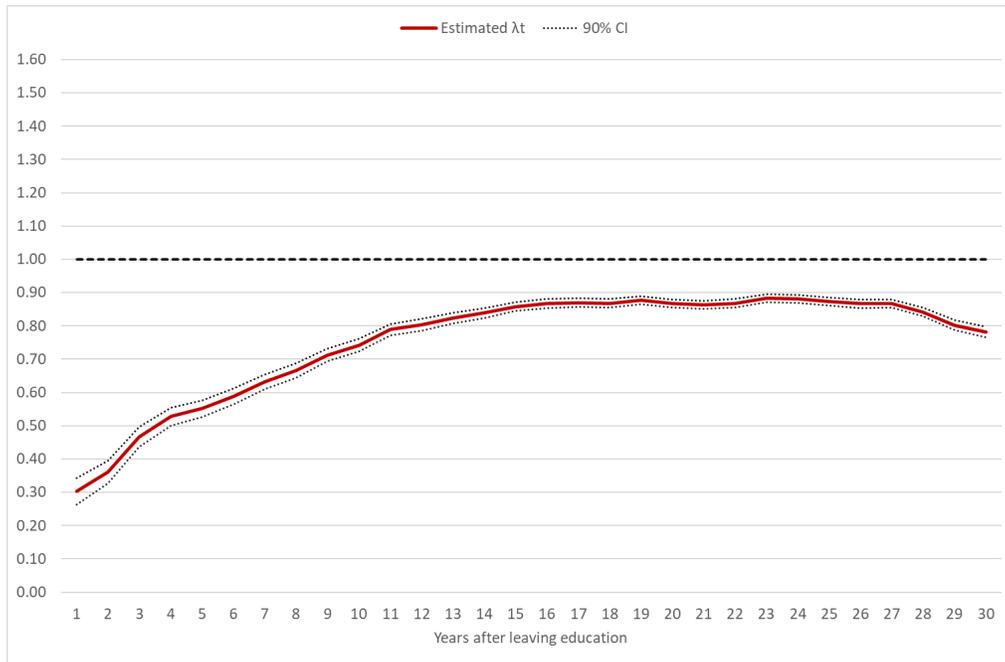
$$\text{rank}(y_{it}^s) = k + \lambda_t \varphi \text{rank}(y_i^f) + (v_{it} + \lambda_t u_i) \quad (10)$$

where the probability limit of the estimated RR is  $\lambda_t \varphi$ , assuming that  $v_{it}$  is uncorrelated with  $\text{rank}(y_i^f)$ . Thus, in this case, the amount of bias depends only on the estimated  $\lambda_t$ . As formally showed by Nybom and Stuhler (2017), unlike the case of the classical measurement error arising when estimating the elasticity, the measurement error affecting RR measures is non classical. This means that the estimated  $\lambda_t$  cannot be greater than 1 and it equals 1 only when the observed rank is equal to the true rank for all individuals. Unfortunately, Nybom and Stuhler (2017) show that in Sweden this is not the case as the estimated  $\lambda_t$  is lower than 1 at any given age with a maximum value of about 0.9 starting from around 35 years old. In our case we present the estimated forward regression for any given distance from the year in which or sample of males individuals left education. The estimated  $\lambda_t$  are presented in the figure 6, from 1 to 30 years of distance from when workers left education.

Confirming previous evidence from Sweden, the estimated  $\lambda_t$  is lower than 1 at any given point of the individuals' career even though the amount of bias seems to be more stable compared to the case of the elasticity. Moreover, consistently with previous evidence by Nybom and Stuhler (2017), the minimum amount of bias seems to be around 10% also in Italy, starting from 15 years of distance from individuals' highest educational achievements.

The results showed in figure 6 suggest that while our highest estimated IGE should not be consistently downward biased due to the lifecycle bias, our highest estimated RR presented in the subsection 4.2 is likely to be strongly affected by left-hand side measurement errors. This means that the degree of intergenerational earnings persistence measured by positional correlation in the earnings distribution between fathers and their sons is underestimated when earnings of sons are observed only 6 years after they left education. This is probably the reason why unlike the case of the estimated IGE

Figure 6: Forward regression of annual ranks selected by years after leaving education on lifetime ranks



**Notes:** Workers that left education between 1979 and 1984 are followed for the 30 subsequent years. Their proxy of lifetime earnings has been calculated by averaging at least 10 positive observations in a 25 years period which begins at the 6th year after they left education and ends with the 30th. In all the estimates we control for the year dummies. **Source:** Authors' elaborations on the AD-SILC dataset.

which is very close to the ones obtained for the US, the RR is closer to the coefficients estimated by [Bratberg et al. \(2017\)](#) in Germany, Norway and Sweden.

## 6 Corrected measures of intergenerational earnings persistence in Italy

In the previous sections we presented different estimated IGEs and RRs observing the sons from 1 to 6 years after they left education and we evaluated the potential amount of lifecycle bias deriving from the fact that we selected sons at early stages of their career. Nevertheless, the estimated “forward regression” suggests that our estimated IGE is unlikely to be strongly downward biased taking sons at 6 years of distance from when they left education. On the contrary, the estimated RR is probably too low due to non-classical measurement errors affecting rank measures of the intergenerational

persistence. In this section, using the estimated coefficients in equation 5 and 9, we correct our highest estimated IGE and RR in order to obtain measures which are comparable with those obtained for other countries in study that take sons at a median age.

Firstly, if we use the estimated parameter from the “forward regression” of the logarithm of yearly earnings on lifetime earnings obtained at the 6th year of distance from individuals’ highest educational achievements, we obtain a corrected IGE and RR of 0.441 and 0.367 respectively (table 4, panel A). However, according to what suggested by Haider and Solon (2006), Nybom and Stuhler (2016) and Nybom and Stuhler (2017), we are perfectly aware that the estimated coefficients in the “forward regressions” can be exploited to correct estimates of persistence affected by the lifecycle bias only if the distance-earnings profile of individuals observed during their entire career is comparable to that of the selected sons. In other terms, one should not use the estimated  $\theta_t$  and  $\lambda_t$  for a given country to correct the downward biased IGE or RR of other countries. Moreover, even considering the estimated parameters of the “forward regression” for the same country for which we want to analyze the degree of the intergenerational persistence, there could be other sources of bias deriving from the fact that the age or distance-income profile may vary from one cohort to another.

In our case, we are not worried by the first source of bias as we properly estimate the “forward regressions” considering male Italian workers to correct the estimated IGEs and RRs of Italian sons. On the contrary, there could be a source of bias deriving from the fact that we estimate the “forward regression” for male Italian workers that left education from 1979 to 1984, thus about 20/25 years before the selected sons. This means, that the age-income profile may be changed over time also because of the increase in the number of years individuals spend in education.

To verify if this potential source of bias is affecting our corrected IGE and RR, we produce new estimates of persistence for a subsample of sons that left education from 2002 and no later than 2004 observed between 10 and 12 years after they left education. Therefore, earnings of individuals that left education in 2002 are averaged considering positive observations between 10 and 12 years of distance from their highest educational achievements and those of sons that left education in 2004 at 10 years of distance. According to these selection rules, we obtain a final subsample of 345 sons observed 10.43 years after they left education on average<sup>10</sup> (3). Their mean age is 30.60

---

<sup>10</sup>As in our case, most of empirical studies that estimate the IGE are forced to use very small samples due to data limitations. See for instance the works by Björklund and Jäntti (1997), Mazumder (2005)

Table 3: Summary statistics: subsample of sons observed from 10 to 12 years after they left education

	Sons			Fathers	
	Earnings	Age	Yrs after leaving educ.	Earnings	Age
Mean	22110.41	30.60	10.43	24097.16	39.20
Sd	13082.83	3.50	0.41	11562.05	5.30
Obs.	345	345	345	345	345

**Notes:** The subsample of sons is obtained by considering individuals that left education from 2002 and no later than 2004. Their earnings are averaged considering positive observations between 10 and 12 years of distance from their highest educational achievements. Monetary values are CPI adjusted at 2012 prices. **Source:** Authors' elaborations on the AD-SILC dataset.

and their CPI adjusted earnings are slightly more dispersed than those of their actual parents observed in a 14 years period and around 39 years old.

These newly estimated measures of persistence are presented in the table 4 (panel B) and are consistently higher than the ones estimated for the full sample of sons observed at 6 years of distance from when they left education. In particular, the estimated IGE is now 0.414 and the estimated RR 0.265. In this case, these estimated measures should be less affected by left-hand side measurement errors as we are closer to the points in career that minimize the lifecycle bias according to the estimated “forward regressions”. In particular, the downward bias should be only 5.60% in the case of the IGE and 23.75% in the case of the RR. Therefore, if the distance-income profiles presented in the figures 5 and 6 properly describe the distance-income profile of our selected sons, we would obtain similar IGEs and RRs either correcting the estimated coefficients obtained observing sons 6 years after they left education or around 10.43 years after.

Reassuringly, both our corrected measures of persistence are very similar in the two cases suggesting that the coefficients estimated from the “forward regressions” are properly applicable to our selected sons. In particular, the corrected IGE equal 0.441 and 0.439 when we correct the estimated coefficient 6 years or 10.43 after sons' left education respectively. On the contrary, the corrected RR is only slightly lower in the 10.43 years of distance case than when the full sample of sons are observed 6 years after they left education.

Using the results obtained after the correction, we can finally compare our estimated and [Schnitzlein \(2016\)](#).

Table 4: Comparable estimated IGEs and RRs

Panel A: Sample of sons observed 6 yrs after leaving education			
	Estimated coeff.	Estimated lifecycle bias	Corrected coeff.
IGE	0.392*** (0.100)	11.04%	0.441
Obs.	758		
RR	0.216*** (0.051)	41.17%	0.367
Obs.	758		
Panel B: Subsample of sons observed 10.43 yrs after leaving education			
	Estimated coeff.	Estimated lifecycle bias	Corrected coeff.
IGE	0.414*** (0.030)	5.60%	0.439
Obs.	345		
RR	0.265*** (0.051)	23.75%	0.348
Obs.	345		

**Notes:** Estimates in panel A are obtained by considering the full sample of sons that left education starting from 2002 and no later than 2008. Their positive earnings are observed 6 years after their highest educational achievements. The IGE is estimated by only controlling for the year dummies as the distance is the same for all sons. Estimates in panel B are obtained by considering the subsample of sons that left education starting from 2002 and no later than 2004. Their earnings are averaged considering positive observations between 10 and 12 years of distance from their highest educational achievements. In this case, the IGE is estimated by controlling for the year and distance dummies. In both panel A and B the estimated RRs are obtained after ranking sons by distance and controlling for the year dummies. **Source:** Authors' elaborations on the AD-SILC dataset.

IGE and RR to other studies for other countries in order to summarize the degree of the intergenerational earnings persistence in Italy and answer to the following question: “Is the degree of intergenerational earnings persistence overestimated in Italy due to the TSTSLS method used in previous studies?” Given that our corrected IGE of 0.439 is extremely close to the ones obtained by Piraino (2007) and Barbieri et al. (2018) (see table 1) we can answer to the latter question with a clear “no”. The degree of the intergenerational earnings persistence in Italy is very high and comparable to those obtained for the UK and the US which are commonly considered as the countries with the lowest level of intergenerational mobility among developed countries. Moreover, also the corrected RR slope is very close to the values obtained for the US, suggesting that only a small fraction of the estimated IGE is related to changes in inequality occurred

across generations whereas a large fraction of the intergenerational persistence is related to the copula of the distribution.

In any case, our estimated measures of mobility should still be a lower bound as we are not able to directly follow the two generations over their entire careers due to data limitations even though we tried to minimize the amount of both left-hand and right-hand side measurement errors. In particular, as showed by Mazumder (2005) even averaging fathers' earnings over 4/5 years may produce downward biased estimates of persistence due right-hand side measurement errors. Nevertheless, our estimates of persistence may be considered comparable to most of studies on intergenerational mobility which use the OLS estimator and are not able to follow the two generations over their entire career.

## 7 Concluding remarks

In this work we present brand new estimates of the degree of intergenerational earnings persistence in Italy taking for the first time actual fathers-sons pairs rather than imputed fathers'-sons' earnings. Using the AD-SILC dataset obtained by merging INPS administrative archives to the 2004 to 2008 Italian samples of the EU-SILC, we are able to link sons to their actual fathers by considering that a large fraction of male individuals are still coresiding with their parents when they have just left education. We then measure fathers' earnings by averaging earnings of individuals with at least 3 positive observations in 14 years period which starts when their sons were 1 years old. Sons' earnings are instead observed from 1 to 6 years after they left education.

According to these selection rules we find an estimated IGE of 0.392 and a RR of 0.216 at the 6th year after sons' highest educational achievements. The first estimated coefficient is not that far from those obtained in previous works for Italy which exploit the TSTSLS and from the ones commonly obtained for the US. On the contrary, the estimated RR appear to be consistently lower than the ones estimated for the US. For this reason we adapt the methodology firstly proposed by Haider and Solon (2006) to a consistent sample of Italian male workers which left education from 1979 and 1984 whose earnings are observed for the subsequent 30 years in order to evaluate the career-income/rank profile.

Results show that while the estimated IGE is unlikely to be consistently underestimated due to the lifecycle bias, the estimated RR is strongly downward biased

observing sons no later than 6 years after they left education. This is the reason why we re-estimate the two measures on intergenerational earnings persistence for a subsample of sons that left education between 2002 and 2004 so that they are followed between the 10th and 12th year after they left education. Then, using these newly estimated coefficients and exploiting the predicted lifecycle bias obtained from the “forward regressions” of yearly earnings on lifetime earnings and yearly ranks on lifetime ranks for any given point in career, we obtain comparable estimates of the IGE and RR which equals 0.439 and 0.348 respectively. According to these final results, it is possible to summarize that the degree of intergenerational earnings persistence in Italy is very high if compared to other developed countries either considering the IGE or the RR as comparable levels of background-related earnings advantages are obtained only in the US and the UK, the less mobile developed countries.

To conclude, contrary to what is often predicted in the empirical literature on intergenerational mobility (Blanden, 2013), the estimated IGEs obtained in previous studies (Barbieri et al., 2018; Mocetti, 2007; Piraino, 2007) which exploited the TSTSLS method do not seem to be strongly upward biased even if these studies were not able to directly link earnings of sons to those of their actual fathers. Therefore, the level of intergenerational earnings mobility in Italy is likely to be very low regardless of the estimation method used. Moreover, given that it is still not possible to directly follow the two generations over their entire careers, our final IGE and RR might still be a lower bound of the true values because of potential residual left/right-hand side measurement errors.

## References

- Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association*, 87(418):328–336.
- Arellano, M. and Meghir, C. (1992). Female labour supply and on-the-job search: an empirical model estimated using complementary data sets. *The Review of Economic Studies*, 59(3):537–559.
- Barbieri, T., Bloise, F., and Raitano, M. (2018). Intergenerational earnings inequality in Italy: New evidence and main mechanisms. *Working Paper CIRET, N. 2/2018*.

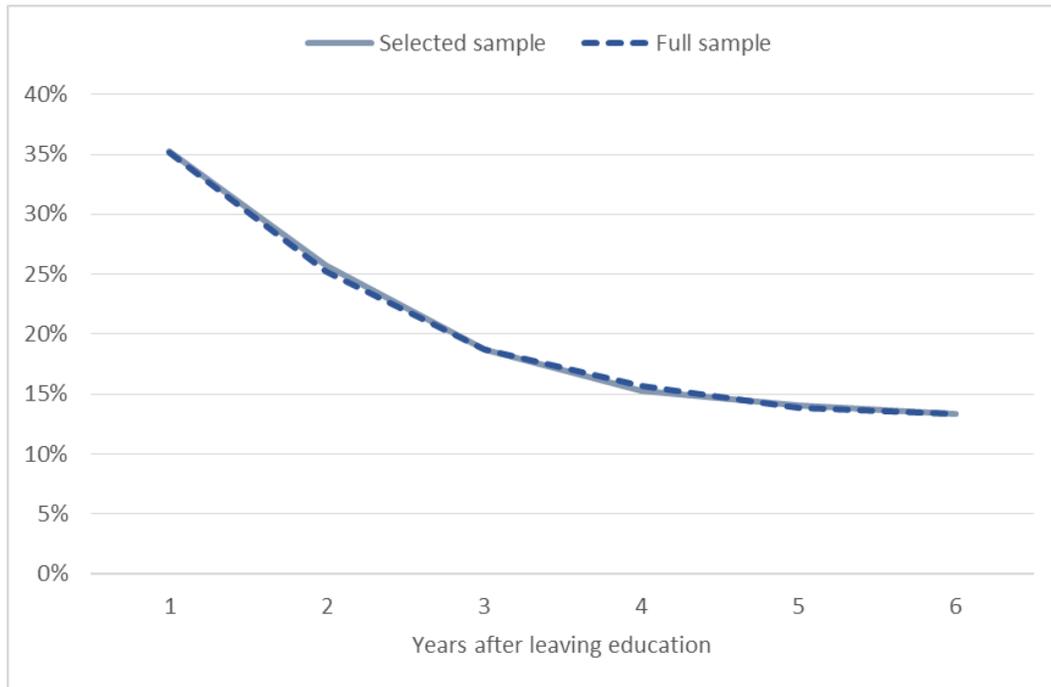
- Becker, G. S. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, 87(6):1153–1189.
- Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of labor economics*, 4(3, Part 2):S1–S39.
- Behrman, J. R. and Taubman, P. (1986). Birth order, schooling, and earnings. *Journal of Labor Economics*, 4(3, Part 2):S121–S145.
- Björklund, A. and Chadwick, L. (2003). Intergenerational income mobility in permanent and separated families. *Economics Letters*, 80(2):239–246.
- Björklund, A. and Jäntti, M. (1997). Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, 87(5):1009–1018.
- Black, S. E. and Devereux, P. J. (2011). *Recent Developments in Intergenerational Mobility*, volume 4 of *Handbook of Labor Economics*. Elsevier.
- Blanden, J. (2013). Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. *Journal of Economic Surveys*, 27(1):38–73.
- Böhlmark, A. and Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: replication and extension for Sweden. *Journal of Labor Economics*, 24(4):879–896.
- Bratberg, E., Anti Nilsen, Ø., and Vaage, K. (2005). Intergenerational earnings mobility in Norway: Levels and trends. *Scandinavian Journal of Economics*, 107(3):419–435.
- Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D. D., and Vaage, K. (2017). A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US. *The Scandinavian Journal of Economics*, 119(1):72–101.
- Cervini-Plá, M. (2015). Intergenerational earnings and income mobility in Spain. *Review of Income and Wealth*, 61(4):812–828.
- Chen, W. H., Ostrovsky, Y., and Piraino, P. (2017). Lifecycle variation, errors-in-variables bias and nonlinearities in intergenerational income transmission: new evidence from Canada. *Labour Economics*, 44:1–12.

- Chetty, R. and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.
- Chetty, R. and Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.
- Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014a). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623.
- Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014b). Is the United States still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, 104(5):141–47.
- Dahl, M. W. and DeLeire, T. (2008). *The association between children’s earnings and fathers’ lifetime earnings: estimates using administrative data*. University of Wisconsin-Madison, Institute for Research on Poverty.
- Gregg, P., Macmillan, L., and Vittori, C. (2017). Moving towards estimating sons’ lifetime intergenerational economic mobility in the UK. *Oxford Bulletin of Economics and Statistics*, 79(1):79–100.
- Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review*, 96(4):1308–1320.
- Lefranc, A. and Trannoy, A. (2005). Intergenerational earnings mobility in France: Is France more mobile than the US? *Annales d’Economie et de Statistique*, (78):57–77.
- Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data. *The Review of Economics and Statistics*, 87(2):235–255.
- Mocetti, S. (2007). Intergenerational earnings mobility in Italy. *The BE Journal of Economic Analysis & Policy*, 7(2):1–25.
- Nybom, M. and Stuhler, J. (2016). Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation. *Journal of Human Resources*, 51(1):239–268.

- Nybohm, M. and Stuhler, J. (2017). Biases in standard measures of intergenerational income dependence. *Journal of Human Resources*, 52(3):800–825.
- Olivetti, C. and Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850-1940. *The American Economic Review*, 105(8):2695–2724.
- Piraino, P. (2007). Comparable estimates of intergenerational income mobility in Italy. *The BE Journal of Economic Analysis & Policy*, 7(2):1–27.
- Raitano, M. and Vona, F. (2018). From the cradle to the grave: The influence of family background on the career path of Italian men. *Oxford Bulletin of Economics and Statistics*, forthcoming.
- Schnitzlein, D. D. (2016). A new look at intergenerational mobility in Germany compared to the U.S. *Review of Income and Wealth*, 62(4):650–667.
- Solon, G. (1992). Intergenerational income mobility in the United States. *American Economic Review*, 82(3):393–408.
- Zimmerman, D. J. (1992). Regression toward mediocrity in economic stature. *American Economic Review*, 82(3):409–429.

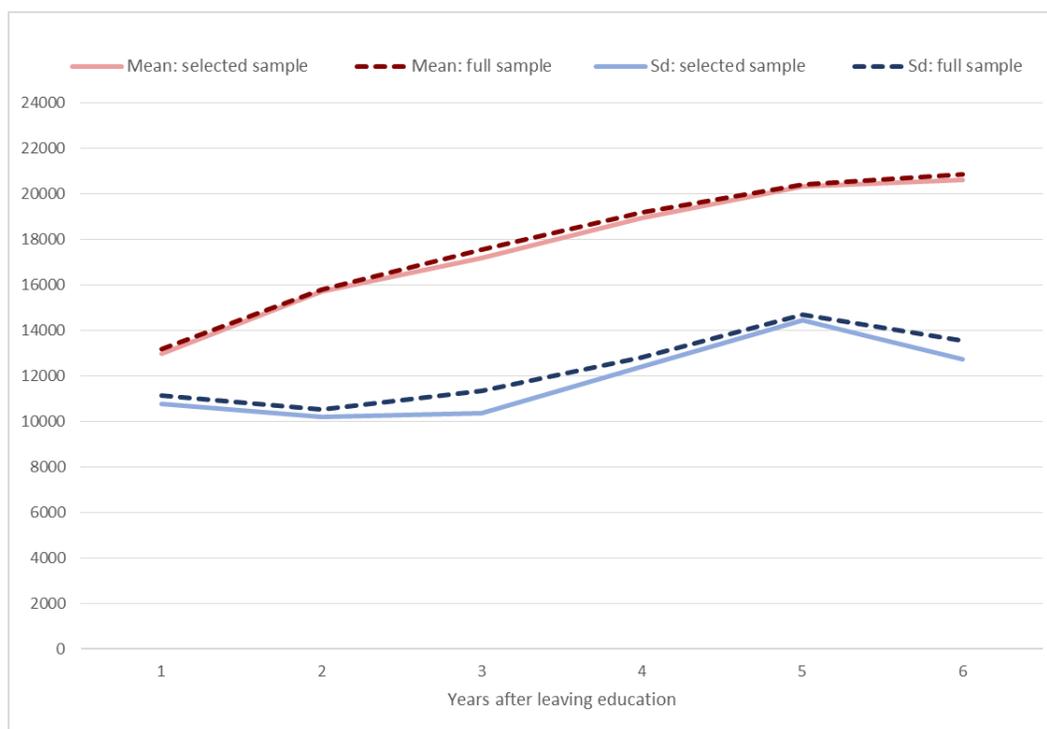
# A Appendix

Figure A.1: Percentage of sons with zero earnings: selected sample v.s. full sample



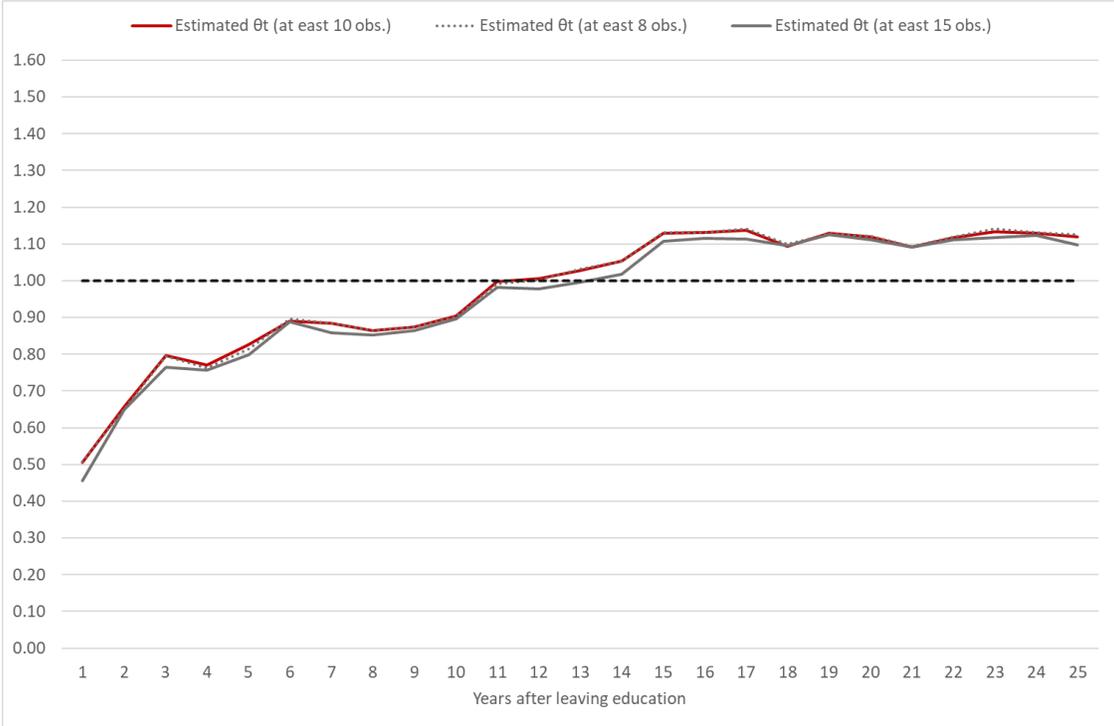
**Source:** Authors' elaborations on the AD-SILC dataset.

Figure A.2: Sons' earnings by years of distance after leaving education: selected sample v.s. full sample



**Notes:** Only individuals with positive earnings are considered. **Source:** Authors' elaborations on the AD-SILC dataset.

Figure A.3: Empirical test of the lifecycle bias: sensitivity test to the number of positive earnings observations used to proxy lifetime earnings



**Notes:** Workers that left education between 1979 and 1984 are followed for the 30 subsequent years. Their proxy for lifetime earnings has been calculated by averaging at least 10 positive observations in a 25 years period which begins at the 6th year after they left education and ends with the 30th (red line). Alternative measures of lifetime earnings have been calculated by averaging at least 8 (gray line) or 15 (dotted line) positive observations in a 20 years period which begins at the 6th year after they left education and ends with the 30th (red line) In all the estimates we control for the year dummies.  
**Source:** Authors' elaborations on the AD-SILC dataset.

Table A.1: Selected sample v.s. full sample: summary statistics

	Selected sample		Full sample	
Age when leaving educ.	20.58	(3.46)	20.77	(3.57)
Weeks of work experience when leaving educ.	32.01	(72.15)	35.87	(80.09)
Years of educ.	10.82	(4.60)	11.07	(4.71)
Work Status:				
<i>Private employee</i>	77.69%		76.41%	
<i>Other private employee</i>	1.86%		1.86%	
<i>Public employee</i>	3.65%		4.20%	
<i>Cococo</i>	5.79%		5.91%	
<i>P. IVA</i>	1.20%		1.08%	
<i>Craft</i>	3.04%		3.65%	
<i>Salesman</i>	3.41%		3.27%	
<i>Agricultural worker</i>	0.89%		0.80%	
<i>Private contractor</i>	2.47%		2.81%	

**Notes:** Only individuals with positive earnings are considered. Standard deviations in parentheses.  
**Source:** Authors' elaborations on the AD-SILC dataset.