I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17476

Heritability in the Labour Market: Evidence from Italian Twins

Sonia Brescianini
Lorenzo Cappellari
Daniele Checchi

NOVEMBER 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17476

Heritability in the Labour Market: Evidence from Italian Twins

Sonia Brescianini
Istituto Superiore di Sanità

Lorenzo Cappellari
Università Cattolica di Milano, LISER and IZA

Daniele Checchi
Università degli Studi di Milano and IZA

NOVEMBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Heritability in the Labour Market: Evidence from Italian Twins*

We use administrative data on educational attainments and life-time earnings to study their correlations among Italian twins. Using the ACE decomposition, we find that heritability in education accounts to almost half of the variance, especially for younger birth cohorts. With respect to labour market outcomes, we find that only for the oldest cohorts there is a greater share of inequality that can be attributed to idiosyncratic factors compared to education, and symmetrically a lower share due to genetics, while the impact of shared environment remains stable among the youngest cohorts. We suggest that increased employment flexibility may be responsible for the decline in the environmental component. Using a larger sample of pseudo-twins (individuals sharing birth date, birth place and family name) we confirm previous results, providing evidence that heritability also drives labour market attachment and prosocial behaviour.

JEL Classification: D31, E21, I24, J31

Keywords: heritability, inequality, labour market outcomes, Italy

Corresponding author:
Daniele Checchi
Università degli Studi di Milano
Via Festa del Perdono 7
Milan, 20122
Italy

E-mail: daniele.checchi@unimi.it

* This research has been made possible under the Visitinps call for projects (3rd wave), which is gratefully aknowledged. We thank INPS DSR and ISS for making their data available for the present analysis. Results have been presented to the GEOM conference (Bari, June 2024). Cappellari gratefully acknowledges funding from Università Cattolica, grant D32-2022 UNEQUAL. Usual disclaims apply.

1. Introduction

Understanding the extent to which observed inequalities depend on one's family background is crucial for unpacking the drivers of socio-economic disparities as well as for assessing their long-term persistence. One prominent approach adopted to investigate this issue is through twin studies. Like other siblings, twins share the family of origin and also experience a common background outside the family, through schools and youth neighbourhoods, with any correlation of their outcomes reflecting those shared influences. Unlike regular siblings, twins also share the date of birth, which potentially reinforces their exposure to shared influences both within and outside the family, making the correlations of their outcomes even stronger. Furthermore, twins share the same in utero experiences and this allows to have a more precise matching compared to siblings. Moreover, when information on twin zygosity is available, it can be leveraged to decompose the between-twins correlation into components due to genetics (or pre-birth influences) and the shared environment (post-birth influences), an approach grounded in behavioural genetics and known as the ACE model (A = Additive genetic factors, C = Common/shared environmental factors, E = Unique environmental factors).

In this paper, we apply the ACE model to investigate the sources of education and income inequality in Italy. Linking data from the Italian Twin Registry with administrative records from the Italian Social Security Institute (INPS), we disentangle the relative contributions of genetics and shared environments to the inequality of outcomes such as years of education, permanent earnings, and permanent working time, with permanent variables defined as averages of their annual counterparts over the life cycle. Twins correlations of education have already been estimated in Italy, although on smaller samples compared to ours. In particular, Branigan et al (2013) published a meta-analysis of data from mostly Western countries reporting a heritability of around 20% for Italy. Later, Silventoinen et al (2020) in an analysis of 28 twin cohorts reported a heritability of around 30% for the Italian cohort.[1] Ours is the first study of this kind that investigates correlations of labour market outcomes. Italy is an interesting setting for this type of analysis because its labour market has undergone major institutional reforms over the past decades aimed at increasing employment flexibility. Our main contribution, therefore, is to understand if and how the sources of inequality have changed in the era of labour flexibility.

Additionally, to broaden the range of outcome variables considered, we exploit social security data that could not be linked to the Twin Registry to complement the analysis with an application of the ACE model to the population of pseudo twins. These are pairs of individuals born on the same day

[1] See Zhelenkova and Panichella 2023 for educational correlations among Italian *siblings*.

in the same municipality and sharing the three letters of the surnames used by the tax authority to compile the individual tax code. Leveraging the sex composition of the pair, we treat them as twins of unknown zygosity. We validate the pseudo-twin approach on education, earnings, and working time data used in the (real) twins analysis and then perform ACE decompositions on variables such as unemployment, absence from work, parental leave, blood donations, and clergy membership.

The use of the ACE approach on twin data to study the impact of family background on socio-economic outcomes has a long history in the economics literature. Early work by Behrman and Taubman (1976) laid the foundations for understanding the role of genetic and environmental factors in shaping economic inequalities. These studies demonstrated the potential of twin data to disentangle the effects of shared family environments from individual-specific factors, providing insights into the sources of variation in education and income. Later work by Behrman and Taubman (1989) expanded on these ideas, emphasizing the importance of both genetic endowments and family environments in determining earnings outcomes.

Björklund, Jäntti and Solon (2005), highlighted the relevance of family background using different sibling types, including twins, to study earnings, showing that shared environmental factors play a crucial role in determining socio-economic outcomes. This approach reinforced the value of twin studies in understanding the complex interplay between genetics and environment.

More recent contributions have expanded the range of outcomes studied, including aspects such as financial behaviour and social preferences. For instance, Cesarini et al. (2009a, 2009b) used twin data to explore the heritability of economic preferences, such as risk-taking and trust, finding significant genetic components underlying these behaviours. Barnea, Cronqvist and Siegel (2010) and Cronqvist and Siegel (2015) further investigated the genetic influence on financial decision-making, including savings behaviour and portfolio choice, illustrating the broad applicability of the twin approach beyond traditional education and income.

Sacerdote (2007) provided a comprehensive review of the literature on twins and adoptees, summarizing key findings on the relative importance of genetic and environmental factors in determining educational attainment and earnings. His work emphasises the consistency of results across different contexts, supporting the robustness of the twin-based approach in understanding socio-economic inequalities.[2]

Some recent studies focus more closely on the heritability of labour market outcomes. Maczulskij (2013) uses twin data to explore the genetic contribution to working in the public sector, finding that

[2] More recently, Fagereng, Mogstad and Rønning (2021) apply the ACE framework to Norwegian administrative data on adoptees to examine the influence of genetic and environmental factors on wealth accumulation, highlighting the persistence of family background effects even in a highly redistributive welfare state.

34–40% of the variation can be attributed to genetic factors, with a significant role mediated by education. This study also shows that wage differentials between the public and private sectors diminish when controlling for genetic and other unobservable factors, emphasizing the importance of genetic influences in shaping career paths. Hyytinen et al. (2019) examine the genetic heritability of lifetime earnings in Finland, finding that genetic factors account for a significant share of earnings variance (54% for men and 39% for women), while shared environment plays a minor role. This underscores the contribution of inherited cognitive and non-cognitive abilities, which influence educational and occupational choices, with implications for intergenerational mobility and policies aimed at reducing economic inequalities. Papageorge and Thom (2019) demonstrate that genetic factors, summarized by a polygenic score, interact with childhood socioeconomic status to influence both educational attainment and labor market outcomes. Their analysis highlights how genetic potential can be either harnessed or constrained by environmental conditions, showing increasing genetic returns in the context of rising wage inequality.

The use of the ACE model in economics and other social or behavioural sciences has faced two main lines of criticism. Firstly, the model has been criticized for the restrictiveness of its underlying assumptions, such as gene-environment independence, additivity, the homogeneity of environments assumed for monozygotic (MZ) and dizygotic (DZ) twins, and the absence of assortative mating or genetic dominance. Various studies have provided insights into the validity of these assumptions (see, among others, Cesarini et al, 2009a, and Fagereng, Mogstad and Rønning 2021). Bingley, Cappellari and Tatsiramos (2024) provide a comparative assessment of the ACE assumptions, concluding that the equal environment assumption most severely biases (upward) heritability estimates.

A second and perhaps more fundamental line of critique questions the usefulness of heritability estimates, arguing that they are uninformative for policy analysis (Goldberger, 1979; Manski, 2011). According to this critique, variance decompositions like those of the ACE model do not provide insights into how outcomes would change under different policy interventions, and therefore heritability estimates cannot inform the potential effectiveness of policies aimed at reducing socio-economic inequalities. In his "eyeglasses analogy," Goldberger suggested that even if a trait like eyesight were found to have high heritability, it would not imply that interventions such as distributing eyeglasses would be ineffective.

In a recent reappraisal of the policy relevance of heritability, Rietveld (2024) acknowledges the limitations of heritability estimates in directly informing policy, particularly in terms of predicting the effects of interventions. However, he argues that heritability estimates can still provide valuable descriptive insights into the origins of socio-economic inequalities and their evolution under changing environmental conditions. Rietveld points out that while heritability estimates do not indicate how

easy or cost-effective it would be to eliminate inequalities, they can be used to assess whether the contribution of genetic and environmental factors to inter-individual differences changes due to policy-induced environmental conditions. For example, estimating the sources of variation in a trait before and after a policy change, such as extending health insurance to cover eyeglasses, could reveal a decrease in the impact of familial resources on outcomes.

These studies collectively provide the background for our analysis, which applies the ACE model to the Italian context. Our work aims to contribute to understanding how genetic and environmental factors influence education and income inequality in Italy, particularly in the context of labour market flexibilization, exploring whether these patterns hold in a country that has experienced significant changes in labour market institutions over recent decades.

Our ACE decompositions show that heritability in education accounts for almost half of the variance, especially for younger birth cohorts. Regarding labour market outcomes, we find that only for the oldest cohorts is a greater share of inequality attributable to idiosyncratic factors compared to education, and symmetrically a lower share due to genetics, while the impact of shared environment remains stable. We suggest that the flexibilization of labour market contracts may be responsible for the decline in the environmental component in the youngest cohorts.

We also provide an advancement in the use of administrative data, at least for a country that is quite strict in privacy protection. Linking data from the Italian Twin Registry with administrative records from the Italian Social Security Institute (INPS), we succeeded in obtaining a more precise measure of labour market outcomes of Italian twins. At the best of our knowledge, this is the first paper studying heritability in the labour market using Italian data.[3]

The paper is organized as follows. In the next section we describe our data. Section 3 explores the correlations among twins in educational attainments and labour market outcomes. Section 4 introduces the ACE model and section 5 illustrates its application to our data. Section 6 explores the robustness of these results when modifying the maintained assumption of absence of assortative mating among parents. Section 7 introduces the results on pseudo twins with respect to a larger set of outcomes, and section 8 concludes.

2. The data

In Italy, personal data is regulated under the general Data Protection Regulation (GDPR, Legislative Decree no. 101 of 10 August 2018) therefore sharing and linking personal information is subject to some restrictions. In particular, according to the Italian law on privacy protection,

[3] There are few papers studying the impact of (unexpected) twin birth on labour market participation: see Ponzo and Scoppa 2024, Barbiellini et al. 2023.

administrative information on earned incomes cannot be publicly revealed. Similarly, biological information on individuals (like being a fraternal twin of someone else) is also privacy protected. Our dataset originates from the merging of two files: the file of twins enrolled in the Italian Twin Register (ITR), run by the Istituto Superiore di Sanità-ISS, and the one on administrative data on payroll taxes from the Istituto Nazionale per la Previdenza Sociale-INPS. This was made possible thanks to an ad hoc agreement signed in January 2022 by the two partners.

The ISS is the official custodian of the Italian Twin Registry. The Italian Twin Registry, established in 2001, is a population-based voluntary registry of twins. At the very beginning, the ITR database consisted of a list of 'possible twin pairs' identified by the Ministry of Finance using the demographic information summarized in the fiscal code, a sort of personal identifier used by the tax system. That database, containing 650,000 'possible twin pairs' with the same surname, date and place of birth and born before the end of 1996 has been used as a starting point. Possible twins were compared to the actual new-born twin pairs recorded by the National Institute of Statistics (ISTAT) in terms of the number of same-sex and opposite-sex twins born between 1981 and 1995. According to ISTAT, 80.955 twin-pairs were born in this period, compared to 112.384 twin-pairs registered in the database, which represent an excess of 39%. This excess, however, steadily decreased over time, from 49% in 1981 to 12% in 1995. Moreover, the excess was lower for same-sex pairs (26% overall) than for opposite-sex pairs (68% overall) (Stazi et al 2002). The 'possible twin pairs' database was used until 2003 when other approaches for recruitment were adopted. Most of the twins are now recruited by applying a population-based strategy in several municipalities. Twins are selected among the residents according to the following criteria: same mother and father, same date and place of birth. Currently, a total of 29.000 twins are enrolled in the ITR: 11.500 monozygotic and 16.700 dizygotic resident throughout the country and belonging to a wide age range (from 0 to 95 years, mean 36.8 years) (Medda et al 2019).

The INPS collects payroll taxes from all workers (private and public employees, self-employed, contract workers), and therefore covers (almost) all source of earnings in the Italian population. The administrative data regarding work careers (in terms of employment and unemployment/layout spell, parental and illnesses leaves) are contained in contributory archives (*estratti conto*) where the data of the present paper come from.

Age cohorts were limited to subjects born between 1964 and 1996, in order to have most of them concluding their education (and possibly entering the labour market by 2021) while still active (and not yet retired). Since the merging procedure required preventive extraction of information from INPS administrative archives, we followed a similar strategy to the ITR creation and extracted individuals sharing family name, born in the same day in the same municipality. This population of

"pseudo-twins" overestimates the true-twin population, because of possible homonymy. In order to minimize such a risk, we have dropped all twinning with more than two individuals. After discarding observations with missing information in the pair, we started with a population of pseudo-twins of 344.226 individuals from INPS, against a potential number of 480.000 twins estimated from Census data over the same time period (assuming a share of 1% twins among new-born over the same period – see table 1). These were matched with 13.600 twins in the ISS registry[4], ending with a working sample of 9.722 twins who experienced at least one job spell in the sample period. The final sample size by year of birth is reported in the final column of Table 1, which also includes the potential population of Italian twins estimated from official statistics.

The match with ISS data differs by age. If we compute the ratio between the last column of table 1 (INPS-ISS matches) over the penultimate one (number of twins in the ISS registry), we observe an inverted U-shape pattern: for older persons born in the 60's the match covers around 70% of registered twins, possibly because a fraction of them has already retired. When we consider individuals in their forties, the coverage rate rises above 80%, then declining in the youngest cohorts to 60%.

Our strategy is exposed to potential biases, since an individual appears in our final sample if:

a) both twins have at least one record in INPS (e.g. a housewife without children and any work experience in the formal labour market throughout her life would not show up), and;

b) both have voluntary enrolled in the twin registry (with overtime changes in the probability of recruitment, as witnessed by the cohorts born in 1983-84-85).

Since both conditions are more likely to be satisfied as long as twins are more similar in terms of work attitudes with respect to randomly selected individuals, the matched file is exposed to the risk of overestimating the correlation in labour market experience among Italian twins. On the other side, the available information on pseudo-twins (which possibly includes false twins) may lead to an attenuation biased estimate of the actual correlation because we are unable to identify identical twins. Thus we have decided to proceed initially with the matched sample, and then to expand our analysis with the extended sample.

The merging of the data was regulated in the agreement, that imposed double hash encryption (in order to make it irreversible) and a further restriction of a minimal threshold of 10 for *n*-tuple replications. This imposes severe limitations in the number of usable variables and in their partitions. For example, we do observe yearly earnings and yearly working time as continuous variables, but we could not use that information because (especially in combination with demographics) it would have

[4] The original file from ISS consisted of 15.463 individuals, but 1.851 had missing information on the other twin, and 12 belonged to a triple twinning. After excluding these cases, we were left with 13.600 twins.

led to singleton observations in the data. We were therefore forced to summarise all the information available on earnings and working time into two variables: the quintile in the distribution of the decennial averages of the yearly income, and the quartile in the distribution of decennial averages of the time fraction spent into employment.[5]

The matched INPS-ISS file seems still representative of the original ISS file, as shown in Table 2, where we have reported descriptive statistics in terms of sex, age, education composition and zygosity. Female are over represented, possibly because they are more available to enrol in the ITR. The fraction of identical twins is around 45% in the twin registry, which is higher than the biological expectation of one third: this is explained by the voluntary enrolment in the registry, where identical twins are more inclined to participate. Looking at the geographical distribution, twins residing in Northern and Southern regions are in a smaller fraction than the total population (42% against 46% for Northern regions, 24% against 33% for Southern regions), thus suggesting that twins in Central regions are more likely to enrol, mainly due to the fact that the ITR is based in Rome and promotes more studies in this area.[6]

If the match does not seem to distort excessively the distribution of characteristics in the ISS file, things are different when comparing the matched file with the pseudo-twins file obtained from INPS administrative data, which in principle offers a better description of the entire population. Looking at Table 3 one can notice that our matched file contains a younger population, where women and the less educated are overrepresented.[7] We do not possess zygosity information in the INPS data, but we can identify (pseudo) fraternal twins in the couple of different sex: these are one fourth in the matched file, but one third in the larger INPS file (as expected from biology form Weinberg's Differential Rule – Fellman 2013).

When we consider income positions, we can notice that individuals are correctly distributed in both files for the youngest cohorts, until the age of 40. Afterwards, matched twins obtain higher earnings than the average pseudo-twins, partly because they work more weeks than average. This is the reflection of non-random attrition in the ISS file combined with non-uniform distribution over different ages: for this reason we will partition the matched population into three age groups (born

[5] More precisely, we collapsed individual information on earnings and contributed weeks at yearly frequency, converting nominal values into real ones using the consumer price index. We then averaged yearly information into decennial values, computed over age intervals of the individual (less than 30, 31-40, 41-50 and above 50). The percentile position was then associated to each individual according to age intervals.

[6] A probit model for being matched using the same variables as regressor confirms negative correlation with female and positive with education and age; statistical significance is also found for few regions (Emilia Romagna, Lombardy Piedmont, Tuscany and Sicily). Available from the authors.

[7] For the ISS file, the years of education are obtained converting the maximal educational attainment (5 years for primary, 8 years for lower secondary, 12 or 13 years for upper secondary and 16 or 18 for college degree, depending on whether the information was collected after or before the age of 25). For the INPS file, the years of education are proxied by the age of first job – 6 (the start of compulsory education).

before 1983, i.e. older than 39; born from 1983 to 1985, i.e. aged between 37 and 39; born after 1985, that is younger than 37) in order to reflect the different phases in recruitment into the twin registry (see table 1). Eventually also notice the non-uniform distribution of worked weeks, with a mass concentration in the top quartiles where workers are employed full-time and full-year. Descriptive statistics by zygosity and age groups are reported in table A1 in the Appendix.

3. Empirical correlations

We consider three main outcomes: years of schooling, income rank (averaged over four observations of individual quintile computed over decennial averages of earnings) and workdays rank (averaged over four observations of individual quartiles computed over decennial averages of worked weeks). Given the longer observation span of older individuals in comparison to younger ones, older cohorts are characterized by greater precision in income/work measures, but they are exposed to the risk of non-random attrition in administrative data on work, due to early retirement, illnesses and loss of jobs.

In Table 4 we report the twin correlations in outcomes. Not surprisingly, outcome correlation among identical twins is higher than among fraternal twins, especially when they are male. Correspondingly, the correlation is the lowest among different sex fraternal twins. Given potential idiosyncratic differences in labour market participation between men and women that could induce lower correlations in labour market outcomes for pairs that include females, we control for sex throughout our analysis.

Given the uneven distribution of our sample over birth cohorts, with a mass of cases concentrated in the three years 1983-1985, we have partitioned the data into three groups: the *young* (born after 1985 – corresponding to 33% of the sample), the *adult* (born between 1983 and 1985 – corresponding to 41% of the sample) and the *old* (born before 1983 – corresponding to 26% of the sample), and we will conduct our investigation on each group separately. These three groups have presumably experienced different economic and institutional settings when entering the labour market. The oldest group is mainly composed by cohorts who completed education before the wave of reforms aimed at increasing employment flexibility hit the Italian labour market between the late 1990s and the early 2000s. The intermediate group, on the other hand, has entered the labour market right after some major reforms had come into effect (Pacchetto Treu in 1997 and Biagi Law in 2003). Finally, the youngest group entered the labour market after labour flexibility had been fully implemented and amid the financial crisis of 2008. In the bottom part of Table 4 we report the twin correlation in outcomes, partitioning the sample by age groups. Correlation halves when passing from

the youngest to the oldest group. While this is expected when looking at labour market experience, due to the idiosyncratic components, it is more surprising when looking at years of education. This could signal that educational choices have changed when moving towards mass education, as experienced by the cohorts born in the nineties and later.

4. The ACE model

For each cohort we have estimated the so-called the ACE model that is popular in behavioural genetics. The ACE posits that outcomes (or phenotypes) are the result of three orthogonal and linearly additive factors:

$$Y_i = A_i + C_{f(i)} + E_i$$

where i is the individual and $f(i)$ denotes her family, Y is the outcome, A is an additive genetic effect, C is a common environmental effect shared by family members, and E is an idiosyncratic effect unique to person i. Each component $x = (A, C, E)$ is drawn from a zero-mean distribution with variance σ_x^2.

This basic formulation of the ACE model rests on several assumptions. Orthogonality of the three factors rules out the possibility of gene-environment correlation, i.e. individuals or families do not sort into environments on the basis of their genes.[8] The linear specification excludes the possibility of gene-environment interactions, a circumstance in which the environment mediates genetic expressions.[9] A third underlying assumption of the ACE is that spouses are not sorted on genes, implying that DZ twins share on average half of their genes, while genetic assortative mating would imply a larger sharing for DZ's.[10] A fourth assumption is that there is no dominance of the gene variants someone receives from one parent on the variants received from the other parent.[11] Finally, the model assumes that the extent of environmental sharing is the same for MZ and DZ twins.[12]

[8] Evidence supporting the gene-environment orthogonality assumption is provided by a number of studies (e.g., Björklund, Jäntti, and Solon, 2005; Fagereng, Mogstad, and Rønning, 2021; Biroli et al., 2022; Collado, Ortuño-Ortín, and Stuhler, 2023).

[9] Studies from molecular genetics that leverage polygenic score information tend to find significant effects by interacting the scores with measures of environmental exposure, see e.g. Biroli et al. (2022).

[10] Existing estimates from polygenic scores of educational attainment indicate that the extent of spousal sorting on genes is at best mild, with a sorting correlation of 0.18 (see Okbay et al. 2022).

[11] Cesarini et al. (2009a) provide evidence that supports this assumption in an ACE model of risk taking.

[12] Bingley, Cappellari and Tatsiramos (2024) relax this assumption in a twin family model of socio-economic outcomes, showing that environmental sharing is stronger for MZ twins and that the canonical ACE with common environment tends to overestimate the contribution of the genetic component to the dispersion of outcomes.

Under this set of assumptions, the model provides sufficient information for the identification of the three variance components σ_A^2, σ_C^2 and σ_E^2. Namely, the outcome variance is:

$$var(Y) = \sigma_A^2 + \sigma_C^2 + \sigma_E^2;$$

the MZ twins covariance is

$$cov(YY')^{MZ} = \sigma_A^2 + \sigma_C^2 \ ;$$

and the DZ twins covariance is

$$cov(YY')^{DZ} = 0.5\sigma_A^2 + \sigma_C^2 \ ;$$

These are three equations in three unknown parameters, which are therefore identified.[13] For example the genetic component σ_A^2 is identified as twice the difference between the MZ and DZ covariances, while the environmental component is identified by subtracting the genetic component from the MZ twins' covariance. The model parameters can be used to compute the degree of heritability, that is the share of cross-sectional dispersion accounted for by the genetic component $\sigma_A^2/(\sigma_A^2 + \sigma_C^2 + \sigma_E^2)$.

To estimate the model we assume normality of the factors and use a Mixed-Model approach.[14] This is essentially a two-equations SURE (one for each twin) in which the moment conditions implied by the model are imposed on the variance-covariance matrix of the errors. Whenever the estimated shared environment was negligible either statistically or substantively, we followed much of the practice in behavioural genetics and turned to a restricted version of the ACE model, the AE model, that constrains the shared environmental component to be equal to zero.[15]

5. Results

Our main results are reported in Table 5 in terms of shares of cross-sectional variance that can be ascribed to each of the three (or two, for AE models) factors.[16] The estimating equations always include sex as controls. The estimated heritability in educational attainment fluctuates over cohorts. For the youngest cohort we estimate heritability at 44%, which is in line with the findings of Silventoinen et al. (2020) who apply the ACE on a dataset of 28 countries. For the other cohorts our

[13] Alternatively, one could assume a standardized distribution of the outcome with $var(Y) = 1$, such as there would be two equations for the twins correlations and two unknowns, that is the shares of cross-sectional variance due to genes and shared environment.

[14] The assumption of normality in the distribution of the outcome variables may be questioned. We have explored semi-parametric approaches based on GMM, but in few instances we experienced convergence issues. When we reached convergence results tend to slightly depart from Mixed-Model results, in particular they display lower heritability. Similarly lower GMM estimates of heritability are reported in Bingley, Cappellari and Tatsiramos (2024).

[15] For further details on the behavioral genetic foundations of the ACE and AE variance decomposition methods see Neale (2009).

[16] Shares are computed from the estimated variance components, that are available on request.

estimates are larger, especially for the intermediate one.[17] The degree of heritability for all cohorts pooled together is equal to 47%. Also, the intermediate cohort displays a notably low degree of shared environmental influences. Finally, idiosyncratic effects account for between 20% and 30% of the cross-sectional variation of educational attainment depending on the cohort.

Moving to labour market outcomes in the lower panels of Table 5, for the oldest cohort there is a greater share of inequality that can be attributed to idiosyncratic factors compared to education, and symmetrically a lower share due to genetics, while the impact of shared environment remains stable. For younger cohorts we see instead that shared environment does not contribute for labour market inequality, which is instead explained in equal proportions by the genetic (49%) and individual components (51%) in the case of earnings, and for the majority accounted for by individual variation in the case of working time (54 and 55%). Estimates for the pooled sample confirm a negligible role for shared environment and an equally important influence from genetics and idiosyncratic environment.

Noteworthy, labour market outcomes are not strictly comparable across cohorts, as they are not observed on the full life-cycle for younger cohorts. Table 6 shows that this asymmetry in the available data does not explain cross-cohorts differences: the oldest cohort is characterized by a greater impact of shared environment and a lower impact of the genetic component compared to younger cohorts also if we limit the observation to outcomes measured soon after labour market entry (i.e. age 20-30), which are available for all cohorts.

As already mentioned, young cohorts in our data entered the labour market after the introduction of a set of reforms aimed at increasing flexibility. The Italian labour market was significantly modified around those years, reducing the employment protection on temporary workers (the OECD EPL index declined from 4.75 over 6.00 in 1997 to 2.00 in 2003). The main reforms took place in 1997 (Treu reform) and in 2003 (Biagi reform), and both were based on expanding the variety of labour market contracts in order to encourage short term work opportunities, similar to what has happened in Germany with the Hertz reform and the "one euro" jobs. The difference in the labour market legislation experienced by the *young-adult* group on one side and by the *old* group on the other side may constitute a labour demand explanation of the different relevance of family environment shown in Table 6.[18] Our results show that this more flexible institutional environment

[17] In the matched data the share of identical twins increases with age, possibly because non-random attrition in the national Registry of Twins: it is 39% in the youngest group, 45 % in the adult group and 54% in the oldest group. We note that while this share decreases monotonically from the oldest to the youngest group, heritability follows a hump-shaped pattern, which is not consistent with the possibility that the oversampling of identical twins drives our estimates.

[18] Rosolia and Torrini (2016) find that entry wages started to decline around the mid-1990s. They argue that this pattern cannot be explained by changes in observable job characteristics. In addition, falling entry wages have not been accompanied by faster subsequent career paths; rather, subsequent career paths have increasingly featured rising earnings dispersion due to both increased workers heterogeneity (consistent with a greater weight of the idiosyncratic component

is associated with a greater relevance of genetic vs shared environmental determinants of inequality. This relevance suggests that labour flexibility has amplified the impact of inherent abilities and traits, partly determined by genetics, on labour market success and career progression.

There may also be other explanations though. A cohort born at the turn of the year 1982 entered primary school in 1988, middle school in 1993, upper secondary school in 1996 and (if not earlier) tertiary education or labour market in 2001. The year 1999 is characterized by an important reform in tertiary education, namely the start of the Bologna process, separating 3-year BA courses from an additional 2-year MA courses. There is evidence that this reform has made college access less dependent on parental background (Di Pietro and Cutillo 2008) which may explain the reduced relevance of the shared environment of younger cohorts. This would represent a labour supply explanation of our findings, but we should bear in mind that such an explanation would apply mainly to college graduates, who represent a minority of the population under study.

We have further explored potential regional differences in these patterns. We were expecting that environmental components being more relevant in more traditional societies, like the Southern ones. However, looking at Table 7, we notice that this is not the case, because we do not find a geographical gradient on educational attainment, nor one associated to the age. Sample size tends to become relevant, since missing information on education plague data for younger twins, while the same applies for employment among older cohorts. In table 8 we have considered labour market outcomes. The environmental component is hardly identified and therefore we have pursued the AE decomposition between genetics and idiosyncratic dimensions. Nevertheless, standard errors are large enough to prevent a consistent ranking across macro regions, leading us to the conclusion that there is no detectable geographical heterogeneity in our sample.[19]

6. Robustness

As discussed, one of the assumptions underlying the ACE model is that of absence of genetic assortative mating, implying that couples are formed by partners whose genes are drawn randomly from the population. Because of this, the extent of genetic similarity of fraternal twins (or, indeed, non-twin siblings) is 50% on average. On the other hand, if the genes of spouses are correlated, the

in twins correlation) and greater temporary earnings instability. They relate such developments to the changes in labour market institutions that took place between the early 1990s and the mid-2000s. Further evidence in this direction is provided by Bianchi and Paradisi (2023) who argue that productivity slow down coupled with pension reforms delaying retirement have limited the career opportunities of more recent cohorts of entrants.

[19] This contrasts what found in other studies, where Southern regions were characterized by lower equality of opportunities given the larger variance accounted by social origins (Checchi and Peragine 2010).

probability that fraternal twins (or non-twin siblings) share their genes increases by the extent of the spousal correlation. Assortative mating, instead, does not affect the genetic resemblance of identical twins, which is always 100%. Assortative mating implies that in the ACE model the genetic component is larger than double the difference between the outcome covariances of identical and fraternal twins. Therefore, failure to account for assortative mating in the ACE model will induce a downward bias in the estimate of the genetic component and an upward bias in the shared environmental component, without affecting the idiosyncratic component that inversely depends on the *sum* of genes and shared environment. The extent of assortative mating is not identified in the ACE. Bingley et al. (2024) leverage data on twins and twins' spouses to overcome the under-identification and estimate a spousal correlation in genes of 0.12, implying that any bias deriving from the omission of assortative mating is modest at best.[20] To address the robustness of our results to the presence of assortative mating in genes, we estimate a version of the ACE model in which we impose a genetic sharing larger than 50% for fraternal twins. We experimented with shares of 55%, 60% and 65%, which is equivalent to calibrate the spousal correlation in genes to 5%, 10% and 15% respectively. We conduct the exercise using years of education as outcome, as the calibrated model did not find convergence on the other outcomes. Results are presented in Table 9, alongside baseline estimates (no genetic spousal correlation, column 1) and indeed show that allowing for assortative mating increase the share of variance imputed to genes, and conversely reduces the impact of shared environment. When we calibrate assortative mating to 5%, the share of genetic variance increases to 52% (compared to 47% in the baseline), while the impact of shared environment declines to 21% (while it is 27% in the baseline). Raising the calibration of assortative mating to 10% further increases the genetic share to 58% and reduces shared environment to 15%. In the last column of the table, which corresponds to a calibrated assortative mating of 15%, close to the upper bound of the estimates available from the molecular genetics literature, we lose precision and estimate shared environment to 6.5% but not statistically different from zero. We take this evidence as indicating that the last calibration is not supported by the data. Overall, this robustness exercise seems to indicate that the omission of assortative mating imparts a modest bias to the estimated ACE decomposition.

The ACE model assumes that outcomes depend on the three latent factors (genetics, shared environment, idiosyncratic environment) whose dispersion can be estimated leveraging twins covariances. Still, not all the relevant determinants of the outcomes of interest are unobservable in our data, and actually some predetermined characteristics that we observe like year and region of birth may well interact with the three ACE factors, and that are likely to reflect the impact of

[20] Molecular genetics studies estimate assortative mating in genes in the range 0.15-0.18, see Okbay et al. (2022).

environmental (rather than genetic) influences. To the extent that these observable characteristics are not orthogonal to ACE factors, we would expect variance decompositions to change when we residualise outcomes on the observables. Table 10 reports variance decompositions obtained after netting out the impact of year and region of birth (in addition to sex that is always controlled for) from the outcomes before performing the variance decomposition. Results for education in panel A go in the expected direction: controlling for year and region of birth reduces the proportion of variance that is attributed to shared environment, while the shares accruing to genetics and idiosyncratic factors increase. This is expected as long as year and region of birth are expressions of shared environments. Moving to labour market outcomes in panels B and C, we report estimates from the restricted AE specification, that assumes that all environmental influences are purely idiosyncratic, because the model with full ACE specification did not converge. This lack of convergence of the model with shared environment suggests that year and region of birth represent a relevant source of shared environmental influences in the process that determines labour market outcomes. This was not the case with years of education, which may depend on common environments (such as schools) whose characteristics are dispersed even with year-region of birth cells. The AE results in panels B and C point towards an increased relevance of idiosyncratic factors after controlling for year and region of birth, and symmetrically a reduction of genetic influences. Overall, while controlling for predetermined observable affects the ACE or AE variance decompositions, their impact is at best limited.

7. Further evidence from pseudo-twins

We now turn to the analysis of the pseudo-twins data described in Section 2. The administrative archives from which these data originate in principle cover the entire population of Italian twins with any pension contributions. However, because pseudo-twin pairs were identified by matching the surname, date and place of birth available in the tax code, these data will likely underestimate twin correlations in outcomes due to homonymy of unrelated individuals. The advantage of using these data is that they allow considering a range of outcomes that is not available in the INPS-ISS twin sample, since they were excluded to minimize the risk of reidentification. Pseudo-twins can be studied in their *labour market attachment* through unemployment or absenteeism (respectively proxied by events of unemployment subsidy or illnesses spells), in their *fertility decisions* (proxied by use of parental leaves),[21] in their *prosocial attitudes* (captured by event of blood donations, since Italian

[21] One should recall that parental leaves for mothers include 2 months before and up to 6 months after the child birth, while parental leaves for father reach 10 days per birth. Thus this correlation involves almost only mothers. The correlations in table 12 are unconditional, while the decomposition in table 13 control for gender.

workers are entitled to one day off in such event) and in their *religious vocations* (proxied by contribution in the clergy pension fund). Another advantage of these data is that they provide an even coverage over birth cohorts with large sample sizes thereby enabling a better reconstruction of life-cycle pattern compared to the data in the twins sample. This additional information is summarized in Table 11. Despite the unavoidable difference in sample size over the life cycle (only individuals from older birth cohorts can be observed at older ages), the means of our variable seem consistent with a life-cycle profile. The risk of unemployment increases with age, as does the absences for illnesses (even though serious illnesses lead to exit from the labour market), fertility is highest in the 30-40 age range, blood donation exhibits an inverted U-shaped and religious vocations remain stable over the life cycle.

Table 12 reports the outcome correlations for pseudo-twins. The first three columns focus on the outcomes that are available also in the twin sample (education, earnings and working time) such as we can benchmark pseudo twins correlations on the correlations of real twins. We distinguish between same-sex and mixed-sex pairs. Because mixed-sex twins are DZ, the correlations estimated on mixed-sex pseudo twins are directly comparable to those of mixed-sex DZ twins. Such direct comparability does not apply to same-sex pseudo twins, as same-sex twins could be both MZ or DZ. Looking at years of education for mixed-sex pairs, the estimated correlation is 0.30, remarkably close to the one estimated on mixed-sex DZ twins (0.34). The discrepancy with mixed-sex DZ twins seems instead larger looking at (permanent) earnings and working time, 0.15 and 0.13 vs 0.21 and 0.23 (respectively) for mixed-sex DZ twins. For same-sex pairs, the pseudo-twins correlation lies between the MZ and DZ estimates in the case of education and earnings, whereas for working time it is close to the DZ twins estimate.

Looking now at the additional outcome variables that are available in the pseudo-twin sample (unemployment, illness absence, take-up of parental leave, blood donation and clergy membership), we see that same-sex correlations are relatively sizeable and always larger than mixed-sex ones. The greatest differences are observed for fertility and pro-social behaviour. Even the probability of a religious vocation is shared among pairs of (potentially identical) pseudo twins.

In the absence of zygosity information in the pseudo-twins sample, we rely on the sex composition of the pair to derive the ACE decomposition of the covariances. To do this we first notice that on average one third of twin births is accounted for by MZ twins (who are all same-sex pairs), one third is made by same-sex DZ twins and the remaining third is given by mixed-sex DZ twins. Therefore, a same-sex (SS) pair of pseudo twins will be composed by MZ twins and DZ twins in equal proportions, such as the ACE covariance decomposition for this pair is given by:

$$cov(YY')^{SS} = 0.75\sigma_A^2 + \sigma_C^2 \ .$$

On the other hand, mixed-sex (MS) pairs will be entirely composed of DZ twins and therefore

$$cov(YY')^{MS} = cov(YY')^{DZ} = 0.5\sigma_A^2 + \sigma_C^2$$

As already mentioned, pairs in the pseudo-twins sample will include spurious twins, such as the estimates of the genetic variance component and shared environmental component will likely underestimate the ones that would be obtained in a sample of genuine twins, while the opposite holds for the idiosyncratic component.

We report the variance decomposition estimated from the pseudo-twins data in Table 13. For education and employment[22], the ACE model algorithm did not reach convergence and therefore we report the corresponding AE estimate. Estimates for education indicate that heritability accounts for a little more than half of the variance in years of education, while the rest is due to purely idiosyncratic variation. It should be noted that idiosyncratic factor explained less than 30% of educational dispersion in the actual twins data, confirming that the pseudo-twins approach tends to overestimate idiosyncratic variation and to underestimate the inequality that comes from shared factors.

Moving to labour market outcomes in Table 13, we notice consistent patterns. As was the case with twins, the share of inequality accounted for by idiosyncratic factors increases comparing labour market outcomes with education, but the pseudo-twins figures are larger than those of twins, being between 60% and 65% for pseudo-twins and between 50% and 60% for twins.

In Table 14 we consider the decomposition of pseudo-twins correlations in labour market outcomes by birth cohorts. For permanent earnings the pattern is analogous to the one found in the twins data, that is shared family environment tends to lose relevance among younger cohorts exposed to more flexible labour markets. For working time, instead we cannot report estimates of the shared environmental component due to lack of converge.

The evidence above suggests that pseudo-twins data are broadly comparable to twins ones and that in the former case variance decompositions tend to put more weight on idiosyncratic components and a lower weight on shared components. On these bases we now turn to investigate the additional outcomes that are observable for pseudo-twins. Decomposition results are reported in Table 15. We use an AE specification throughout the table due to lack of convergence of the ACE specification on these outcomes. For the same reason, we set the weight of the genetic variance component equal to unity for same-sex pseudo-twins in the case of pro-sociality and religious vocation. We find that the genetic component exhibits a declining weight as long as we move towards responsible individual choices: it is as high as 25-30% in the case of unemployment or sick leaves, but it goes to 20% in the

[22]In the administrative data there is no information on educational attainment, which is then proxied by the age of first job – 6 (age of start of compulsory education), capped at 25.

case of blood donations or to 13% in case of fertility, approaching zero in the case of religious vocation.

8. Conclusions

The present paper provides for the first time estimates of heritability in education and labour market outcomes using administrative data for Italy. Using both correlation analysis and the ACE model, we show that the genetic component matters, contributing to a large fraction of inequality in education and labour market outcomes, in a range comprised between 30 to 50%. For our age cohorts, the common environment component matters for education (in a range between one third and one fourth of total variance), but less and less for labour market outcomes in young cohorts.

We asked ourselves about the potential explanation of the disappearance of the shared environment component, paralleled by an increase of the genetic component. Since the younger cohorts (entering the labour market at the end of last century) have experiences a more flexible labour market, it is possible that labour market flexibility has amplified idiosyncratic dimensions (that could also be termed "unobservable ability") like behavioural traits, psychological traits and the like, that are evidently correlated among twins. However, sample size and representativeness prevent us from providing more stringent tests using a diff-in-diff strategy.

One could also read the decline in the shared component with a positive eye, since it would correspond to a reduced inequality of opportunities. Conversely, it may be questioned whether genes are to be counted as circumstances outside individual responsibility. In case of positive answer, then what would matter in terms of inequality of opportunity would be the inequality explained by A+C components, which in our data tend to remain stable. Other studies using different data argues against a decline in measure inequality of opportunities in the Italian case. Thus we leave this as an open question, requiring more data to apply cohort analysis to twins outcomes.

The paper also explores the possibility of extending twin analysis to larger datasets of pseudo-twins (i.e. individuals sharing part of the socials security code containing information on date and place of birth and three letters of the family name). We find that pseudo-twins results mimic real-twins ones at a large extent, even though we are unable to say whether these differences are to be attributed to sample selection in the real-twins data (where enrolment is voluntary) and/or in the pseudo-twins data (where homonymy may inflate the relevant population), since by construction it is impossible to link the two samples. The richness of administrative information on pseudo-twins allows us to extend the ACE model to other social outcomes. We find that the genetic component exhibits a declining weight

as long as we consider responsible individual choices, from absenteeism (25-30%) to blood donations (20%) or fertility (13%).

References

Barbiellini Amidei, Federico, Sabrina Di Addario, Matteo Gomellini and Paolo Piselli. (2023). Female labour force participation and fertility in Italian history. Centro Studi Luca D'Agliano Development Studies Working Papers N. 484, February

Barnea, A., H. Cronqvist and S. Siegel (2010). Nature or Nurture: What Determines Investor Behavior*? Journal of Financial Economics*, 98, 583-604.

Behrman, J.R. and P. Taubman (1976). Intergenerational Transmission of Income and Wealth. *American Economic Review*, 66(2), 436-440.

Behrman, J.R. and P. Taubman (1989). Mostly in the Genes? Nature-Nurture Decomposition Using Data on Relatives. *Journal of Political Economy*, 97(6), 1425-1446.

Bianchi, N. and M.Paradisi. (2023). Countries for Old Men: An Analysis of the Age Wage Gap. mimeo

Bingley, Paul, Lorenzo Cappellari, and Konstantinos Tatsiramos. (2024). On the Origins of Socio-Economic Inequalities: Evidence from Twin Families. LISER Working Paper 2024-03.

Biroli, P., T.J. Galama, S. von Hinke, H. van Kippersluis, C.A. Rietveld, and K. Thom (2022). The Economics and Econometrics of Gene-Environment Interplay, *Tinbergen Institute Discussion Papers*, 2022-019/V.

Björklund, A., M. Jäntti and G. Solon (2005). Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types. In: Bowles, S., Gintis, H., Groves, M.O. (Eds.), *Unequal Chances: Family Background and Economic Success*. Princeton University Press, Princeton, 145-164.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese. (2013). Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Social Forces* 92(1) 109–140, September doi: 10.1093/sf/sot076

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein and B. Wallace (2009a). Genetic Variation in Preferences for Giving and Risk Taking. *Quarterly Journal of Economics*, 124(2), 809-842.

Cesarini, D., M. Johannesson, P. Lichtenstein and B. Wallace (2009b). Heritability of Overconfidence. *Journal of the European Economic Association*, 7(2-3), 617-627.

Checchi, D. and V. Peragine, (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, vol. 8(4), pp 429-450

Collado, M.D., I. Ortuño-Ortín and J. Stuhler (2023). Estimating Intergenerational and Assortative Processes in Extended Family Data. *Review of Economic Studies*, 90(3), 1195-1227.

Cronqvist, H. and S. Siegel (2015). The Origins of Savings Behavior. *Journal of Political Economy*, 123(1), 123-169.

Di Pietro, G. and Cutillo, A. (2008). Degree Flexibility and University Drop-out: The Italian Experience, *Economics of Education Review*, 27(5), 546-555.

Fagereng, A., M. Mogstad and M. Rønning (2021). Why Do Wealthy Parents Have Wealthy Children? *Journal of Political Economy*, 129(3), 703-756.

Fellman J. (2013). Statistical analyses of monozygotic and dizygotic twinning rates. *Twin Research and Human Genetics* 16: pp 1107–1111.

Goldberger, A. S. (1979). Heritability. *Economica*, 46(184), 327–347.

Guo, G., and J. Wang. (2002). The mixed or multilevel model for behaviour genetic analysis. *Behavior Genetics* 32/1, pp 37-49.

Hyytinen, A., Ilmakunnas, P., Johansson, E., Toivanen, O. (2019). Heritability of lifetime earnings, *The Journal of Economic Inequality*, 17, 319–335

Lang, V. (2017). ACELONG: Stata module to fit multilevel mixed-effects ACE, AE and ADE variance decomposition models. Statistical Software Components S458402, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s458402.html

Maczulskij, T. (2013) Employment sector and pay gaps: Genetic and environmental influences, *Labour Economics*, 23, 89-96.

Manski, C. F. (2011). Genes, Eyeglasses, and Social Policy. *Journal of Economic Perspectives*, 25(4), 83-94.

Medda E., Toccaceli V., Fagnani C., Nisticò L., Brescianini S., Salemi M., Ferri M., D'Ippolito C., Alviti S., Arnofi A., Stazi M.A. (2019) The Italian twin registry: an update at 18 years from its inception. *Twin Research and Human Genetics* 22(6): 572–578.

Neale, M. C. (2009). Biometrical models in behavioral genetics. In: Y.-K. Kim. *Handbook of Behavior Genetics*. New York: Springer, pp 15-33.

Papageorge, N.W and Thom, K (2019): Genes, education, and labor market outcomes: Evidence from the Health and Retirement Study, *Journal of the European Economic Association*, 18, 1351–1399

Ponzo, Michela and Vincenzo Scoppa (2024). Human Capital Investments and Family Size in Italy: IV Estimates Using Twin Births as an Instrument. *BE J. Econ. Anal. Policy* 24(2): 425–46

Okbay A., Wu Y., Wang N., Jayashankar H., Bennett M., Nehzati S.M., …, Young A.I. (2022). Polygenic Prediction of Educational Attainment Within and Between Families from Genome-Wide Association Analyses in 3 Million Individuals. *Nature Genetics*, 1–13.

Rabe-Hesketh, S., A. Skrondal, and H. K. Gjessing. (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics* 64/1, pp 280–288.

Rietveld, C.A. (2024). Heritability and Public Policy Reconsidered, Again. *Tinbergen Institute Discussion Papers*, 2024-012/V.

Rosolia, A. and R.Torrini. (2016). The generation gap: a cohort analysis of earnings levels, dispersion and initial labour market conditions in Italy. *Questioni di Economia e Finanza Occasional papers* Number 366 – November

Sacerdote, B. (2007). How Large Are the Effects of Changes in Family Environment? A Study of Korean American Adoptees. *Quarterly Journal of Economics*, 122(1), 119-157.

Silventoinen, Karri, Aline Jelenkovic, Reijo Sund, Antti Latvala, Chika Honda, Fujio Inui, Rie Tomizawa, et al. (2020). Genetic and Environmental Variation in Educational Attainment: An Individual-Based Analysis of 28 Twin Cohorts. *Scientific Reports* 10 (1): 12681.

Stazi, M. A., Cotichini, R., Patriarca, V., Brescianini, S., Fagnani, C., D'Ippolito, C., Cannoni, S., Ristori, G., & Salvetti, M. (2002). The Italian twin project: From the personal identification number to a national twin registry. *Twin Research and Human Genetics*, 5, 382–386

Table 1 – Sample size

birth year	age in 2022	new born (ISTAT)	estimated twins	pseudo twins (INPS)	real twins (ISS)	matched INPS-ISS
1964	58	1 016 120	20 528	11 830	150	105
1965	57	990 458	20 009	11 976	188	134
1966	56	979 940	19 797	12 178	200	145
1967	55	948 772	19 167	12 384	172	120
1968	54	930 172	18 791	12 030	166	121
1969	53	932 466	18 838	12 670	173	141
1970	52	901 472	18 212	11 956	192	150
1971	51	906 182	18 307	12 864	216	165
1972	50	888 203	17 943	12 490	200	144
1973	49	874 546	17 668	13 020	160	123
1974	48	868 882	17 553	13 524	214	165
1975	47	827 852	16 724	13 356	178	143
1976	46	781 638	15 791	12 486	184	149
1977	45	741 103	14 972	11 950	142	106
1978	44	709 043	14 324	11 256	164	116
1979	43	670 221	13 540	10 398	153	111
1980	42	640 401	12 937	10 182	168	142
1981	41	623 103	12 588	10 200	152	115
1982	40	619 097	12 507	9 808	200	142
1983	39	601 928	12 160	9 622	1592	1193
1984	38	587 871	11 876	9 104	2314	1750
1985	37	577 345	11 664	9 280	1710	1272
1986	36	555 445	11 221	8 916	578	441
1987	35	551 539	11 142	8 786	266	202
1988	34	569 698	11 509	9 208	322	246
1989	33	560 688	11 327	9 098	338	250
1990	32	569 255	11 500	8 996	314	232
1991	31	562 787	11 369	8 834	342	258
1992	30	567 841	11 472	8 540	424	279
1993	29	549 484	11 101	7 782	546	328
1994	28	533 050	10 769	7 264	622	361
1995	27	525 609	10 618	6 378	456	227
1996	26	528 103	10 669	5 838	404	146
Total		**23 690 314**	**478 592**	**344204**	**13 600**	**9 722**

Source: New born: nati vivi (legittimi+naturali) from Istat Serie Storiche - Tavola 2.5.1 - Nati vivi e nati morti per filiazione e sesso - Anni 1926-2014

Estimated twins: assuming 1% of deliveries, computed as $=2 \times \frac{0.01}{0.99}$

Pseudo twins: from INPS archives, individuals sharing family name, date and place of birth, with active records.

Real twins: from ISS Twin registry

Table 2 – Descriptive statistics for ISS and matched INPS-ISS files

	(1) Match INPS-ISS			(2) Actual twins in ISS registry		
	(a) mean	(b) sd	(c) obs	(a) mean	(b) sd	(c) obs
Demographics						
female	0.57	0.50	9 722	0.57	0.49	13 600
age	38.81	7.47	9 722	38.27	7.74	13 600
years of education	12.54	2.73	7 501	12.81	2.75	10 068
age of leaving family	22.67	4.80	7 006	22.58	4.78	9 403
	% fraction		obs	% fraction		obs
Maximal educational attainment						
primary	0.73		55	0.92		93
lower secondary	12.00		900	13.32		1 341
upper secondary (declared after 26)	16.38		1 229	16.10		1 621
upper secondary (declared before 25)	55.30		4 148	54.82		5 519
college degree (declared before 25)	1.39		104	1.48		149
college degree (declared after 26)	14.20		1 065	13.36		1 345
total	**100.00**		**7 501**	**100.00**		**10 068**
Zygosity						
identical twins (monozygote)	46.60		4 530	45.63		6 206
fraternal twins (dizygote) same sex	29.64		2 882	29.96		4 074
fraternal twins (dizygote) different sex	23.76		2 310	24.41		3 320
total	**100.00**		**9 722**	**100.00**		**13 600**
Region of residence						
Piedmont	5.95		578	5.54		754
Valle d'Aosta	0.13		13	0.19		26
Lombardy	22.76		2 213	22.03		2 996
Liguria	0.96		93	0.91		124
Veneto	7.81		759	7.63		1 038
Trentino-Alto Adige	0.97		94	1.00		136
Friuli-Venezia Giulia	4.08		397	4.10		558
Emilia-Romagna	3.95		384	3.40		462
Tuscany	3.11		302	2.84		386
Umbria	1.99		193	1.84		250
Marche	1.59		155	1.53		208
Abruzzo	1.05		102	1.05		143
Molise	0.57		55	0.54		74
Lazio	20.46		1 989	19.63		2 670
Campania	6.88		669	7.51		1 022
Calabria	3.44		334	3.81		518
Basilicata	0.53		52	0.57		78
Apulia	5.41		526	5.63		766
Sicily	6.70		651	7.87		1 070
Sardinia	1.68		163	1.69		230
Abroad				0.67		91
total	**100.00**		**9 722**	**100.00**		**13 600**

Table 3 – Descriptive statistics for INPS and matched INPS-ISS files

	(1) Match INPS-ISS			(2) Pseudo twins (INPS)		
	(a) Mean	(b) sd	(c) obs	(a) Mean	(b) sd	(c) obs
Demographics						
female	0.57	0.50	9 722	0.457	0.498	344 204
age	38.81	7.47	9 722	43.695	9.147	344 204
years of education	12.54	2.73	7 501	15.862	4.510	343 267
age of leaving family	22.67	4.80	7 006			
age of first job				22.14	5.28	343 267
	% fraction		obs			
Zygosity						
identical twins (monozygote)	46.60		4530	63.30		217 878
fraternal twins (dizygote) same sex	29.64		2882			
fraternal twins (dizygote) different sex	23.76		2310	36.70		126 326
total	**100.00**		**9722**	**100.00**		**344 204**
	Age < 30	Age 31-40	Age 41-50	Age > 50	Age < 30	Age 31-40	Age 41-50	Age > 50
Income position (%)								
first quintile	22.40	16.12	13.13	13.40	21.58	19.17	20.56	20.00
second quintile	17.15	19.66	14.15	13.29	18.53	21.09	18.65	20.56
third quintile	20.29	22.40	19.89	15.65	21.87	21.65	21.86	20.45
fourth quintile	17.61	19.98	21.45	22.75	19.14	18.93	18.57	18.92
fifth quintile	22.55	21.85	31.38	34.91	18.89	19.17	20.35	20.06
total	**9 323**	**8 064**	**2 247**	**888**	**324 829**	**289 170**	**199 372**	**78 203**
Worked time position (%)								
first quartile	32.08	24.69	20.65	18.92	25.15	24.78	24.54	24.36
second quartile	26.05	30.85	24.34	22.07	26.62	27.76	26.72	27.77
third quartile	21.97	44.46	55.01	28.94	24.76	47.46	48.74	23.71
fourth quartile	19.90			30.07	23.46			24.16
total	**9 323**	**8 064**	**2 247**	**888**	**324 829**	**289 170**	**199 372**	**78 203**

Table 4 – Rank correlations (Spearman) between twins

	(1) Years education	(2) Permanent income quintile	(3) Permanent workdays quartile
Identical twins – all sample	0.71	0.52	0.46
Fraternal twins – all sample	0.43	0.27	0.26
Identical twins – male	0.72	0.56	0.49
Fraternal twins – same sex: male	0.44	0.31	0.32
Identical twins – female	0.70	0.48	0.42
Fraternal twins – same sex: female	0.53	0.30	0.26
Fraternal twins – different sex	0.34	0.21	0.23
Identical twins - all sample – born after 1983	0.78	0.68	0.73
Fraternal twins - all sample – born after 1983	0.48	0.37	0.47
Identical twins - all sample – born 1983-1985	0.50	0.52	0.51
Fraternal twins - all sample – born 1983-1985	0.25	0.21	0.37
Identical twins - all sample – born before 1985	0.46	0.45	0.40
Fraternal twins - all sample – born before 1985	0.24	0.25	0.29

Note: bootstrapped 50 replications – all significant at 0.01

Table 5 – ACE decomposition (percent) of lifecycle outcomes

(A)
Years of education

	(1) age < 37 born after 1985		(2) age 37-39 born 1983-1985		(3) age > 39 born before 1983		(4) entire sample	
A genetics	43.68	***	62.88	***	50.6	***	46.88	***
C environment	35.77	***	6.66	***	22.71	***	26.55	***
E idiosyncratic	20.54	***	30.45	***	26.69	***	26.56	***
observations	1200		5524		3344		10068	

(B)
life-cycle earnings (quintile average over 4 decades)

	(1) age < 37 born after 1985		(2) age 37-39 born 1983-1985		(3) age > 39 born before 1983		(4) entire sample	
A genetics	48.44	***	49.95	***	23.46	***	49.20	***
C environment	0.79				27.72	***	3.59	
E idiosyncratic	50.77	***	50.05	***	48.8	***	47.19	***
observations	2970		4215		2537		9722	

(C)
life-cycle employment (quartile average over 4 periods)

	(1) age < 37 born after 1985		(2) age 37-39 born 1983-1985		(3) age > 39 born before 1983		(4) entire sample	
A genetics	43.85	***	38.29	***	21.22	**	39.50	***
C environment	2.21		6.73		18.64	**	6.86	*
E idiosyncratic	53.94	***	54.97	***	60.13	***	53.62	***
observations	2970		4215		2537		9722	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 6 – ACE decomposition (percent) of labour market outcomes, by decades

(1)

Earnings quintiles

	(a) Age < 37				(b) Age 37-39				(c) Age > 39							
	1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		3rd decade (age 41-50)		4th decade (age 51-60)	
A genetics	42.39	***	42.82	***	37.38	***	44.25	***	20.97	***	25.89	***	12.15		24.57	*
C environment	5.08				8.3				21.16	***	21.65	***	31.86	***	21.05	*
E idiosyncratic	52.53	***	57.18	***	54.32	***	55.75	***	57.86	***	52.45	***	55.98	***	54.37	***
observations	2942		1349		4031		4214		2350		2501		2247		888	

(2)

Workdays quartiles

	(a) Age < 37				(b) Age 37-39				(c) Age > 39							
	1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		3rd decade (age 41-50)		4th decade (age 51-60)	
A genetics	41.95	***	10.47		40.23	***	21.28	***	20.83	*	32.19	***	24.43	***	26.95	***
C environment	4.02		16.05		7.34		5.79		18.56	**	2.69		3.63			
E idiosyncratic	54.03	***	73.48	***	52.43	***	72.92	***	60.6	***	65.12	***	71.93	***	73.04	***
observations	2942		1349		4031		4214		2350		2501		2247		888	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 7 – ACE and AE decomposition (percent) of educational attainment by geographical macroarea

Education (controlling for gender) – ACE decomposition						
	(a) North					
	Age < 37		Age 37-39		Age > 39	
A genetics	65.27	***	58.42	***	46.49	***
C environment	21.05	***	9.24		27.01	***
E idiosyncratic	13.67	***	32.32	***	26.49	***
observations	462		2 259		1 238	
	(b) Centre					
	Age < 37		Age 37-39		Age > 39	
A genetics	0.00		60.99	***	41.00	***
C environment	50.30	***	5.60		27.57	***
E idiosyncratic	49.70	***	33.40	***	31.42	***
observations	308		1 158		1 393	
	(c) Sud					
	age < 37		age 37-39		age > 39	
A genetics	43.80	***	68.59	***	76.79	***
C environment	42.24	***	4.27		4.65	
E idiosyncratic	13.97	***	27.12	***	18.65	***
observations	402		2 103		682	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 8 – AE decomposition (percent) of labour market outcomes, by geographical macroarea

	permanent income rank (quintile average over 4 decades, controlling for gender)			workdays rank (average over 4 periods, controlling for gender)		
	(a) North					
	age < 37	age 37-39	age > 39	age < 37	age 37-39	age > 39
A genetics	48.61 ***	45.48 ***	47.60 ***	45.49 ***	41.23 ***	42.58 **
E idiosyncratic	51.39 ***	54.52 ***	52.40 ***	54.51 ***	58.77 ***	57.42 ***
observations	1 703	1 823	1 005	1 703	1 823	1 005
	(b) Centre					
	age < 37	age 37-39	age > 39	age < 37	age 37-39	age > 39
A genetics	39.96 ***	42.39 ***	54.97 ***	37.79	37.15 ***	40.49 *
E idiosyncratic	60.04 ***	57.61 ***	45.03 ***	62.21 ***	62.85 ***	59.51 ***
observations	780	921	1095	780	921	1095
	(c) Sud					
	age < 37	age 37-39	age > 39	age < 37	age 37-39	age > 39
A genetics	53.63 ***	45.57 ***	55.83 ***	57.39 ***	48.64 ***	37.29
E idiosyncratic	46.37 ***	54.43 ***	44.17 ***	42.61 ***	51.36 ***	62.71 ***
observations	487	1471	437	487	1471	437

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 9: ACE decomposition with varying degrees of hypothesized genetic correlation among DZ twins ($corrA_{DZ}$)

	Years of education							
	(1)		(2)		(3)		(4)	
	$corrA_{DZ}$ =0.50		$corrA_{DZ}$=0.55		$corrA_{DZ}$=0.60		$corrA_{DZ}$=0.65	
A genetics	46.88	***	52.09	***	58.6	***	66.95	***
C environment	26.55	***	21.34	***	14.82	***	6.48	
E idiosyncratic	26.56	***	26.56	***	26.56	***	26.56	***
observations	10068		10068		10068		10068	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 10: ACE and AE decompositions controlling for year and region of birth

(A)
Years of education

	(1) controlling for gender		(2) controlling for gender and birth year		(3) controlling for gender, birth year and region of birth	
A genetics	46.88	***	50.53	***	50.56	***
C environment	26.55	***	20.80	***	20.58	***
E idiosyncratic	26.56	***	28.67	***	28.84	***
observations	10068		10068		10068	

(B)
Life-cycle earnings (quintile average over 4 decades)

	(1) controlling for gender		(2) controlling for gender and birth year		(3) controlling for gender, birth year and region of birth	
A genetics	53.22	***	49.61	***	46.16	***
E idiosyncratic	46.77	***	50.38	***	53.83	***
observations	9722		9722		9722	

(C)
Life-cycle employment (quartile average over 4 periods)

	(1) controlling for gender		(2) controlling for gender and birth year		(3) controlling for gender, birth year and region of birth	
A genetics	47.33	***	44.02	***	41.85	***
E idiosyncratic	52.67	***	55.97	***	58.14	**
observations	9722		9722		9722	***

Table 11 – Additional information available for pseudo-twins: frequency of specific events

	Obs	Mean	Std. Dev.
Unemployment/layoff events			
age<30	324 829	0.102	0.191
age 31-40	289 170	0.155	0.265
age 41-50	199 372	0.163	0.277
age>50	78 203	0.180	0.313
Absence for illness events			
age<30	324 829	0.041	0.103
age 31-40	289 170	0.071	0.157
age 41-50	199 372	0.087	0.187
age>50	78 203	0.098	0.230
Parental leave events			
age<30	324 829	0.019	0.074
age 31-40	289 170	0.056	0.151
age 41-50	199 372	0.035	0.141
age>50	78 203	0.030	0.151
Pro-social (absence for blood donation)			
age<30	324 829	0.004	0.037
age 31-40	289 170	0.013	0.086
age 41-50	199 372	0.019	0.118
age>50	78 203	0.016	0.114
Religious vocation (contributing to clergy pension fund)			
age<30	324 829	0.000	0.021
age 31-40	289 170	0.001	0.028
age 41-50	199 372	0.001	0.031
age>50	78 203	0.001	0.030

Table 12 – Pseudo twins correlations

	Years education	Permanent income decile	Permanent workdays quartile	Uunemploy ment/lay-off events	Absence for illness events	Parental leave events	Pro-social (blood donations)	Religious vocation (clergy)
Same-sex	0.545	0.425	0.363	0.274	0.242	0.271	0.191	0.086
Mixed-sex	0.300	0.158	0.134	0.113	0.113	-0.051	0.050	-0.006

Note: bootstrapped 50 replications – all significant at 0.01

Table 13 – ACE/AE decomposition (percent) of lifecycle outcomes – pseudo twins

	Years of education		Life-cycle earnings		Life-cycle employment	
A genetics	52.57	***	34.45	***	34.21	***
C environment			5.05	***		
E idiosyncratic	47.42	***	60.48	***	65.78	***
observations	344 204		344 204		345 136	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 14 – ACE decomposition (percent) of lifecycle outcomes by birth cohort– pseudo twins

	(A) Life-cycle earnings			
	1 born in 1982 or before		2 born after 1982	
A genetics	33.46	***	34.96	***
C environment	4.21	**	2.43	**
E idiosyncratic	62.31	***	62.59	***
observations	226 558		117 646	
	(B) Life-cycle employment			
	1 born in 1982 or before		2 born after 1982	
A genetics	29.15	***	32.25	***
E idiosyncratic	70.84	***	67.74	***
observations	226 558		117 646	

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table 15 – AE decomposition (percent) of additional outcomes – pseudo twins

	Unemployment		Absence		Parental leave		Blood donations		Religious vocation	
A genetics	30.11	***	24.34	***	12.25	***	19.87	***	7.73	***
E idiosyncratic	69.88	***	75.65	***	87.74	***	80.12	***	92.26	***

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1 –
Number of observations 344 204 - For blood donations and religious vocation we impose a unit correlation of genetic factors among same sex pseudo twins

Appendix

Table A1 – Descriptive statistics for ISS

	(a) identical twins (ISS)			(b) fraternal twins (ISS) – same sex			(c) fraternal twins (ISS) – other sex		
Demographics	mean	sd	obs	mean	sd	obs	mean	sd	obs
Female	0.60	0.48	6206	0.57	0.49	4074	050	0.50	3220
Age	39.15	7.84	6206	37.39	7.73	4074	37.68	7.35	3220
Years of education	12.96	2.80	4574	12.73	2.70	2774	12.59	2.63	2373
Age of leaving family	22.99	5.12	4921	22.22	4.54	2593	22.13	4.31	2236
Income quintile	3.03	1.29	4530	2.97	1.32	2882	2.97	1.31	2310
Work duration quartile	2.23	0.87	4530	2.18	0.86	2882	2.19	0.85	2310

	(a) age < 37 born after 1985			(b) age 37-39 born 1983-1985			(c) age > 39 born before 1983		
Demographics	mean	sd	obs	mean	sd	obs	mean	sd	obs
Female	0.52	0.49	4612	0.57	0.49	5616	0.63	0.48	3372
Age	30.68	3.28	4612	37.97	0.76	5616	49.11	5.40	3372
Years of education	12.87	2.47	1200	12.25	1.97	5524	13.71	3.56	3344
Age of leaving family	21.27	3.01	1128	19.90	1.15	5217	27.61	5.12	3058
Income quintile	2.71	1.37	2970	3.01	1.27	4215	3.32	1.20	2537
Work duration quartile	2.04	0.93	2970	2.22	0.83	4215	2.37	0.75	2537

Table A2 – AE decomposition (percent) of lifecycle outcomes

	(1) Years of schooling		
	(a) Age < 37 born after 1985	(b) Age 37-39 born 1983-1985	(c) Age > 39 born before 1983
A genetics	79.83 ***	69.88 ***	73.86 ***
E idiosyncratic	20.17 ***	30.12 ***	26.14 ***
	(2) Life-cycle earnings (average over 4 periods)		
	(a) Age < 37 born after 1985	(b) Age 37-39 born 1983-1985	(c) Age > 39 born before 1983
A genetics	49.37 ***	49.95 ***	53.63 ***
E idiosyncratic	50.63 ***	50.05 ***	46.37 ***
	(3) Life-cycle employment (average over 4 periods)		
	(a) Age < 37 born after 1985	(b) Age 37-39 born 1983-1985	(c) Age > 39 born before 1983
A genetics	46.44 ***	45.98 ***	42.10 **
E idiosyncratic	53.56 ***	54.02 ***	57.90 ***

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table A3 – AE decomposition (percent) of labour market outcomes, by decades

(1)
Earnings quintile

	(a) Age < 37 born after 1985				(b) Age 37-39 born 1983-1985				(c) Age > 39 born before 1983							
	1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		3rd decade (age 41-50)		4th decade (age 51-60)	
A genetics	48.30	***	42.82	***	46.81	***	44.25	***	44.50	***	49.70	***	47.26		47.63	*
E idiosyncratic	51.70	***	57.18	***	53.19	***	55.75	***	55.50	***	50.30	***	52.74	***	52.37	***

(2)
Workdays quartiles

	(a) Age < 37 born after 1985				(b) Age 37-39 born 1983-1985				(c) Age > 39 born before 1983							
	1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		1st decade (age 20-30)		2nd decade (age 31-40)		3rd decade (age 41-50)		4th decade (age 51-60)	
A genetics	46.66	***	29.61		48.43	***	28.18	***	41.61	*	35.25	***	28.60	***	26.96	***
E idiosyncratic	53.34	***	70.39	***	51.57	***	71.82	***	58.39	***	64.75	***	71.40	***	73.04	***

Note: controls include gender – statistical significance: *** p<0.01, ** p<0.05, * p<0.1