
WORKING PAPERS

On the Origins of Socioeconomic: Evidence from Twin Families

Paul **BINGLEY**¹
Lorenzo **CAPPELLARI**^{2,3}
Konstantinos **TATSIRAMOS**⁴

¹ VIVE Copenhagen, Denmark

² Luxembourg Institute of Socio-Economic Research (LISER), Luxembourg

³ Università Cattolica Milano, Italy

⁴ University of Luxembourg, Luxembourg

LISER Working Papers are intended to make research findings available and stimulate comments and discussion. They have been approved for circulation but are to be considered preliminary. They have not been edited and have not been subject to any peer review.

The views expressed in this paper are those of the author(s) and do not necessarily reflect views of LISER. Errors and omissions are the sole responsibility of the author(s).

On the Origins of Socioeconomic Inequalities: Evidence from Twin Families

Paul Bingley

VIVE Copenhagen

Lorenzo Cappellari

Università Cattolica Milano and
Luxembourg Institute of Socio-
Economic Research (LISER)

Konstantinos Tatsiramos

University of Luxembourg

June 2024

Using Danish Twins Registry and population data, we link twins with their relatives to evaluate the controversial assumptions of the classic twin model and decompose socioeconomic inequality into genetic (heritability) and environmental factors. We reject the equal environments assumption, finding that the classic model overestimates heritability. Heritability explains 9% of variation in education and 14-16% in earnings, income, and wealth, helping to fill the ‘missing heritability’ gap between the classic twin model and Genome-Wide Association Studies. Shared environments account for 26-42% of these variances and 45-81% of intergenerational persistence. These findings reconcile estimates from twin and adoptee studies.

Keywords: nature, nurture, family background, genes, environment, inequality
JEL codes: D31; D63; E21; E24; I24

Acknowledgments: This research was financed by the Luxembourg National Research Fund (FNR), grant reference [C21/SC/15771764]. We thank Kaare Christensen, Mikael Lindahl, Petter Lundborg, Erik Plug, and Jan Stuhler for their discussions about this project. We are also grateful for the comments received at the ESSGN and SoLE Conferences. An earlier version was circulated as IZA Discussion Paper No. 16520. All errors and omissions are the responsibility of the authors.

1. Introduction

Inherited endowments and family background are pivotal in shaping key determinants of individual success, such as human capital, abilities, and skills (Becker and Tomes, 1979). Consequently, a comprehensive understanding of economic inequality requires tracing individuals back to their social origins. A considerable body of research has investigated the origins of observed inequality in socioeconomic outcomes (Black and Devereux, 2011; Sacerdote, 2011; Mogstad and Torsvik, 2023, for reviews). This research has employed two approaches to infer the relative influences of genes and environment.

One approach, grounded in behavioral genetics, known as the ACE model, decomposes cross-section inequality of a population at a given point in time into the proportions explained by additive genetic inputs (A), common (or shared) environments (C), and idiosyncratic (non-shared) environments (E). Empirical evidence from applying the ACE model on twins and adoptees finds a limited or no role for shared environments in explaining the cross-sectional variation in socioeconomic outcomes such as educational attainment and income (see Sacerdote, 2011, for a discussion).¹ However, despite frequent use in economics, the ACE model has received critiques because it relies on several restrictive assumptions, which can lead to biased estimates of its variance components (e.g., Goldberger, 1979; Evans and Martin, 2000).²

An alternative approach disentangles genetic and family environmental effects of the intergenerational elasticity (IGE) of socioeconomic outcomes by regressing the child's outcomes on the parent's outcomes separately for adopted and non-adopted children. Unlike studies applying the ACE model, intergenerational regressions on adoptees find substantial family environmental effects on children's outcomes (Black and Devereux, 2011; Sacerdote,

¹ The ACE model using *twins* has been applied to study earnings (Behrman and Taubman, 1976; Taubman, 1976; Behrman and Taubman, 1989; Björklund, Jäntti and Solon, 2005), education (Behrman and Taubman, 1976; Branigan et al., 2013; Silventoinen et al., 2020), preferences for risk taking (Cesarini et al., 2009a), overconfidence (Cesarini et al., 2009b), investor behavior (Barnea, Cronqvist, and Siegel, 2010) and savings behavior (Cronqvist and Siegel, 2015). The ACE model using *adoptees* has been applied to study education and family income (Sacerdote, 2007) and wealth (Fagereng, Mogstad, and Rønning, 2021).

² Beyond socioeconomic outcomes, the ACE has been used to decompose the relative contributions of genetic and environmental factors to individual differences in a variety of outcomes, such as cognitive abilities (Plomin, 1988), personality traits (Nichols, 1978), psychiatric disorders (Kendler, 2001). Polderman et. al. (2015) perform a meta-analysis of twin studies on a very wide variety of human traits.

2011; Mogstad and Torsvik, 2023).³ This alternative approach offers valuable insights into the intergenerational aspect of the nature versus nurture debate and have gained favor among economists. This popularity is largely due to the ability of adoptee studies, under quasi-random assignment of adoptees to families, to identify the treatment effect of being raised in different families (Mogstad and Torsvik, 2023).

More recently, a small but growing literature is emerging linking genetic data to socioeconomic outcomes (see Biroli et al., 2022; Benjamin et al., 2024). Whereas twin and adoptee studies compare similarities between different types of family members to understand the relative contribution of genetic and environmental factors, this research uses genome-wide association studies (GWAS) to identify specific genetic variants associated with outcomes of interest to economists.⁴

This paper uses a twin family design that links twins with their spouses and children.⁵ The twin family design offers two important advantages. First, it provides a unified framework that integrates studies using the ACE model or intergenerational regressions on adoptees to assess the influence of genetic and environmental factors on inequality within and between generations. Second, it allows to critically assess the sensitivity of the results of the ACE model to its strong assumptions using information only on close family members (twins, their spouses, and children). Studies using different types of sibling pairs to assess the sensitivity of the results of the ACE model either restrict the analysis to cross-sectional inequality (Björklund, Jäntti and Solon, 2005), or relax only specific assumptions, such as the correlation between genes and the environment (Fagereng, Mogstad, and Rønning, 2021). The twin family design offers an alternative approach to testing the assumptions of the ACE model without the need for

³ The outcomes considered in adoptees studies include education or earnings (Sacerdote, 2002; Björklund, Lindahl, and Plug, 2006; Holmlund, Lindahl, and Plug, 2011; Lundborg, Plug, and Rasmussen, 2021), and wealth (Black et al., 2020; Fagereng, Mogstad, and Rønning, 2021). Beyond the conventional parent-child correlations, Adermon, Lindahl, and Palme (2021) estimate long-run intergenerational persistence in human capital using the extended family and adoptees, while Collado, Ortuño-Ortín, and Stuhler (2023) compare different degrees of kinship within the same or close generations to quantify the importance of family background on educational attainment based on an explicit intergenerational and assortative model.

⁴ This approach has been applied to study skills and educational attainment (Rietveld et al., 2013; Okbay et al. 2022; Houmark, Ronda and Rosholm, 2024), as well as earnings and wealth (Papageorge and Thom, 2019; Barth, Papageorge and Thom, 2020).

⁵ Nance and Corey (1976) introduced the twin family model, which has been applied to study outcomes such as smoking during pregnancy, (D’Onofrio et al. 2003), depression (Silberg et al. 2010), anxiety (Eley et.al. 2015). Behrman and Rosenzweig (2002) use schooling differences between twin parents and between their children to estimate the intergenerational transmission of schooling.

genetic data, whereas GWAS studies rely on detailed genetic information to test these assumptions.

Our focus for the main part of the analysis is on education attainment, measured by years of schooling, because of the role of human capital in shaping a range of subsequent outcomes (e.g., Blanden, Doepke, Stuhler, 2023) and the attention that educational attainment has received in the broader social sciences, in the behavioral genetics and social-science genomics literatures. We reconstruct the twins' families for large samples by combining data from the Danish Twin Registry (identifying twins and their zygosity) with population registers.⁶ We begin by estimating the parameters of the ACE model using only the twin pairs, also known as the Classic Twin Design (CTD), which compares the resemblance of monozygotic (MZ) twins, who share all of their genes at conception, with dizygotic (DZ) twins, who share, on average, half of their segregating genes. We find that, in the CTD, genetic factors account for 34% of the variation in years of schooling; this percentage is also known as the degree of heritability. Shared environments account for 24% of the variation in education, with the remaining variation attributed to non-shared environments.

We then estimate more general models and critically assess the sensitivity of the results derived from the CTD model to violations of its assumptions. These assumptions include (i) no assortative mating based on genetic factors, (ii) absence of genetic dominance, (iii) no gene-environment correlation, (iv) no gene-environment interactions, and (v) equally shared environments (EEA), which asserts that MZ and DZ twins experience similar environmental influences not shaped by their genetic similarities (Evans and Martin, 2000). We maintain the assumption of no gene-environment correlation, as empirical studies focusing on socioeconomic outcomes consistently find no evidence supporting such correlations (e.g., Björklund, Jäntti, and Solon, 2005; Fagereng, Mogstad, and Rønning, 2021; Biroli et al., 2022; Collado, Ortuño-Ortín, and Stuhler, 2023).

We find a positive correlation (0.12) between genetic factors related to education among twins and their spouses, although we find no evidence of genetic dominance. To examine gene-environment interactions, we exploit an educational reform as a natural experiment to test whether genetic factors vary with exposure to enhanced educational opportunities introduced

⁶ As unmarried co-parenting is very common in Denmark, throughout the paper we use the word “spouses” to also include co-parents, i.e., “spouses” in our study are not necessarily married.

by a schooling expansion. In a fully saturated model, which interacts the genetic, shared, and non-shared components with the exogenous environmental change, we find that the additive genetic component of the variance in years of schooling remains constant. However, shared and non-shared environmental components are significantly lower for those exposed to the reform. Consequently, as a result, heritability estimates are higher among those who received additional educational opportunities.

Finally, we reject the EEA for MZ and DZ twins, demonstrating significant differences in the degree of shared environments between these twin pairs.⁷ By allowing for varied shared environments between twin pairs, we find that heritability accounts for 9% of the total variance in years of schooling, a substantial decrease from the 34% estimated by the CTD. Conversely, shared environments account for 42% of the variance, significantly higher than the 24% estimated by the CTD, with the remainder attributed to non-shared environmental factors. These findings underscore that the CTD's EEA is overly restrictive and leads to overestimating the role of additive genetic factors in explaining educational inequalities. We reach similar conclusions by extending our analysis to earnings, income, and wealth. Genetic factors account for 16% of the total variance in earnings, 14% for assets, and 14% for income, whereas shared environments explain 26% of the total variance across these measures.

Additionally, using the parameter estimates from the twin family design, we can decompose the intergenerational elasticities of earnings, income, wealth, and education. Shared environments contribute substantially to transmitting these socioeconomic outcomes across generations. Specifically, shared environments account for 55% of the rank IGE in earnings, 61% of the rank IGE in income, 71% of the rank IGE in assets, and 81% of the IGE in years of education.

Our study speaks to the literature that examines the relative contribution of genetic and environmental factors in shaping socioeconomic inequalities (Black and Devereux, 2011; Sacerdote, 2011; Mogstad and Torsvik, 2023). We reconcile the estimates from the two prevalent approaches in the field. Studies using the ACE model tend to find that genetic factors explain a larger proportion of cross-sectional inequality than shared environments, while

⁷ The most common methodologies for testing the EEA depend on measures of environmental similarity such as the twin's or their parents' perception of their zygosity, their frequency of contact during youth and their similarity in appearance (Felson, 2014). The main concern with such tests is that they require prior knowledge about which environmental factors influence the outcome of interest.

research focusing on the decomposition of intergenerational elasticity finds that the contribution of family environments is at least as large as that of genetic factors. Our analysis demonstrates that shared environmental factors account for a substantial proportion of both the cross-sectional variance (26% to 42%) and the IGE of socioeconomic outcomes (55% to 81%), surpassing the contributions of genetic factors, which are responsible for only 9% to 16% of the cross-sectional variance and 19% to 45% of the intergenerational correlations. Additionally, we build a model that requires data from only a few relatives—up to six family members—with known genetic relatedness. This model enables testing the CTD’s assumptions, offering a novel perspective to examining a broad set of socioeconomic outcomes.

Our study also contributes to the social-science genomics literature, which typically reports much lower heritability estimates than those found in twin and adoptee studies using the ACE model. By demonstrating that the ACE model tends to overestimate heritability, we show that the discrepancy between the ACE model and the social-science genomics literature, known as the "missing heritability" problem, is not as large as previously suggested.

The rest of the paper is structured as follows. Section 2 describes the data. Section 3 presents the CTD and decomposes the cross-sectional variation of education into genetic, shared environmental, and non-shared environmental factors. Section 4 presents extensions of the CTD that relax and test its assumptions of no assortative mating based on genetic factors (4.1), absence of genetic dominance (4.2), no gene-environment interactions (4.3), and the equal environment assumption (4.4). Section 5 reports results for earnings, income, and wealth. Section 6 concludes.

2. Data

We use data from the Danish Twins Registry combined with administrative population registers. Since the civil registration system was established in 1968, every resident in Denmark has been registered with a unique personal identification number that is used in all national registers, thereby enabling accurate linkage. The Twins Registry has identified more than 170,000 twins born since 1870 through parish and hospital records (Skytthe et al., 2002). Zygosity is established for same-sex twins according to responses to four survey questions about twin similarity, a method validated with an overall accuracy of 96% (Christiansen et al., 2003).

We sample all MZ and DZ twin pairs in the Twins Registry. We aim to construct a dataset of twin families; each centered around a single twin pair and including the children of the twins and their co-parents. Using links from children to parents registered shortly after birth—links that originate from municipal and parish records—we find first-born children of the twins and the co-parents of the children (twins' spouses). Parent-child links are complete for births from 1955, which defines the earliest birth year for which we sample children of twins in our data. Record incompleteness also implies that, for parents born before 1935, the first registered child may not be the firstborn, so we exclude families in which twins or spouses are born before 1935. To include individuals who are old enough to have completed their education and to be observed in the labor market, we also exclude children of twins born after 1984.⁸

Table 1 presents the sample that we use in estimation, split according to family role (whether a twin, spouse, or child), gender, and zygosity of the twin pair. The sample contains 80,205 individuals, of whom 33,915 are twins, 26,682 are twin spouses, and 19,608 are children of twins.⁹ These individuals belong to 17,325 twin families, of which 43% have two children born before 1985, 28% have one child born before 1985, and the remaining 30% have no child born before 1985. Table 1 shows that, on average, twins and their spouses are born in the early 1950s, while the children of the twins are born in the early 1970s. In our sample, the average age at first birth (not shown) is 26 for men and 24 for women.

We match the information on twin families with administrative data from population registers. Educational institutions report qualifications to the Ministry of Education; see Jensen and Rasmussen (2011). Statistics Denmark calculates the highest level of education based on information from the Ministry about prerequisites and normed times for completing each qualification. The qualification that would take the longest time for an individual to obtain by the shortest possible route defines the highest level of education. Using this definition, we measure educational attainment as the highest level by age 29. Descriptive statistics show an increase in average attainment and a reduction in its dispersion, in line with educational trends in Denmark (Karlson and Landersø, 2021).

⁸ We exclude any family with parents born after 1966, as these parents would not have reached age 18 before 1984. We also exclude the very small number of families in which both parents are twins.

⁹ We do not observe a co-twin for 735 twins because they emigrated or passed away outside our observation period. However, we observe children and spouses for some of these missing twins.

3. The Classic Twin Design

The CTD is the version of the ACE model most often used for decomposing trait variation into additive genetic and shared environmental factors (Keller and Coventry, 2005). By comparing the similarities between MZ twins with those between DZ twins, the CTD provides insights into how much of the variation in a trait across a population can be explained by genetics versus the environment shared by twins growing up together.

3.1 Model and Identification

Let y_i denote the long-term outcome of person i in deviations from the population mean and consider the following factorization

$$y_i = a_i + c_{f(i)} + e_i, \quad h(i) \in \{T_1, T_2\} \quad (1)$$

where $f(i)$ denotes i 's family of origin, while $h(i)$ denotes the person's role in the model; within the framework of the CTD, all persons are twins and are denoted by T_1 or T_2 . In the model of Equation (1), a_i is an additive genetic factor, $c_{f(i)}$ is an environmental factor shared among the offspring of family f , and e_i is a non-shared environmental factor representing individual idiosyncratic variation in outcomes. Factors are drawn from zero mean distributions with variances σ_a^2 , σ_c^2 and σ_e^2 .

The CTD relies on five assumptions to identify the three variance components. The first assumption is the absence of *assortative mating* based on *additive genetic factors*. This assumption implies that mating is random, so DZ twins share, on average, half of their segregating genetic endowments. Positive assortative mating implies that DZ twins share more than half of their segregating genetic endowment due to non-random selection of partners based on genetic similarities. The second assumption is *no genetic dominance*, i.e. that gene variants do not interact but add up to affect outcomes. The third assumption is that the three components of Equation (1) enter the model linearly. This assumption rules out the presence of *gene-environment interactions*, which occur when the effect of an individual's genetic makeup on an outcome varies across different environmental contexts. The fourth assumption of the CTD is that MZ and DZ twins share environments to the same extent, irrespective of their genetic similarity, implying that the environmental factor $c_{f(i)}$ is drawn from a distribution that is common for MZ and DZ twins. The fifth assumption is that there is no gene-environment

correlation, which implies that individuals do not actively choose or are not placed into environments based on their genetic predispositions.

Under these assumptions, information on variances and covariances of outcomes identifies the variances of the three components in equation (1). More specifically, the total variance of outcome y_i is

$$\text{var}(y_i) = \sigma_a^2 + \sigma_c^2 + \sigma_e^2. \quad (2)$$

The covariance of outcomes between twins is

$$\text{cov}(y_i, y_{i'})_z^{T_1 T_2} = 0.5^{(1-z)} \sigma_a^2 + \sigma_c^2, \quad (3)$$

where the indicator z is equal to one for MZ twins and equal to zero for DZ twins.¹⁰ Equation (3) encompasses two moment restrictions depending on the value of z . For MZ twins ($z = 1$), who share all of their genes at conception, the loading of the genetic factor is equal to 1. For DZ twins ($z = 0$), who share only half of their segregating genes on average, the loading of the genetic factors is equal to 0.5 in the absence of assortative mating and genetic dominance. Equation (2) and the two moment restrictions of equation (3) identify the three variance components of the CTD model exactly. Therefore, the variance of outcome y_i can be decomposed into genetic, shared environmental, and idiosyncratic factors. Using this decomposition, the degree of heritability is measured by the proportion of the genetic component within the total outcome variance, which is defined as $\sigma_a^2 / (\sigma_a^2 + \sigma_c^2 + \sigma_e^2)$.

We estimate the parameters of the model by Minimum Distance, matching empirical variances and covariances to the corresponding moments generated by the model (i.e., Equations 2 and 3).¹¹ Specifically, we use an Equally Weighted Minimum Distance estimator in which we weight the minimization by the identity matrix and correct standard errors using the empirical matrix of fourth moments.

3.2 Results

We report the parameter estimates of the variance components for education in Panel A, Column (1) of Table 2. We find that the additive genetic factor (σ_a^2) is larger than the shared

¹⁰ For brevity, in the main text we use compact notation. In Appendix Table A1, we present the covariances of each relationship for all the models that we estimate.

¹¹ Equation (3) implies that the genetic variance component is equal to twice the difference in the covariance between MZ and DZ twins, allowing heritability to be alternatively estimated as twice the difference in correlation between MZ and DZ twins.

environmental factor (σ_c^2). In Panel B, we use these estimates to decompose cross-sectional inequality, finding that heritability accounts for 34% of the total variance. In contrast, the shared environment accounts for 24%, with the remaining variance attributed to the idiosyncratic environmental factor that twins do not share.¹²

Heritability estimates in the behavioral genetics literature rely on the use of Multilevel Mixed-effects models. One key distinction between the Minimum Distance and Mixed-effects estimators lies in the latter's reliance on distributional assumptions, typically normality, for deriving the likelihood function (Kim, 2009). For example, using twin registers from several countries, Silventoinen et al. (2020) applied the CTD to investigate educational attainment using Multilevel Mixed-effects models, estimating heritability at 43%. When we apply the Mixed-effects CTD model to our data, heritability is estimated at 48% (instead of 34% with Minimum Distance). Shared environmental factors account for 15% of the variation in educational attainment (instead of 24% with Minimum Distance), with the remaining variance attributable to non-shared environments. Overall, our heritability estimate from the CTD is consistent with the evidence in the literature that, on average, 40% of the variance in educational attainment can be explained by additive genetic factors (Branigan et al., 2013).

4. Extensions of the Classic Twin Design

In this section, we extend the CTD to allow for assortative mating based on genetic factors (section 4.1), genetic dominance (section 4.2), gene-environment interactions (section 4.3), and heterogeneous shared environments (section 4.4). Fagereng, Mogstad, and Rønning (2021) investigate the extent of *gene-environment correlation* within the ACE framework by exploiting the quasi-random assignment of Korean adoptees in Norway. Their results suggest that this correlation is not statistically different from zero for various outcomes, including education, while it is negative for net wealth. Negative and insignificant gene-environment correlations are also reported by Björklund, Jäntti, and Solon (2005), who apply the ACE model to various sibling types, Collado, Ortuño-Ortín, and Stuhler (2023), who analyze siblings and siblings-in-law to the sixth degree, and Biroli et al. (2022), based on a GWAS analysis.

¹² In our data, the empirical correlation for years of education is 0.58 for MZ twins and 0.43 for DZ twins. Consequently, estimating heritability as twice the difference in the correlation between MZ and DZ twins yields a heritability estimate of 30%, which is close to the Minimum Distance estimate (34%).

Drawing on this evidence, we develop our model, maintaining no gene-environment correlation.

With only three moment restrictions, we can no longer identify the model if we relax any of the other four CTD assumptions. However, by including information from the twins' relatives, we can add moment restrictions and relax the assumptions while still identifying the model. A twin family design provides this information by including the twins' spouses and children.

4.1 Assortative Mating

Assortative mating is the tendency for individuals to mate with similar partners. Researchers have found evidence of assortative mating on education (Güell, Rodríguez Mora, and Telmer, 2015; Eika, Mogstad, and Zafar, 2019; Collado, Ortuño-Ortín, and Stuhler, 2023), earnings (Gonalons-Pons, 2017), income (Greenwood et al., 2014), and wealth (Fagereng, Guiso, and Pistaferri, 2022). In contrast to these studies of assortative mating on socioeconomic outcomes, we are interested in the related but distinct phenomenon of assortative mating based on genetic factors (Neale, 2009). The CTD posits the absence of genetic assortative mating, assuming that couples are formed by individuals whose genes are randomly drawn from the population. Assortative mating on genetic factors would lead to higher genetic resemblance between DZ twins than expected, which would cause the CTD to underestimate heritability and overestimate shared environmental factors (Freese and Jao, 2017).

4.1.1 Model and Identification

To allow for assortative mating based on genetic factors, we extend the CTD to include the spouses of twins, denoted by S . A person in the model can now be either a twin or a twin spouse, i.e., $h(i) \in \{T_1, T_2, S_1, S_2\}$. Let δ be the assortative mating parameter, defined as the correlation of genetic factors among spouses (see Bowles and Gintis, 2002).¹³ With assortative mating, the spousal covariance is given by

$$\text{cov}(y_i, y_{i'})^{T_j S_j} = \delta \sigma_a^2 + \gamma_S, \quad j = 1, 2 \quad (4)$$

¹³ While Bowles and Gintis (2002) do not present estimates of δ , they conjecture that a reasonable value for it could be 0.2 (p. 11).

where γ_S parameterizes environmental sharing among spouses. Unlike twins, spouses are born in different families and therefore draw the environmental component $c_{f(i)}$ from different distributions.

The correlation of genetic factors between spouses does not affect the genetic similarity among MZ twins. However, assortative mating increases the genetic similarity of DZ twins by a factor of 0.5δ , such that the twin covariance becomes

$$\text{cov}(y_i, y_{i'})_z^{T_1 T_2} = (0.5(1 + \delta))^{(1-z)} \sigma_a^2 + \sigma_c^2, \quad (5)$$

Equation (5) extends equation (3) of the CTD to incorporate assortative mating based on genetic factors. The model resulting from equations (2), (4), and (5) is not identified because it provides four moment restrictions for five parameters. To identify this extended model, we use information on the covariances between siblings-in-law. These covariances depend on environmental factors and, assuming assortative mating on genes, also depend on genetic factors, with the strength of the genetic connectedness depending on the degree of zygosity of the twins. The covariances between siblings-in-law add four equations (two for each type of zygosity) and only one environmental parameter.¹⁴ The following two equations give the covariances for siblings-in-law

$$\text{cov}(y_i, y_{i'})_z^{T_j S_k} = \delta(0.5(1 + \delta))^{(1-z)} \sigma_a^2 + \gamma_{SL} \quad j, k = 1, 2 \quad j \neq k \quad (6)$$

$$\text{cov}(y_i, y_{i'})_z^{S_1 S_2} = \delta^2(0.5(1 + \delta))^{(1-z)} \sigma_a^2 + \gamma_{SL}, \quad (7)$$

where γ_{SL} parameterizes environmental sharing between siblings-in-law. Equation (6) defines the covariance between twins and co-twins' spouses. Equation (7) defines the covariance between the spouses of the twins.

4.1.2 Results

We find that the correlation in genetic factors between spouses (δ) related to education is 0.12 (s.e. 0.03), which we report in Panel A, Column (2) of Table 2, together with the parameter estimates of the variance components. This estimate is close to the spousal correlation of the educational attainment polygenic score of 0.18 reported in Okbay et al. (2022) and 0.13 in

¹⁴ We use the term ‘‘siblings-in-law’’ to describe the relationship to either a spouse’s co-twin or a co-twin’s spouse, regardless of their marital status.

Conley et al. (2016).¹⁵ We report the decomposition of cross-sectional inequality in Panel B, Column (2) of Table 2. Allowing for assortative mating on genes, we find a higher heritability estimate (45%) compared with the CTD (34%) and a lower percentage of educational inequality attributed to shared environments (14% compared to 24% in the CTD). This finding is consistent with theoretical expectations that not accounting for assortative mating in the CTD results in underestimating heritability and overestimating shared environmental factors. We reach this conclusion while maintaining the other assumptions of the CTD, a conclusion to which we return in Section 4.4.

4.2 Genetic Dominance

Individuals receive two versions of each gene, one from each parent. Genetic dominance refers to the relationship between these two versions, whereby one (the dominant) version masks the expression of the other (Neale, 2009). A type of gene-gene interaction, dominance, is a relationship exclusive to two versions of the same gene. Thus, full siblings can exhibit a degree of genetic dominance because of the combinations of genes they receive from their parents. In contrast, other relatives exhibit much less dominance because they inherit genetic combinations less stably.

The CTD asserts that genetic variation has an additive effect, with no genetic dominance or gene-gene interactions. Genetic dominance introduces non-additive genetic effects, where the presence of a dominant version of a gene can disproportionately influence a trait beyond what would be expected from simply adding up the effects of individual gene variants. This phenomenon can lead to reduced trait similarity among DZ twins because they might not share the same dominant gene variant, even though they share 50% of their genetic material. Consequently, the CTD might attribute the reduced similarity among DZ twins to a lack of shared additive genetic factors. However, because the CTD model uses the difference in outcome similarity between MZ and DZ twins to estimate heritability, not accounting for the role of genetic dominance could lead to overestimating the purely additive genetic contribution (Neale, 2009).

¹⁵ This estimate of $\delta=0.12$ also implies that genetic similarity accounts for about 13% of the spousal correlation in education, computed as $\delta\sigma_a^2/(\delta\sigma_a^2 + \gamma_S)$.

4.2.1 Model and Identification

To focus on the implications of allowing for genetic dominance, we maintain the other assumptions of the CTD. Under these assumptions, and in particular no assortative mating in genes, the impact of dominance would be confined to DZ twins (Keller et al., 2009). In the presence of genetic dominance, we extend the factorization of education to include an orthogonal zero-mean dominant genetic factor d with variance σ_d^2

$$y_i = a_i + c_{f(i)} + d_i + e_i. \quad (8)$$

Cesarini et al. (2009a) refer to the extension of the ACE model, which includes adding a genetic dominance component, as the ACDE model. By incorporating genetic dominance into our analysis, we introduce a fourth component into the variance decomposition of the CTD

$$\text{var}(y_i) = \sigma_a^2 + \sigma_c^2 + \sigma_d^2 + \sigma_e^2. \quad (9)$$

Given that MZ twins are genetically identical, sharing all their genes, their covariance function fully reflects the dominance factor. However, for DZ twins who share, on average, only half of their segregating genes, the dominance factor enters their covariance function with a weight of 0.25. This reduced weight is the probability that parents transmit the same (dominant) gene variant to each twin. Since each parent has a 50% chance of passing on a dominant gene variant and considering the requirement for both parents to transmit the same gene variant for dominance to be expressed, which has a 50% chance, the resulting probability is 0.25. Therefore, the covariance for twins becomes

$$\text{cov}(y_i, y_{i'})_z^{T_1 T_2} = 0.5^{(1-z)} \sigma_a^2 + \sigma_c^2 + 0.25^{(1-z)} \sigma_d^2. \quad (10)$$

The extended CTD that allows dominance is no longer identified because we now have four parameters to estimate (additive genetic, shared environmental, non-shared environmental, and dominance) but only three moment restrictions. To identify this extended model, it is necessary to introduce additional moment restrictions with the following two requirements: first, they should depend on shared environmental and additive genetic factors, excluding dominance effects; second, given the common environment, they should allow for variation in additive genetic factors across different types of relationships. Covariances between spouses or siblings-in-law do not satisfy these requirements because, under random mating (an assumption maintained in this subsection), these covariances only depend on environmental factors. Similarly, parent-child covariances do not meet these requirements, as they do not encompass variation in additive genetic factors. Parent-child covariances add only

one moment restriction and one common environment parameter, which is insufficient for identification.

Additional moment restrictions that satisfy the necessary criteria for identifying the CTD extended to include dominance are found in relationships between uncles or aunts and nephews or nieces (avuncular relationships) and between cousins. Under the assumption of random mating, these moment restrictions do not depend on genetic dominance. The covariances between these pairs depend only on shared environmental and additive genetic factors, where the degree of genetic linkage in these relationships varies according to the twins' zygosity. Each of these pairs includes at least one member from the offspring generation, and an individual in this extended model can now be either a twin (T) or a twin's child (C), i.e., $h(i) \in \{T_1, T_2, C_1, C_2\}$. This approach leverages the unique genetic architecture of avuncular and cousin relationships within twin families to introduce additional moment restrictions. The covariance for avuncular relations is given by

$$\text{cov}(y_i, y_{i'})_z^{T_j C_k} = 0.5^{(1-z)} 0.5 \sigma_a^2 + \gamma_A, \quad j, k = 1, 2 \quad j \neq k \quad (11)$$

where γ_A parameterizes environmental sharing in avuncular relationships. The covariance for cousins is given by:

$$\text{cov}(y_i, y_{i'})_z^{C_1 C_2} = 0.5^{(1-z)} 0.25 \sigma_a^2 + \gamma_C, \quad (12)$$

where the parameter γ_C captures environmental sharing among cousins. This extended CTD model is identified because equations (11) and (12) add four moment restrictions—two for each zygosity—and two environmental parameters.

4.2.2 Results

We find that the variance component of the dominance factor σ_d^2 is relatively small and statistically insignificant, which we report in Panel A, Column (3) of Table 2. While the estimate of the additive genetic component σ_a^2 is slightly smaller than the CTD estimate, this difference is insignificant. Furthermore, the decomposition of cross-sectional inequality, which we report in Panel B, Column (3) of Table 2, suggests that there is very little change in the percentage of educational variance attributable to genetic factors in the extended CTD with dominance (32%) compared to the CTD (34%). These three insignificant differences support the CTD's assumption that genetic dominance does not play a significant role. This finding of

no dominance is consistent with Cesarini et al. (2009a) based on an ACE model of risk-taking and Okbay et al. (2022) based on a GWAS analysis of educational attainment.

4.3 Gene-Environment Interactions

The CTD imposes a simplistic division of influences into purely genetic and purely environmental (Heckman, 2007). This dichotomy ignores complex interactions between genes and environment and the potential that genes and environment could correlate (Plomin, DeFries, and Loehlin, 1977). To allow for gene-environment interactions—whereby the environment influences the expression of genetic traits—we exploit a change to the social environment experienced by the twins in our sample. Exploiting this exogenous environmental shift has the advantage of assessing how policy changes alter the heritability and environmental estimates within the same population. By examining the *change* in heritability estimates, we can provide insights into the potential policy relevance of heritability estimates from the ACE model (Rietveld, 2024). This focus on the change rather than the level of heritability estimates also mitigates the risk of misinterpreting heritability estimates as deterministic and immutable explanations for differences in socioeconomic outcomes, which could undermine efforts to address inequalities through social and economic policies (Goldberger, 1979; Goldberger and Manski, 1995; Manski, 2011).

Specifically, we focus on a 1937 school reform, which mandated municipalities, the local government units responsible for primary and lower secondary schools, to extend educational offerings to include eighth- and ninth-grade teaching. However, the implementation of this reform was gradual, so several municipalities had yet to offer eighth and ninth grades in 1958 when a second reform required all remaining municipalities to offer these grades. This phased, varied implementation of educational reform across municipalities creates a natural experiment, allowing us to observe the effects of environmental changes on genetic expression within our twin families.¹⁶

The 1937 Danish school reform impacted the cohorts attending school during the 1940s and 1950s, corresponding to the cohorts of the twins in our sample. To evaluate the impact of this educational expansion, we use the distance to the nearest school offering eighth and ninth-

¹⁶ While municipalities were required to offer eighth- and ninth-grade teaching, attendance was voluntary. Seven years of teaching was mandatory until a 1972 reform mandated nine years.

grade education as a measure of exposure. We determine this distance using the twins' birth year and the parish where their birth was registered.¹⁷ We categorize twin pairs into two groups according to whether their distance to the nearest relevant school at age 14—the typical age for enrolling in eighth grade—was above or below the median distance. Of the 33,914 twins in our main sample, information on the distance to the nearest school offering eighth and ninth grades is available for 32,124 twins.¹⁸ Using this information, we assign 16,756 twins to the group within each cohort most exposed to the environmental shift induced by the educational reform.

When interacted with the model's parameters, this binary partition of the twins' sample allows us to assess the effects of exposure to an exogenous shift of the educational environment on the components of variance in the CTD. Let $R_{f(i)}$ be a dummy indicating whether the twins born in family f were exposed (E) or not exposed (NE) to the educational reform. Allowing for interactions with the educational reform, equation (1) of the CTD model becomes:

$$y_i = \begin{cases} a_i^{NE} + c_{f(i)}^{NE} + e_i^{NE} & \text{if } R_{f(i)} = 0 \\ a_i^E + c_{f(i)}^E + e_i^E & \text{if } R_{f(i)} = 1 \end{cases} \quad (13)$$

In equation (13), each of the three factors that determine outcome y is allowed to change depending on whether family f was exposed to the educational expansion, resulting in a variance decomposition that depends on reform exposure:

$$\text{var}(y_i) = \begin{cases} \sigma_{aNE}^2 + \sigma_{cNE}^2 + \sigma_{eNE}^2 & \text{if } R_{f(i)} = 0 \\ \sigma_{aE}^2 + \sigma_{cE}^2 + \sigma_{eE}^2 & \text{if } R_{f(i)} = 1 \end{cases} \quad (14)$$

If there are gene-environment interactions, we should observe the genetic variance components σ_{aE}^2 and σ_{aNE}^2 of equation (14) to differ between the twins exposed to the schooling environment induced by the reform compared to their non-exposed counterparts. However, the contribution of the different factors in explaining the variance of an outcome in equation (14) may differ between the two groups even if there are no gene-environment interactions. That is, we may observe heritability to vary because the shared or the non-shared environmental components differ between the two groups defined by the exogenous shift of the educational environment.

¹⁷ Parish of residence is not registered until 1970. For calculating distance to school, we assume that individuals reside in their parish of birth registration at age 14. We calculate distances as the crow flies between the parish where the birth was registered and parishes offering eighth-grade teaching using parish church coordinates.

¹⁸ Parish of birth registration information is missing for Southern Jutland, where ecclesiastical administration is organized differently.

The parametrization of equations (13) and (14) fully interacts each of the three factors with the environmental change induced by the reform. This is a general (saturated) model that nests as special cases models where only some factors are allowed to shift. One such special case is a model where only the genetic component interacts with the reform indicator. This special case is the variance components' counterpart of regression models that include polygenic scores interacted with environmental measures.

4.3.1 Results

We report estimates of the additive genetic, shared environmental, and idiosyncratic variance components in Panel A, Column 1, of Table 3. The additive genetic variance component is marginally lower among families exposed to the reform than their non-exposed counterparts; however, this difference is not statistically significant (p -value = 0.566). In contrast, the environmental and idiosyncratic variance components for families exposed to the reform are considerably lower, and the difference to the non-exposed families is statistically significant (p -values= 0.000). These findings suggest that expanding education opportunities influences educational inequality mostly by limiting the contribution of shared and non-shared environmental factors rather than altering the contribution of the genetic component. This evidence does not support gene-environment interactions in the context of educational achievement, indicating that while the environment can be modified to reduce disparities, the contribution of genetic factors to education persists independently of these institutionally induced environmental changes. These findings are consistent with evidence of no gene-environment interactions reported by Black et al. (2020) on interactions between biological and adoptive parental wealth.

We report the decomposition of cross-sectional inequality for the exposed and non-exposed groups in Panel B, Column 1 of Table 3. For the exposed, who live closer to the affected schools, the environmental factors (especially the shared environment) account for a lower percentage of the variance in education compared to their non-treated counterparts who live further away. Conversely, genetic factors account for a larger percentage among the treated group. Given the evidence from Panel A, which indicates that the genetic variance component does not change across regimes, the increased proportion of the variance attributed to genetic factors among the treated is due to a decrease in the contribution of environmental factors to

the variance. As the reform diminishes the role of environmental differences, the relative proportion of variance attributable to genetics appears larger, not because the contribution of genetic factors has grown, but because the contributions of shared and idiosyncratic environments have decreased. These findings are consistent with the notion that as educational opportunities expand, diminishing environmental differences, genetic factors will account for a larger proportion of the variation in years of schooling (Trzaskowski et al., 2014; Mogstad and Torsvik, 2023).

This section's analytical framework can also be used to derive estimates of gene-environment interactions under the assumption that the environmental shift produced by the reform impacts only the genetic component within the CTD model, leaving other factors unchanged. This approach mimics empirical specifications used in GWAS studies, which explore gene-environment interactions by leveraging information on polygenic scores interacted with socioeconomic environment measures in regression models (e.g., Biroli et al., 2022). We report the results of this exercise in Columns 2 and 3 of Table 3. When we fix the shared and unique environmental components to be equal across different reform regimes, we find that the genetic variance component is significantly lower among the exposed twin families (p -value = 0.000). This finding leads to a corresponding reduction of heritability from 49% to 23%.¹⁹

4.4 Equal Environments Assumption

The CTD posits that MZ and DZ twins share the components of their environment that are not influenced by their genetics to the same degree (Evans and Martin, 2000). When MZ twins are treated more similarly by others (parents, peers, society) because they look alike, the EEA assumption is violated. For example, suppose parents are more likely to dress MZ twins similarly or expect them to have similar interests purely because they look alike. In that case, they create a more similar environment for MZ twins than for DZ twins, which is not a direct consequence of the twins' genetic makeup but instead of societal perceptions and reactions to

¹⁹ When we allow the genetic and environmental components of the CTD to shift with the educational reform, while keeping the idiosyncratic component fixed across reform regimes, we obtain a significant reduction in the genetic variance component (p -value=0.000) and a marginally significant decrease in environmental variance components (p -value=0.109). This restricted specification results in heritability decreasing from 47% to 28%. These findings underscore the importance of fully interacting the CTD variance components when testing for gene-environment interactions.

their physical similarity. Violating this EEA leads to overestimating genetic effects (see also Taubman, 1976).

4.4.1 Model and Identification

To allow for the possibility that the degree to which the environment is shared between twins may vary by zygosity, we use information on the twins, their spouses, and children so that $h(i) \in \{T_1, T_2, S_1, S_2, C_1, C_2\}$, where T represents the twins, S their spouses, and C their children. This twin family model leverages both horizontal and vertical relationships. Horizontal relationships include twins, cousins, and siblings-in-law, all within the same generation. Vertical relationships span different generations and include those between twins or their spouses, and their children or their “niblings” (nephews and nieces). More generally, besides zygosity, the degree to which the environment may be shared between twins or, indeed, between other relatives may also vary with the gender composition of the twin pair. Gender composition is a salient feature of the pair that may affect how parents and peers treat or perceive twins. We account for gender-related environmental variation in the model through gender composition, measured as the number of males in a pair of relatives, denoted by the index $g \in \{0, 1, 2\}$. Therefore, we relax the EEA of the CTD by allowing twins to draw the environmental factor $c_{f(i)}$ from zero mean distributions, whose variance ($\sigma_{c_{gz}}^2$) depends on zygosity (z) and gender (g) composition of the pair. The covariance for twins then becomes

$$\text{cov}(y_i, y_{i'})_{gz}^{T_1 T_2} = 0.5^{(1-z)} \sigma_a^2 + \sigma_{c_{gz}}^2 \quad . \quad (15)$$

Equation (15) encompasses five moment restrictions: two moment restrictions for MZ twin pairs, categorized by gender as two males or two females, and three moment restrictions for DZ twin pairs, including pairs of the same gender and mixed-gender pairs. The model defined by equations (2) and (15) is not identified because it has seven parameters (the idiosyncratic variance component σ_e^2 , the genetic factors variance component σ_a^2 , and the five shared environmental components $\sigma_{c_{gz}}^2$) but only six moment restrictions.

To identify this extended CTD, we use additional moment restrictions derived from intergenerational relationships, which include the covariances between twins (or their spouses) and their niblings. As in equation (15) for the twins, we allow for variation in the shared environments based on the twins' zygosity and the gender composition of the specific pairs. For example, an uncle who is an MZ twin may share a closer relationship and thus have a more

similar environment with his co-twin's children than a DZ twin uncle. The model also allows for same-gender pairs to experience more environmental sharing due to common interests and activities compared to mixed-gender pairs.

The covariance for twins and their niblings is given by

$$\text{cov}(y_i, y_{i'})_{gz}^{T_j C_k} = 0.5^{(1-z)} 0.5 \sigma_a^2 + \gamma_{Agz}, \quad j, k = 1, 2 \quad j \neq k \quad (16)$$

where γ_{Agz} parameterizes differential environmental sharing within avuncular relationships. The genetic relationship between children and their consanguineous uncles and aunts varies depending on whether their parent is an MZ or DZ twin. Children of MZ twins ($z = 1$) share genes with their consanguineous uncles and aunts—who are the MZ co-twins of their parents—to the same extent ($0.5\sigma_a^2$) as they do with their parents. However, for children of DZ twins ($z = 0$), the genetic sharing with their consanguineous uncles and aunts—who are the DZ co-twins of their parents—is only half as large ($0.25\sigma_a^2$). Equation (16) adds six moment restrictions (each combination of zygosity and gender composition of the avuncular pair) and six parameters (the environmental covariances γ_{Agz}) to the model.

The covariance for the twin's spouses and their niblings is given by

$$\text{cov}(y_i, y_{i'})_{gz}^{S_j C_k} = \gamma_{Agz}. \quad j, k = 1, 2 \quad j \neq k \quad (17)$$

The covariance between the spouses of twins and their niblings depends only on environmental sharing, as these spouses have no genetic relationship with the twins' nieces and nephews. This lack of genetic similarity arises because the spouses are not blood relatives of the twins' nieces and nephews unless there is assortative mating based on genetic factors.

Conditional on the gender composition of the pair (g) and the zygosity (z) of twins, our identifying equal environment assumption is that the degree of environmental sharing (γ_{Agz}) between twins and their niblings (equation 16) is the same as that between the spouses of twins and their niblings (equation 17). This version of the EEA posits that the environmental influences shared between, for instance, male MZ twins (consanguineous uncles) and their nephews are comparable to those shared between male spouses of female MZ twins (uncles by marriage) and their nephews. As a result, this EEA suggests that nephews and uncles experience the same environmental influences, whether the uncle is consanguineous or related by marriage. In our model, this EEA corresponds to the parameter γ_{A21} being the same in equations (16) and (17), where $g = 2$ refers to male pairs and $z = 1$ refers to the families of

MZ twins (see Appendix Table A1). Similarly, the environmental sharing between female MZ twins (consanguineous aunts) and their nieces is the same as that between female spouses of male MZ twins (aunts by marriage) and their nieces. This assumption corresponds to the parameter γ_{A01} being the same in equations (16) and (17), where $g = 0$ refers to female pairs and $z = 1$ refers to families of MZ twins. The same framework for equal environmental sharing extends to mixed-gender pairs ($g = 1$) and to families with DZ twins ($z = 0$), with a separate environmental parameter for each case.

Crucially, our identification allows for variability by zygosity and gender composition in the degree of environmental sharing between avuncular relationships, such as uncle-nephew and aunt-niece pairs, within each family. This variability means that we do not impose that uncles and nephews share environments to the same degree as aunts and nieces. We do not assume, for example, that MZ uncles share environments to the same degree with their nephews as DZ uncles. Therefore, this version of the EEA is weaker than that typically used in the CTD. The differences in treatment and perception that might bias results are less likely to be pronounced or influential in avuncular relationships compared to the more salient differences between MZ and DZ pairs.

Equation (17) provides sufficient additional information for identifying this extended CTD model because it adds six moment restrictions that correspond to different combinations of gender composition and zygosity without adding new parameters. As a result, the model now comprises 13 parameters (idiosyncratic variance, genetic variance, and 11 shared environmental components) and 18 moment restrictions (equation 2 and equations 15-17).²⁰

Since we add at least as many equations as parameters in the rest of the model, these additions do not affect identification. Including the children of twins allows us to estimate parent-child covariances in outcomes, contributing to the intergenerational elasticity (IGE), which measures the intergenerational transmission of socioeconomic outcomes.

²⁰ With five overidentifying moment restrictions we could still not fully relax our assumption that, conditional on zygosity and gender composition, the degree of environmental sharing between twins' spouses and their niblings is the same as that between twins and their niblings. Such relaxation would require inclusion of six additional environmental parameters for the twin spouse-nibling relationships of Equation (17). Nevertheless, we can test whether the data support partially relaxing the identifying assumption. Specifically, by maintaining the assumption for a given gender composition (i.e. a given value of g) of the uncle/aunt-nibling pair, we can test its validity for the other two gender compositions. The results of these tests, which are detailed in the next section, are supportive of our assumption, indicating that the environmental sharing between uncles/aunts and niblings does not significantly vary if uncles and aunts are twins or twins' spouses.

The covariance for parents and children is given by

$$\text{cov}(y_i, y_{i'})_g^{P_j C_j} = 0.5\sigma_a^2 + \gamma_{I_g}, \quad P = S, T \quad j = 1, 2 \quad (18)$$

where γ_{I_g} parameterizes differential environmental sharing in parent-child relationships. This covariance contains three moment restrictions, each depending on the gender composition of the parent-child pair (both male, both female or mixed-gender) and adds three environmental sharing factors to the parameter set. Given that parents always transmit, on average, 50% of their segregating genes to their children, these moment restrictions and the corresponding environmental sharing factors do not vary with the twins' zygosity.

The covariance for cousins is given by

$$\text{cov}(y_i, y_{i'})_{gz}^{C_1 C_2} = 0.5^{(1-z)} 0.25\sigma_a^2 + \gamma_{C_{gz}}, \quad (19)$$

where the degree of environmental sharing between cousins ($\gamma_{C_{gz}}$) depends on their gender composition and the zygosity of their twin parents, resulting in six moment restrictions (one for each combination of z and g).

This extended model also includes relationships between spouses and between siblings-in-law. Assuming no assortative mating based on genetic factors, these relationships are influenced solely by shared environments. Thus, the covariance for twins and their spouses, which does not depend on the zygosity of the twins, is given by

$$\text{cov}(y_i, y_{i'})^{T_j S_j} = \gamma_S, \quad j = 1, 2 \quad (20)$$

while the covariance for siblings-in-law—who are connected through being married to twins or being the co-twin's spouse—is given by

$$\left. \begin{array}{l} \text{cov}(y_i, y_{i'})_{gz}^{T_j S_k} \\ \vdots \\ \text{cov}(y_i, y_{i'})_{gz}^{S_1 S_2} \end{array} \right\} = \gamma_{SL_{gz}} \cdot \quad j, k = 1, 2 \quad j \neq k \quad (21)$$

4.4.2 Results

Cross-Sectional Inequality

Table 4 presents parameter estimates from the twin family model for educational inequality, in which we extend the CTD by relaxing the EEA. The parameter estimates in Column (1) show substantial heterogeneity in environmental sharing across different familial relationships. We

find that shared environments significantly differ between MZ and DZ twins, rejecting the EEA, with a p-value of 0.011 for male twins and 0.006 for female twins.

As we discussed in Section 4.4.1, we identify different environmental parameters for MZ and DZ twins based on the assumption that the degree of environmental sharing between twins and their niblings is the same as that between twins' spouses and their niblings, conditional on the gender composition of the pair and the zygosity of the twins. That is, we estimate six shared environmental parameters, one for each combination of gender composition and zygosity (γ_{Agz} , for $g \in \{0, 1, 2\}$ and $z \in \{0, 1\}$), and we assume that each of these parameters is the same for avuncular pairs that include twins (the uncle or aunt is consanguineous) and avuncular pairs that include twin spouses (the uncle or aunt is by marriage).

Because the model is overidentified, we can also estimate versions in which we maintain the EEA for only one gender combination of the avuncular pair and test its validity for the other two gender combinations. For example, we can maintain the assumption for only mixed-gender pairs in MZ families (γ_{A11}) and allow the shared environmental parameter for male pairs to differ between consanguineous (γ_{A21}) and by marriage ($\tilde{\gamma}_{A21}$) uncles, and the shared environmental parameter for female pairs to differ between consanguineous (γ_{A01}) and by marriage aunts ($\tilde{\gamma}_{A01}$). We can then partially test our identifying assumption that shared environments do not differ for avuncular pairs in which the uncle (aunt) is consanguineous or by marriage by testing that $\gamma_{A21} = \tilde{\gamma}_{A21}$ and $\gamma_{A01} = \tilde{\gamma}_{A01}$. Similarly, we can maintain the assumption for only male pairs or for only female pairs in MZ families and allow the shared environmental parameters to differ for the other two gender combinations. Extending the same tests for DZ families, we have overall 6 different tests of the identifying assumption in which we never reject the hypothesis that shared environmental parameters between avuncular pairs that include twins and avuncular pairs that include twin spouses are the same, with p-values in the range of 0.3-0.8.

While results from the twin family model with heterogeneous shared environmental components in this section point to the rejection of the CTD's EEA, the evidence from Section 4.1 highlights the importance of assortative mating on genes. To assess which of these assumptions is more consequential for our analysis, we report the variance components in Column (2) of Table 4 after calibrating the heterogeneous environment model to allow for assortative mating based on genetic factors, using the 0.12 correlation between spouses

reported in Table 2.²¹ The similarity of variance components to those in Column (1), obtained under the assumption of no assortative mating, suggests that the observed significant assortative mating on genes under the common environment assumption, discussed in Section 4.1, may result from not fully accounting for EEA violations.

We present the decomposition of educational inequality from the twin family model in Column (1) of Table 5, which allows shared environments to vary by zygosity and gender.²² Heritability accounts for 9% of the total variance, with shared environments explaining 42% and the remaining variance attributed to non-shared environmental factors. Column (2) shows that allowing for assortative mating reduces only marginally the heritability estimate to 8%. These findings from the twin family model demonstrate that CTD's EEA is overly restrictive, leading to a biased variance decomposition that overestimates the contribution of additive genetic factors. This evidence of violation of the CTD's EEA, leading to overestimation of heritability, is consistent with Björklund, Jäntti, and Solon (2005), who decompose the variance in earnings and find also substantially lower heritability estimates when they allow shared environments to vary across sibling types. Our heritability estimate is also consistent with the evidence by Collado, Ortuño-Ortín, and Stuhler (2023), who, using an extended family design, find about 7% of the variance in years of schooling is explained by variation in latent genetic factors.

This evidence of bias in the CTD also contributes to the debate on “missing heritability” in molecular genetics studies, which refers to the observation that genetic variants identified through GWAS explain only a fraction of the heritability of complex traits estimated by traditional methods, such as family and twin studies (Maher, 2008; Eichler, Flint, and Gibsonet, 2010; Young, 2019). For example, recent evidence from Okbay et al. (2022) indicates that the educational attainment polygenic score explains 12-16% of the variance in educational attainment. This percentage includes direct genetic effects on educational attainment (heritability) and indirect effects from inherited socioeconomic and environmental factors

²¹ In the model of this Section with fully flexible environmental variation, we calibrate assortative mating because the siblings-in-law moment restrictions, which identify assortative mating in the simpler model discussed in section 4.1, introduce four distinct environmental parameters, each corresponding to a different combination of zygosity and gender composition, which implies that we cannot identify separately the assortative mating parameter.

²² For the decomposition of cross-sectional education inequality, we use equation (2) substituting the common environmental component σ_c^2 with the average environmental component across zygosity and gender. In Appendix Table A1, Column (1), we also report gender-specific inequality decompositions.

(Trejo and Domingue, 2018). Okbay et al. (2022) find that 31% of the explanatory power of the polygenic score on the variance of educational attainment is attributable to direct genetic effects, which implies that heritability explains 4-5% of the variance of educational attainment, a much lower estimate than those found in studies using the CTD (around 40%).²³

Intergenerational Elasticity, Siblings and Other Relatives Correlations

An important feature of the twin family design is that it enables us to decompose both cross-sectional educational inequality—a within-generation measure—and the IGE of education into its environmental and genetic factors. The IGE is given by

$$\beta = \frac{0.5\sigma_a^2 + \bar{y}_I}{\sigma_a^2 + 0.5(\bar{\sigma}_c^2 + \bar{\sigma}_s^2) + \sigma_e^2} = \underbrace{\frac{0.5\sigma_a^2}{\sigma_a^2 + 0.5(\bar{\sigma}_c^2 + \bar{\sigma}_s^2) + \sigma_e^2}}_{\text{genetic component}} + \underbrace{\frac{\bar{y}_I}{\sigma_a^2 + 0.5(\bar{\sigma}_c^2 + \bar{\sigma}_s^2) + \sigma_e^2}}_{\text{environmental component}}, \quad (22)$$

where σ_s^2 is the variance of shared environment for spouses and overbars denote the average of environmental parameters over gender (for twins, twins' spouses, and their children) and zygosity (for twins). We report the IGE of education in Table 5, which equals 0.24. The decomposition suggests that shared environmental influences account for 81% of the IGE of education. This finding aligns with decompositions from GWAS studies, which indicate that 75 to 85% of the parent-child educational correlation in the U.S. (Conley et al., 2015; Liu, 2018) and Norway (Isungset et al., 2022) reflects environmental transmission.

More generally, the moment restrictions of the model provide a gene-environment decomposition of the *covariances* in education among family members. Scaling with the variance, we can decompose the education correlations of relatives. For example, we can decompose the twins' correlation as follows

$$\text{corr}(y_i, y_{i'})_{gz}^{T_1 T_2} = \frac{0.5^{(1-z)}\sigma_a^2 + \sigma_{c_gz}^2}{\sigma_a^2 + \sigma_{c_gz}^2 + \sigma_e^2} = \underbrace{\frac{0.5^{(1-z)}\sigma_a^2}{\sigma_a^2 + \sigma_{c_gz}^2 + \sigma_e^2}}_{\text{genetic component}} + \underbrace{\frac{\sigma_{c_gz}^2}{\sigma_a^2 + \sigma_{c_gz}^2 + \sigma_e^2}}_{\text{environmental component}} \quad (23)$$

Table 6 presents an overview of correlation decompositions for all relative types. Columns (1) and (2) show that the predicted educational correlations closely match the empirical correlations. This close match is expected, given the many pair-specific

²³ Kong et al. (2018) using Icelandic data find that the polygenic score explains 5% of the variance of educational attainment, with direct effects explaining only 2.5%. Using the same data but estimating heritability using a method that examines how much more or less related a pair is than would be expected from the relatedness of the parents, Young, Frigge, and Gudbjartsson (2018) find an estimate of heritability equal to 17%.

environmental sharing parameters.²⁴ Columns (3)-(5) of Table 6 report the proportions of the educational correlations attributed to shared environments, with each column calibrating the model with a different degree of assortative mating. In the absence of assortative mating ($\delta=0$), shared environments account for at least 80% of the predicted educational correlation. For MZ twins, the proportion of educational correlation accounted for by shared environment is about 85%, while it is around 90% for DZ twins.²⁵ Generally, the proportions of shared environment tend to be larger in DZ families than in MZ ones and do not appear to vary much with gender composition within zygosity.

When we calibrate the model using the 0.12 correlation of genetic factors among spouses (reported in Table 2), we find that the proportion of the predicted educational correlation attributed to environmental factors changes slightly (Column 4). This calibration exercise supports our earlier findings from decomposing cross-sectional inequality using the twin family model, regardless of whether we account for assortative mating.

Another fact to note about the calibration exercise of Column (4) is that additive genetic factors contribute merely 2% to the spousal correlation in years of education. This contribution differs substantially from the 13% contribution found using the CTD model accounting for assortative mating (see footnote 15). The finding that genetic assortative mating loses relevance after allowing for environmental heterogeneity suggests that the findings from the CTD augmented with assortative mating largely reflect the omission of environmental differences across pairs connected by varying degrees of genetic resemblance under assortative mating, e.g., spouses and siblings-in-law.

Finally, we find that assortative mating has the largest effect in decomposing the educational correlation between cousins. However, as shown in Column (5) of Table 6, affecting correlations for other relatives requires calibrating the assortative mating correlation to 0.30. This degree of assortative mating would imply that spouses share genetic factors

²⁴ To assess the predictive performance of our model, we randomly select half of the twin families, re-estimate the model with this subsample, and predict educational correlations. We then compare these predictions with their empirical counterpart estimated from the other half of the sample. The results of this exercise, reported in Appendix Table A3, show that the 95% confidence interval of the predicted correlations from the first subsample always contains the empirical correlations estimated from the second subsample.

²⁵ Fletcher et al. (2023) residualize measures of educational attainment on polygenic scores using data from the UK Biobank, finding that polygenic scores account for 20% of the sibling correlation in education.

related to education more than cousins with MZ parents, leading us to consider the 0.30 assortative mating correlation as an upper bound for this calibration exercise.

5. Results for Economic Outcomes

In this section, we estimate the twin family model for earnings, income, and wealth, which decomposes cross-sectional inequality and intergenerational persistence in market outcomes. Extending our analysis beyond educational attainment, we aim to shed light on broader economic dimensions that shape individuals' financial status and mobility across generations. In Section 5.1, we describe the data; in Section 5.2, we discuss the decomposition results for cross-sectional inequality; in Section 5.3, we present the results for the decomposition of the (rank-rank) intergenerational elasticity.

5.1 Data Description

We observe annual pre-tax labor earnings from income tax returns. Employers report earnings for each employee to the tax authorities, who send these reports to the employees every March for verification of earnings in the previous calendar year. Using the Statistics Denmark Income Statistics Register (Baadsgaard and Quitzau, 2011), we calculate the sum of earnings from all employment each year from 1980-2018.

Given the size of the twin population, we cannot estimate a model of permanent earnings with multi-person life-cycle dynamics (e.g., Bingley, Cappellari, and Tatsiramos, 2021). Therefore, we follow a simpler empirical approach: to measure permanent earnings as the average earnings percentile over ages 30 to 50.²⁶ Averaging individual earnings over this specific age range is an intuitive way of separating permanent earnings from transitory shocks, helping mitigate life-cycle biases (Haider and Solon, 2006; Bohlmark and Lindquist, 2006). Intergenerational analyses are especially prone to life-cycle bias (Nybom and Stuhler, 2016; Bingley and Cappellari, 2019). By using earnings percentiles rather than earnings levels, we also avoid the biases associated with modeling fluctuations of earnings levels in the cross-section and over the life cycle.

²⁶ To calculate permanent earnings, we require at least five annual earnings observations for each person. Percentiles are first estimated by year, and then recentered by gender and person type (whether a twin, child or spouse).

We also use average percentiles between ages 30 and 50 for permanent income and wealth.²⁷ We calculate disposable income by subtracting taxes from personal gross income and adding transfers. We measure wealth by using taxable assets. The wealth register contains information on end-of-year financial assets, non-financial assets, and liabilities.²⁸ Because real estate capital offsets most debt, the value of assets is very close to the value of liabilities for about half of households. To abstract from this collateralization, we ignore debt and consider the sum of total financial and non-financial assets.²⁹ As registration of pension wealth from defined contribution plans began in 2012, we excluded these plans from our measures for consistency.³⁰

5.2 Cross-Sectional Inequality

As for education, we first estimate the CTD model and use the parameter estimates to decompose cross-sectional inequality for earnings, income, and assets. Panel A of Table 7 shows that additive genetic factors account for 60% of the variation in earnings, 56% in disposable income, and 42% in assets. These results align with those from the analysis of educational inequality using the CTD (shown in Panel B of Table 2), where additive genetic factors were also found to explain a much larger proportion of cross-sectional inequality compared to shared environments. Our findings for assets, applying the ACE model using twins, are broadly in line with those of Fagereng, Mogstad, and Rønning (2021), who, applying the ACE model using Norwegian biological and adopted siblings, report a heritability estimate of 33% for financial assets, with shared environmental factors accounting for 9% of the variance.

We report the cross-sectional decomposition from the twin family design in Panel B of Table 7. Relaxing the CTD's EEA reduces the contribution of additive genetic factors to cross-

²⁷ Requiring at least five observations in the 30-50 age range and excluding cases with missing information reduces the sample size for income and wealth (with 79,255 and 79,259 individuals, respectively) compared to the sample used for analyzing educational inequality (80,205 individuals). For earnings, the sample is reduced to 72,417 individuals, with the loss concentrated among females in the parental generation.

²⁸ Financial institutions report wealth to the tax authorities. Until 1996, the authorities used this information for wealth taxation. However, since the abolition of the wealth tax, the authorities have used these wealth reports to check whether income is consistent with net wealth changes (see, e.g., Jakobsen et al., 2020).

²⁹ See Boserup, Kopczuk, and Kreiner (2018) for the importance of wealth and assets over the life cycle in Denmark.

³⁰ Appendix Table A4 presents descriptive statistics for earnings, income, and wealth.

sectional inequality: for earnings, the percentage drops from 60% to 16%; for disposable income, from 56% to 14%; and for assets, from 42% to 14%. Similar to our findings for education, this evidence suggests that the equal environments assumption of the CTD substantially overstates the influence of genetic factors on the inequality of earnings, income, and assets. Our decompositions reveal that contrary to the conclusions reached with the CTD, the contribution of shared environmental factors is much larger than that of genetic factors in explaining the inequality of socioeconomic outcomes. Specifically, shared environments explain 26% of the variance in earnings, income, and assets. The heritability estimate for earnings is relatively close to the 20% estimate in Björklund, Jäntti, and Solon (2005), although we find relatively higher estimates for the importance of shared environments (26% versus 16%, respectively).

We summarize this evidence in Figure 1, which shows that heritability accounts for 14% to 16% of cross-sectional inequality observed in earnings, income, and assets. These estimates suggest that additive genetic factors explain a higher portion of these outcomes' variance than education, where heritability accounts for only 9%. On the other hand, shared environmental factors explain slightly more than one-third of the variation in earnings, income, and assets, in contrast to contributing to half of the education variance.

5.3 Intergenerational Elasticity

Using the twin family model estimates, we report in Panel B of Table 7 the decomposition of the IGE of earnings rank, income rank, and assets rank. We calculate the IGE of earnings and income to be 0.18 and the IGE of assets to be 0.25. Shared environmental factors account for 55% of the correlation in earnings, 61% in income, and 71% in assets. Figure 2 summarizes these findings for all outcomes, highlighting that shared environmental factors account for a higher percentage (81) of the intergenerational correlation in education.³¹ These findings are consistent with the evidence from the adoptee studies that the family environment accounts for half or more of the IGE for these outcomes (e.g., Björklund, Lindahl, and Plug, 2006; Holmlund, Lindahl, and Plug, 2011; Black et al., 2020; Fagereng, Mogstad, and Rønning, 2021).

³¹ In Appendix Table A2, Columns (2)-(4), we report decompositions of cross-sectional inequality and IGEs by gender.

We also extend the analysis to different types of family relationships and report in Table 8 the empirical and predicted correlations and the proportion attributed to shared environments. We observe that correlations for earnings, income, and assets generally tend to be lower than for education, a more pronounced pattern for non-twin pairs.³² The contribution of shared environmental factors to the correlations of outcomes other than education is lower, ranging between 45 and 85% (compared with a contribution of shared environment consistently above 80% in the case of education). This evidence confirms the findings from the IGE decomposition that the contribution of genetic factors is greater for these market outcomes. Moreover, while the variation of shared environmental factors for education is mainly associated with zygosity, for market outcomes, shared environmental factors are more sensitive to gender composition. In particular, shared environmental factors are generally higher for male pairs than for female or mixed-gender pairs. This pattern may reflect gender differences in participation in labor and capital markets independent of shared environments.

6. Conclusions

Twin studies, frequently used in economics to decompose cross-sectional inequality into genetic and environmental factors, have faced critiques for relying on restrictive assumptions that can bias their variance component estimates (Goldberger, 1979). In contrast, adoptee studies have gained favor among economists, offering valuable insights into the intergenerational transmission aspect of the nature versus nurture debate. In this paper, we develop a unified framework by linking twins with their spouses and children to assess genetic and environmental factors' influence on inequality within and across generations.

Using this framework, we test the assumptions invoked in twin studies, combining data from the Danish Twin Registry with population registers. We find no evidence of genetic dominance, but the equal environment assumption is rejected, leading to an overestimation of the influence of genetic effects. Allowing for differentially shared environments between twin pairs, shared environmental factors account for a substantial proportion of the cross-sectional variance (26% to 42%) and the intergenerational correlation (55% to 81%) of education,

³² Correlations between twins and their spouses are lower for disposable income than for earnings and assets. This difference may reflect that disposable income includes transfers that may be effectively gender-specific. For instance, parental leave, which might be taken mostly by women in the parent cohorts of our study.

earnings, income, and wealth. These contributions surpass those of genetic factors, responsible for only 9% to 16% of the cross-sectional variance and 19% to 45% of the intergenerational correlations, and remain robust when considering assortative mating.

We also examine gene-environment interactions using an educational reform as a natural experiment to determine if genetic factors vary with exposure to improved educational opportunities. In a fully saturated model that interacts genetic, shared, and non-shared components with the exogenous environmental change, we find that the additive genetic component of the variance in years of schooling remains constant. However, shared and non-shared environmental components are significantly lower for those exposed to the schooling expansion.

These findings underscore the substantial role of shared environments in shaping socioeconomic inequality within and across generations, bridging the gap between estimates from various methodologies, including twin and adoptee ACE studies, intergenerational regressions on adoptees, and recent socio-genomics research. Our finding that twin and adoptee studies overestimate heritability due to violating the equal environment assumption contributes to the “missing heritability” debate, indicating that the discrepancy between the extended ACE model and GWAS findings is smaller than previously thought.

By extending the classic twin design to include spouses and children, we can estimate its parameters without relying on the restrictive assumptions traditionally associated with twin studies. This advance, coupled with the growing availability of multi-generational surveys and administrative datasets, especially those with genetic information, provides social scientists with a robust framework for understanding the origins of inequality. Incorporating additional relatives—such as non-twin siblings, half-siblings, step-siblings, and grandparents—into our model would offer a more comprehensive understanding of the role of shared environments in explaining inequality. These extensions represent promising avenues for future research.

REFERENCES

- Adermon, A., M. Lindahl and M. Palme (2021). Dynastic Human Capital, Inequality, and Intergenerational Mobility. *American Economic Review*, 111 (5), 1523-48.
- Baadsgaard, M., and J. Quitzau (2011). Danish Registers on Personal Income and Transfer Payments. *Scandinavian Journal of Public Health*, 39(7), 103-105.
- Barnea, A., H. Cronqvist and S. Siegel (2010). Nature or Nurture: What Determines Investor Behavior? *Journal of Financial Economics*, 98, 583-604.
- Barth, D., N.W. Papageorge and K. Thom (2020). Genetic Endowments and Wealth Inequality. *Journal of Political Economy*, 128(4), 1474-1522.
- Becker, G.S., and N. Tomes (1979) An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87(6), 1153-1189.
- 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.
- Behrman, J.R. and M.R. Rosenzweig (2002). Does Increasing Women’s Schooling Raise the Schooling of the Next Generation? *American Economic Review*, 92(1), 323–334.
- Benjamin, D.J., D. Cesarini, P. Turley and A.S. Young (2024). Social-Science Genomics: Progress, Challenges, and Future Directions, NBER Working Paper No. 32404.
- Bingley, P., and L. Cappellari (2019). Correlation of brothers’ earnings and intergenerational transmission. *Review of Economics and Statistics*, 101(2), 370–383.
- Bingley, P., L. Cappellari and K. Tatsiramos (2021). Family, Community, and Long-Term Socio-Economic Inequality: Evidence from Siblings and Youth Peers. *Economic Journal*, 131(636), 1515–1554.
- 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.
- Björklund, A., M. Lindahl and E. Plug (2006). The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data. *Quarterly Journal of Economics*, 121(3), 999-1028.
- Blanden, J., M. Doepke and J. Stuhler (2023). Educational Inequality. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.) *Handbook of the Economics of Education*. North-Holland, 405-497.

- Black, S.E., Devereux, P.J. (2011). Recent Developments in Intergenerational Mobility. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4B, pp. 1487-1541). Elsevier.
- Black, S.E., P.J. Devereux, P. Lundborg, and K. Majlesi (2020). Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours. *Review of Economic Studies*, 87(4), 1683–1725.
- Boserup, S.H., W. Kopczuk and C.T. Kreiner (2018). Born with a Silver Spoon? Danish Evidence on Wealth Inequality in Childhood. *Economic Journal*, 128(602), F514-F544.
- Bowles, S. and H. Gintis (2002). The Inheritance of Inequality. *Journal of Economic Perspectives*, 16(3), 3–30.
- Böhlmark, A. and M.J. Lindquist (2006). Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden. *Journal of Labor Economics*, 24(4), 879–900.
- Branigan, A.R., K.J. McCallum and J. Freese (2013). Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Social Forces*, 92(1), 109-40.
- 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.
- Christiansen L., H. Frederiksen, K. Schousboe, A. Skytthe, N. von Wurmb-Schwark, K. Christensen et al. (2003). Age- and sex differences in the validity of questionnaire-based zygosity in twins. *Twin Research*, 6(4), 275–278.
- Cronqvist, H. and S. Siegel (2015). The Origins of Savings Behavior. *Journal of Political Economy*, 123(1), 123-169.
- 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.
- Conley, D., B.W. Domingue, D. Cesarini, C. Dawes, C.A. Rietveld and J.D. Boardman. (2015). Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?. *Sociological Science*, 2, 82-105.
- D’Onofrio, B. M., E., Turkheimer, L.J. Eaves, L. A., Berg, K., M.H. Solaas, et al. (2003). The Role of the Children of Twins Design in Elucidating Causal Relations Between Parent Characteristics and Child Outcomes. *Journal of Child Psychology and Psychiatry*, 44, 1130–1144.
- Eichler, E., J. Flint, G. Gibson et al. (2010). Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease. *Nature Review Genetics*, 11, 446–450.

- Eika, L., M. Mogstad and B. Zafar (2019). Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy*, 127(6), 2795-2835.
- Eley, T. T., T. A. McAdams, F. V. Rijsdijk, P. Lichtenstein, J. Narusyte, D. Reiss, E. L. Spotts, J. M. Ganiban, J. M. Neiderhiser (2015). The intergenerational transmission of anxiety: a children-of-twins study. *American Journal of Psychiatry*, 172(7), 630-637.
- Evans D.M. and N.G. Martin (2000). The Validity of Twin Studies. *Gene Screen*, 1(2), 77–79.
- Fagereng, A., L. Guiso and L. Pistaferri (2022). Assortative Mating and Wealth Inequality. *NBER Working Paper No. 29903*.
- Fagereng, A., M. Mogstad and M. Rønning (2021). Why Do Wealthy Parents Have Wealthy Children? *Journal of Political Economy*, 129(3), 703-756.
- Felson J. (2014). What Can We Learn from Twin Studies? A Comprehensive Evaluation of the Equal Environments Assumption." *Social Science Research*, 43, 184–199.
- Fletcher, J., Q. Lu, B. Mazumder, and J. Song (2023). Understanding Sibling Correlations in Education: Molecular Genetics and Family Background. IZA Discussion Paper No. 15862.
- Freese, J. And Y.-H. Jao (2017). Shared Environment Estimates for Educational Attainment: A Puzzle and Possible Solutions. *Journal of Personality*, 85(1), 79–89.
- Gonalons-Pons, P., and C.R. Schwartz (2017). Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage? *Demography*, 54(3), 985-1005.
- Goldberger, A. S. (1979). Heritability. *Economica*, 46(184), 327–347.
- Goldberger, A. S. and C.F. Manski (1995). Review Article: The Bell Curve by Herrnstein and Murray. *Journal of Economic Literature*, 33(2), 762–776.
- Greenwood, J., N. Guner, G. Kocharkov and C. Santos (2014). Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, 104(5), 348–53.
- Güell, M, J.V. Rodríguez Mora and C.I. Telmer (2015). The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. *Review of Economic Studies*, 82(2), 693–735.
- Haider, S.J. and G. Solon (2006). Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, 96(4), 1308–1320.
- Heckman, J. J. (2007). The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences*, 104(33), 13250–13255.
- Holmlund, H., M. Lindahl, and E. Plug. (2011). The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods. *Journal of Economic Literature*, 49(3), 615-51.

- Houmark, M.A., V. Ronda and M. Rosholm (2024). The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills. *American Economic Review*, 114(2), 385-425.
- Isungset, M.A., D. Conley, H.D. Zachrisson, E. Ystrom, A. Havdahl, P.R. Njølstad and T. H. Lyngstad (2022). Social and Genetic Associations with Educational Performance in a Scandinavian Welfare State. *Proceedings of the National Academy of Sciences*, 119(25), 2201869119.
- Karlson, K., and R. Landersø (2012). The Making and Unmaking of Opportunity: Educational Mobility in 20th Century-Denmark. *IZA Discussion Paper* 14135.
- Kendler, Kenneth S. (2001). Twin Studies of Psychiatric Illness: an Update. *Archives of General Psychiatry*, 58(11), 1005-1014.
- Kim, YK. (2009). *Handbook of Behavior Genetics*. Springer New York, NY.
- Jakobsen, K., K. Jakobsen, H. Kleven and G. Zucman (2020). Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark. *Quarterly Journal of Economics*, 135(1), 329–388.
- Jensen, V.M. and A.W. Rasmussen (2011). Danish Education Registers. *Scandinavian Journal of Public Health*, 39(7), 91-94.
- Keller M.C. and W.L. Coventry (2005). Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design. *Twin Research and Human Genetics*, 8(3), 201-213.
- Keller, M.C., S.E. Medland, L.E. Duncan, P.K. Hatemi, M.C. Neale, H.H.M. Maes, and L.J. Eaves (2009). Modeling Extended Twin Family Data I: Description of the Cascade Model. *Twin Research and Human Genetics*, 12(1), 8-18.
- Kong A., G. Thorleifsson, M.L. Frigge, B.J. Vilhjalmsson, A.I., T.E. Thorgeirsson, . . . K. Stefansson (2018). The Nature of Nurture: Effects of Parental Genotypes. *Science*, 359, 424–428.
- Lundborg, P., E. Plug and A.W. Rasmussen (2021). On the Family Origins of Human Capital Formation: Evidence from Donor Children. *IZA Discussion Paper* No. 14708.
- Maher B. (2008). Personal Genomes: The Case of the Missing Heritability. *Nature*, 456, 18–21.
- Manski, C. F. (2011). Genes, Eyeglasses, and Social Policy. *Journal of Economic Perspectives*, 25(4), 83-94.
- Mogstad, M. and G. Torsvik (2023). Family Background, Neighborhoods, and Intergenerational Mobility. In: *Handbook of the Economics of Family*, (Eds.) Shelly Lundberg and Alessandra Voena, Amsterdam: North Holland, Vol 1, Chapter 7.
- Nance, W.E. and L.A. Corey (1976). Genetic Models for the Analysis of Data from the Families of Identical Twins, *Genetics*, 83(4), 811–826.

- Neale, M.C. (2009). Biometrical Models in Behavioral Genetics. In: Kim, YK. (eds) *Handbook of Behavior Genetics*. Springer, New York, NY.
- Nichols R.C. (1978). Twin Studies of Ability, Personality, and Interests. *Homo*, 29, 158-173.
- Nybohm, M. and J. Stuhler (2016). Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation. *Journal of Human Resources*, 51(1), 239-268.
- 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.
- Papageorge, N.W. and K. Thom (2019). Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study. *Journal of the European Economic Association*, 18(3), 1351-1399.
- Plomin R., J.C. DeFries and J.C. Loehlin (1977). Genotype-Environment Interaction and Correlation in the Analysis of Human Behavior. *Psychological Bulletin*, 84, 309–322.
- Plomin R. (1988). The Nature and Nurture of Cognitive Abilities. In R. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 4, pp. 1-33). Lawrence Erlbaum Associates, Hillsdale, NJ.
- Polderman, T.J.C., B. Benyamin, C.A. de Leeuw, P.F. Sullivan, A. van Bochoven, P.M. Visscher and D. Posthuma (2015). Meta-analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies, *Nature Genetics*, 47, 702-709.
- Rietveld, C.A., S.E. Medland, J. Derringer, J. Yang, T. Esko, N.W. Martin, H-J. Westra, K. Shakhbazov, A. Abdellaoui, A. Agrawal et al. (2013). GWAS of 126,559 individual identifies genetic variants associated with educational attainment. *Science*: 1235488
- Rietveld, C.A. (2024). Heritability and Public Policy Reconsidered, Again. *Tinbergen Institute Discussion Papers*, 2024-012/V.
- Sacerdote, B. (2002). The Nature and Nurture of Economic Outcomes. *American Economic Review*, 92(2), 344-348.
- 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.
- Sacerdote, B. (2011). Nature and Nurture Effects on Children’s Outcomes: What Have We Learned from Studies of Twins and Adoptees? In: *Handbook of Social Economics*, (Eds.) Jess Benhabib, Matthew Jackson and Alberto Bisin, Amsterdam: North Holland, Vol 1A, Chapter 1.
- Skytthe A., K. Kyvik, N.V. Holm, J.W. Vaupel and K. Christensen (2002). The Danish Twin Registry: 127 Birth Cohorts of Twins. *Twin Research*, 5(5), 352–7.
- Taubman, P. (1976). The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins. *American Economic Review*, 66(5), 858–870.

- Young, A.I., M.L, Frigge, D.F. Gudbjartsson *et al.* (2018). Relatedness Disequilibrium Regression Estimates Heritability without Environmental Bias. *Nature Genetics*, 50, 1304–1310.
- Willoughby, E. A., M. McGue, W.G. Iacono and J. J. Lee (2021). Genetic and environmental contributions to IQ in adoptive and biological families with 30-year-old offspring. *Intelligence*, 88, 101579.

Table 1. Sample Descriptive Statistics

	MZ Families		DZ Families	
	Men	Women	Men	Women
	Number of individuals			
Twin	3782	3551	14039	12542
Spouse	2921	2964	10220	10577
Child	2070	1912	7932	7694
	Average year of birth			
Twin	1951	1954	1950	1952
Spouse	1951	1953	1949	1952
Child	1972	1973	1972	1972
	Years of education			
Twin	12.3	12.3	11.9	11.8
	(3.0)	(2.9)	(3.1)	(3.0)
Spouse	12.6	12.0	12.4	11.9
	(2.9)	(2.9)	(3.1)	(2.9)
Child	13.4	13.9	13.5	13.8
	(2.5)	(2.4)	(2.4)	(2.4)

Note: Standard deviations are in parentheses.

Table 2. Parameter Estimates and Decomposition of Cross-Sectional Education Inequality - Classic Twin Design and Extensions.

	(1) Classic Twin Design (CTD)		(2) CTD with Assortative Mating		(3) CTD with Genetic Dominance	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Panel A. Variance Components						
Additive Genetic	419.97	51.09	495.66	59.45	343.06	89.24
Assortative Mating on Genes			0.12	0.03		
Genetic Dominance					51.28	58.06
Shared Environment						
Twins	297.68	33.27	167.30	52.55	323.32	38.28
Spouses			140.77	54.51		
Parent-Child					131.56	44.15
Uncle/Aunt-Nephew/Niece					118.41	28.10
Cousins					101.21	17.20
Twins-spouses			439.25	19.98		
Siblings-in-law			316.88	14.75		
Idiosyncratic Environment	516.98	20.60	516.48	20.75	516.98	20.60
Panel B. Total Variance and Variance Proportions						
Total Variance	1234.63	8.36	1234.63	8.36	1234.63	8.36
Proportion Genetic	0.34	0.04	0.45	0.05	0.32	0.04
Proportion Shared Env.	0.24	0.03	0.14	0.04	0.26	0.03
Proportion Idiosyncratic Env.	0.42	0.02	0.41	0.02	0.42	0.02

Note: The table reports parameter estimates and standard errors of the variance components in Panel A and the decomposition of education cross-sectional inequality in Panel B. Column (1) refers to the CTD using twin pairs; Column (2) refers to the extended version of the CTD when we allow for assortative mating based on genetic factors; Column (3) refers to the extended version of the CTD when we allow for genetic dominance; Column (4) refers to the extended version of the CTD when we allow for gene-environment interactions.

Table 3. Gene-environment interactions - Parameter Estimates and Decomposition of Cross-Sectional Educational Inequality

	(1) Full interaction		(2) Only genes interacted		(3) Only genes and shared env. interacted	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Panel A. Variance Components						
Additive Genetic						
Exposed	449.44	57.16	228.13	52.88	272.60	51.05
Non-Exposed	511.84	92.48	733.16	57.96	688.69	71.28
Difference	-62.40	108.72	-505.03	22.16	-416.09	59.60
Shared Environment			253.96	34.69		
Exposed	125.51	37.21			213.93	34.20
Non-Exposed	382.40	58.57			293.98	49.85
Difference	-256.89	69.39			-80.05	49.46
Idiosyncratic Environment			497.92	22.32	497.92	22.32
Exposed	365.29	22.40				
Non-Exposed	630.55	38.62				
Difference	-265.26	44.65				
Panel B. Total Variance and Variance Proportions						
Total Variance - Exposed	940.24	10.79	980.01	12.82	984.45	13.16
Proportion Additive Genetic	0.48	0.06	0.23	0.05	0.28	0.05
Proportion Shared Environment	0.13	0.04	0.26	0.04	0.22	0.03
Proportion Idiosyncratic Environment	0.39	0.02	0.51	0.02	0.51	0.02
Total Variance – Non-Exposed	1524.80	12.67	1485.04	14.82	1480.59	15.00
Proportion Additive Genetic	0.34	0.06	0.49	0.04	0.47	0.05
Proportion Shared Environment	0.25	0.04	0.17	0.02	0.20	0.03
Proportion Idiosyncratic Environment	0.41	0.03	0.34	0.02	0.34	0.02

Note

Table 4. Parameter Estimates for Twin Family Design

	(1) Twin Family Design		(2) Twin Family Design with Assortative Mating	
	Coeff.	s.e.	Coeff.	s.e.
Additive Genetic	110.81	37.81	102.46	31.7
Shared Environment				
MZ twins (males)	643.42	(46.61)	651.77	(42.47)
MZ twins (females)	570.27	(46.08)	578.62	(41.76)
DZ twins (males)	533.75	(30.23)	525.63	(30.61)
DZ twins (females)	428.68	(29.20)	420.56	(29.61)
DZ twins (male-female)	394.36	(24.82)	386.24	(25.18)
Spouses (males)	537.49	(32.83)	534.86	(31.73)
Spouses (females)	455.64	(30.95)	453.01	(29.77)
Father-Son	255.93	(22.16)	247.81	(22.77)
Mother-Daughter	262.26	(21.22)	254.14	(21.85)
Father-Daughter/Mother-Son	227.98	(20.31)	219.86	(20.98)
MZ Twin Uncle-Nephew	264.37	(30.92)	256.50	(31.57)
MZ Twin Aunt-Niece	217.09	(27.45)	209.21	(27.98)
MZ Twin Uncle (Aunt)-Niece (Nephew)	218.73	(21.52)	210.86	(22.26)
DZ Twin Uncle-Nephew	174.09	(14.93)	165.89	(15.82)
DZ Twin Aunt-Niece	177.43	(13.48)	169.23	(14.34)
DZ Twin Uncle (Aunt)-Niece (Nephew)	166.29	(10.45)	158.08	(11.59)
Male Cousins (MZ)	189.25	(42.71)	177.56	(43.54)
Female Cousins (MZ)	187.90	(38.96)	176.21	(39.73)
Male-Female Cousins (MZ)	174.37	(31.72)	162.69	(32.76)
Male Cousins (DZ)	129.45	(23.36)	118.88	(24.11)
Female Cousins (DZ)	83.49	(20.63)	72.92	(21.47)
Male-Female Cousins (DZ)	104.08	(15.53)	93.51	(16.66)
Twins-spouses	503.86	(8.69)	491.56	(9.51)
Brothers-in-law (MZ-S)	354.59	(38.86)	353.12	(38.94)
Sisters-in-law (MZ-B)	376.07	(34.62)	374.60	(34.68)
Brother-Sister in-law (MZ-BS)	405.88	(19.12)	393.59	(19.43)
Brothers-in-law (DZ-S)	323.92	(16.73)	318.54	(16.85)
Sisters-in-law (DZ-B)	329.93	(15.25)	324.54	(15.37)
Brother-Sister in-law (DZ-BS)	337.73	(11.50)	332.34	(11.66)
Idiosyncratic Environment				
Male parents	628.15	(21.19)	639.13	(18.61)
Female parents	573.31	(20.11)	584.29	(17.42)
Idiosyncratic plus Shared Environment				
Male children	726.42	39.48	734.77	33.68
Female children	658.78	39.17	667.13	33.3

Note: The table reports parameter estimates and standard errors of the variance components of education from the twin family design in Column (1) and the twin family design, calibrating assortative mating based on genetic factors to a value equal to 0.12 in Column (2). MZ: monozygotic twins; DZ: dizygotic twins; MZ-S=MZ twins are sisters; MZ-B=MZ twins are brothers; MZ-BS=sister (brother) in-law is wife (husband) of MZ brother (sister); DZ-S=DZ twins are sisters or brother-in-law is husband of DZ sister; DZ-B=DZ twins are brothers or sister-in-law is wife of DZ brother; DZ-BS=sister (brother) in-law is wife (husband) of DZ brother (sister).

Table 5. Education Inequality Decompositions Based on the Twin Family Design.

	(1) Twin Family Design		(2) Twin Family Design with assortative mating	
	Coeff.	s.e.	Coeff.	s.e.
Cross-Sectional Inequality				
Proportion Additive Genetic	0.09	0.03	0.08	0.03
Proportion Shared Environment	0.42	0.02	0.42	0.02
Proportion Idiosyncratic Environment	0.49	0.01	0.50	0.01
Intergenerational Elasticity (IGE)				
Proportion Shared Environment	0.24	0.01	0.23	0.01
	0.81	0.06	0.82	0.06

Note: The table reports the decomposition of education cross-sectional inequality and intergenerational elasticities from the twin family design in Column (1) and the twin family design calibrating assortative mating based on genetic factors to a value equal to 0.16 in Column (2).

Table 6. Education Correlations Decompositions Based on the Twin Family Design.

	(1)	(2)	(3)	(4)	(5)
	Empirical correlation	Predicted correlation	Predicted Shared Environmental Proportion		
			$\delta=0$	$\delta=0.12$	$\delta=0.3$
MZ twins (males)	0.58	0.58	0.85	0.86	0.87
MZ twins (females)	0.59	0.59	0.84	0.85	0.86
DZ twins (males)	0.45	0.46	0.91	0.89	0.87
DZ twins (females)	0.42	0.42	0.89	0.87	0.84
DZ twins (male-female)	0.36	0.37	0.88	0.86	0.83
Father-Son	0.30	0.30	0.82	0.80	0.75
Mother-Daughter	0.34	0.34	0.83	0.80	0.76
Father-Daughter/Mother-Son	0.29	0.29	0.80	0.78	0.73
MZ Twin Uncle-Nephew	0.28	0.28	0.91	0.88	0.83
MZ Twin Aunt-Niece	0.26	0.26	0.89	0.85	0.80
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.25	0.25	0.89	0.86	0.80
DZ Twin Uncle-Nephew	0.18	0.18	0.93	0.88	0.79
DZ Twin Aunt-Niece	0.20	0.20	0.93	0.88	0.79
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.18	0.18	0.92	0.88	0.78
Male Cousins (MZ)	0.26	0.26	0.87	0.82	0.72
Female Cousins (MZ)	0.28	0.28	0.87	0.82	0.72
Male-Female Cousins (MZ)	0.25	0.25	0.86	0.81	0.70
Male Cousins (DZ)	0.17	0.17	0.90	0.83	0.66
Female Cousins (DZ)	0.13	0.13	0.86	0.75	0.50
Male-Female Cousins (DZ)	0.15	0.15	0.88	0.79	0.59
Twins-Spouses	0.41	0.41		0.98	0.94
Brothers-in-law (MZ-S)	0.28	0.27		1.00	0.98
Sisters-in-law (MZ-B)	0.33	0.33		1.00	0.98
Brother-Sister in-law (MZ-BS)	0.33	0.33		1.00	0.98
Brothers-in-law (DZ-S)	0.25	0.25		0.98	0.95
Sisters-in-law (DZ-B)	0.29	0.29		0.98	0.95
Brother-Sister in-law (DZ-BS)	0.28	0.28		0.98	0.95

Note: The table reports empirical correlations and predictions from the twin family design, with environmental proportions computed under different assumptions for assortative mating based on genetic factors. All predictions are statistically significant with a p-value=0.000. Empty cells correspond to cases where the environmental proportion is one by construction. MZ: monozygotic twins; DZ: dizygotic twins; MZ-S=MZ twins are sisters; MZ-B=MZ twins are brothers; MZ-BS=sister (brother) in-law is wife (husband) of MZ brother (sister); DZ-S=DZ twins are sisters or brother-in-law is husband of DZ sister; DZ-B=DZ twins are brothers or sister-in-law is wife of DZ brother; DZ-BS=sister (brother) in-law is wife (husband) of DZ brother (sister).

Table 7. Inequality Decompositions for Other Outcomes.

	(1)		(2)		(3)	
	Earnings		Income		Assets	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Panel A. Classic Twin Design</i>						
Cross-Sectional Inequality						
Proportion Additive Genetic	0.60	0.05	0.56	0.04	0.42	0.04
Proportion Shared Environment	-0.07	0.03	-0.04	0.03	0.05	0.03
Proportion Idiosyncratic Environment	0.47	0.02	0.48	0.02	0.53	0.02
<i>Panel B. Twin Family Design</i>						
Cross-Sectional Inequality						
Proportion Additive Genetic	0.16	0.04	0.14	0.04	0.14	0.04
Proportion Shared Environment	0.26	0.03	0.26	0.03	0.26	0.03
Proportion Idiosyncratic Environment	0.58	0.02	0.60	0.02	0.60	0.02
Intergenerational Elasticity (IGE)	0.18	0.01	0.18	0.01	0.25	0.01
Proportion Shared Environment	0.55	0.13	0.61	0.13	0.71	0.08

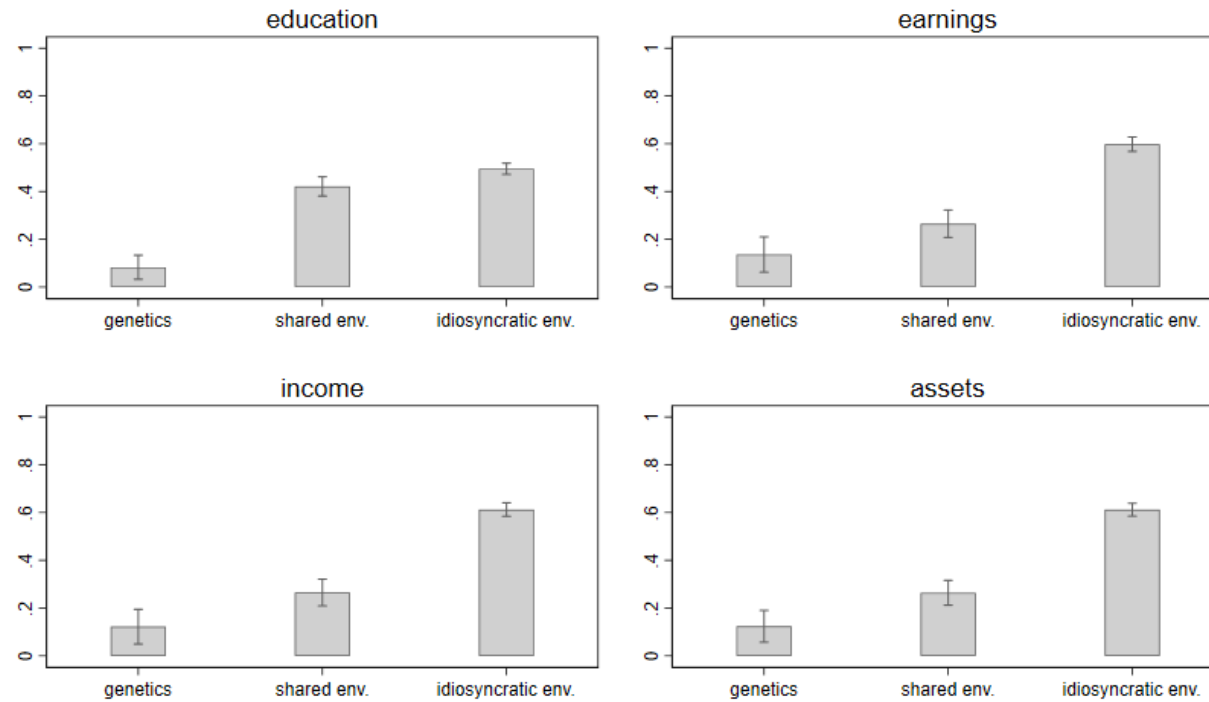
Note: The table reports the decompositions of cross-sectional inequality for earnings (Column 1), disposable income (Column 2), and assets (Column 3) from the CTD in Panel A, and the decompositions for these outcomes of cross-sectional inequality and intergenerational elasticities from the twin family design in Panel B.

Table 8. Correlation Decompositions for Other Outcomes Based on the Twin Family Design.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Earnings			Income			Assets		
	Empirical Corr.	Predicted Corr.	Predicted Shared Envir. Proportion	Empirical Corr.	Predicted Corr.	Predicted Shared Envir. Proportion	Empirical Corr.	Predicted Corr.	Predicted Shared Envir. Proportion
MZ twins (males)	0.50	0.50	0.72	0.53	0.53	0.76	0.55	0.55	0.78
MZ twins (females)	0.61	0.61	0.69	0.50	0.50	0.69	0.38	0.38	0.56
DZ twins (males)	0.27	0.27	0.75	0.28	0.28	0.77	0.34	0.34	0.82
DZ twins (females)	0.29	0.29	0.68	0.28	0.28	0.72	0.23	0.23	0.64
DZ twins (male-female)	0.15	0.15	0.46	0.14	0.14	0.50	0.20	0.20	0.64
Father-Son	0.21	0.21	0.68	0.21	0.21	0.72	0.31	0.31	0.81
Mother-Daughter	0.17	0.17	0.47	0.17	0.17	0.54	0.18	0.18	0.59
Father-Daughter/Mother-Son	0.15	0.15	0.49	0.15	0.15	0.55	0.21	0.21	0.69
MZ Twin Uncle-Nephew	0.16	0.16	0.78	0.16	0.16	0.81	0.24	0.24	0.87
MZ Twin Aunt-Niece	0.14	0.14	0.67	0.15	0.15	0.75	0.11	0.11	0.65
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.13	0.13	0.70	0.13	0.13	0.74	0.15	0.15	0.77
DZ Twin Uncle-Nephew	0.08	0.08	0.78	0.08	0.08	0.81	0.14	0.14	0.89
DZ Twin Aunt-Niece	0.09	0.09	0.76	0.09	0.09	0.77	0.10	0.10	0.81
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.08	0.08	0.75	0.08	0.08	0.78	0.09	0.09	0.81
Male Cousins (MZ)	0.22	0.22	0.85	0.25	0.25	0.89	0.14	0.14	0.80
Female Cousins (MZ)	0.18	0.18	0.75	0.19	0.19	0.80	0.20	0.20	0.83
Male-Female Cousins (MZ)	0.08	0.08 [^]	0.50 ^{^^}	0.13	0.13	0.75	0.18	0.18	0.83
Male Cousins (DZ)	0.09	0.09	0.82	0.11	0.11	0.87	0.08	0.08	0.82
Female Cousins (DZ)	0.09	0.09	0.76	0.08	0.08	0.77	0.11	0.11	0.85
Male-Female Cousins (DZ)	0.08	0.08	0.77	0.09	0.09	0.82	0.07	0.07	0.78
Spouses	0.18	0.18		0.06	0.06		0.22	0.22	
Brothers-in-law (MZ-S)	0.21	0.21		0.23	0.23		0.23	0.23	
Sisters-in-law (MZ-B)	0.20	0.20		0.20	0.20		0.23	0.23	
Brother-Sister in-law (MZ- BS)	0.16	0.16		0.15	0.15		0.18	0.18	
Brothers-in-law (DZ-S)	0.15	0.15		0.14	0.14		0.19	0.19	
Sisters-in-law (DZ-B)	0.15	0.16		0.14	0.14		0.15	0.15	
Brother-Sister in-law (DZ-BS)	0.11	0.11		0.10	0.10		0.12	0.12	

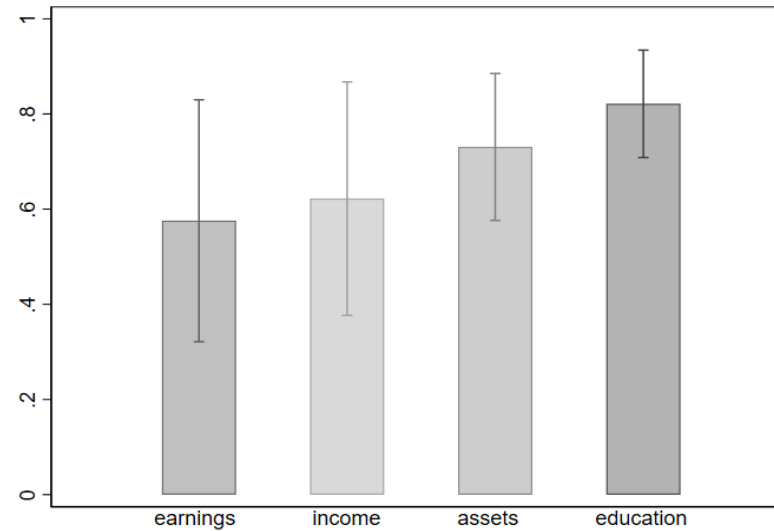
Note: The table reports empirical correlations and predictions from the twin family design as well as the predicted proportion of shared environment. All predictions are statistically significant with a p-value<0.005, except [^] (p-value=0.075) and ^{^^} (p-value=0.111). Empty cells correspond to cases where the environmental proportion is one by construction. MZ: monozygotic twins; DZ: dizygotic twins; MZ-S=MZ twins are sisters; MZ-B=MZ twins are brothers; MZ-BS=sister (brother) in-law is wife (husband) of MZ brother (sister); DZ-S=DZ twins are sisters or brother-in-law is husband of DZ sister; DZ-B=DZ twins are brothers or sister-in-law is wife of DZ brother; DZ-BS=sister (brother) in-law is wife (husband) of DZ brother (sister).

Figure 1.
Decomposition of Cross-Sectional Inequality in the Twin Family Design.



Note: The figure reports the decomposition of inequality of education, earnings, income, and assets into genetic factors, common environment, and idiosyncratic (non-shared) environment. The estimates are based on the twin family twin, where we include twin spouses and their children and allow for differential shared environments and assortative mating based on genetic factors.

Figure 2.
Decomposition of Intergenerational Correlations in the Twin Family Design.



Note: The figure reports the proportion of shared environments when decomposing intergenerational correlations of education, earnings, income, and assets. The estimates are based on the twin family design, where we include twin spouses and their children and allow for differential shared environments and assortative mating based on genetic factors.

APPENDIX

Appendix Table A1. List of Covariance Restrictions.

Equation	Model	Familial Relationship	Extended layout
3	Canonical	MZ twins	$cov(y_i, y_{i'})_{11}^{T_1 T_2} = \sigma_a^2 + \sigma_c^2$
3	Canonical	DZ twins	$cov(y_i, y_{i'})_{00}^{T_1 T_2} = 0.5\sigma_a^2 + \sigma_c^2$
5	Assortative mating	MZ twins	$cov(y_i, y_{i'})_{11}^{T_1 T_2} = \sigma_a^2 + \sigma_c^2$
5	Assortative mating	DZ twins	$cov(y_i, y_{i'})_{00}^{T_1 T_2} = 0.5(1 + \delta)\sigma_a^2 + \sigma_c^2$
6	Assortative mating	MZ twin co-twin's spouse	$cov(y_i, y_{i'})_{11}^{T_j S_k} = \delta\sigma_a^2 + \gamma_{SL}$
6	Assortative mating	DZ twin co-twin's spouse	$cov(y_i, y_{i'})_{00}^{T_j S_k} = \delta(0.5(1 + \delta))\sigma_a^2 + \gamma_{SL}$
7	Assortative mating	MZ twin's spouses	$cov(y_i, y_{i'})_{11}^{S_1 S_2} = \delta^2\sigma_a^2 + \gamma_{SL}$
7	Assortative mating	DZ twin's spouses	$cov(y_i, y_{i'})_{00}^{S_1 S_2} = \delta^2(0.5(1 + \delta))\sigma_a^2 + \gamma_{SL}$
10	Dominance	MZ twins	$cov(y_i, y_{i'})_{11}^{T_1 T_2} = \sigma_a^2 + \sigma_c^2 + \sigma_d^2$
10	Dominance	DZ twins	$cov(y_i, y_{i'})_{00}^{T_1 T_2} = 0.5\sigma_a^2 + \sigma_c^2 + 0.25\sigma_d^2$
11	Dominance	MZ twin co-twin's child	$cov(y_i, y_{i'})_{11}^{T_j C_k} = 0.5\sigma_a^2 + \gamma_A$
11	Dominance	DZ twin co-twin's child	$cov(y_i, y_{i'})_{00}^{T_j C_k} = 0.25\sigma_a^2 + \gamma_A$
12	Dominance	Cousins born to MZ twins	$cov(y_i, y_{i'})_{11}^{C_1 C_2} = 0.25\sigma_a^2 + \gamma_C$
12	Dominance	Cousins born to DZ twins	$cov(y_i, y_{i'})_{00}^{C_1 C_2} = 0.125\sigma_a^2 + \gamma_C$
13	Differential environment	Male MZ twins	$cov(y_i, y_{i'})_{21}^{T_1 T_2} = \sigma_a^2 + \sigma_{c21}^2$
13	Differential environment	Female MZ twins	$cov(y_i, y_{i'})_{01}^{T_1 T_2} = \sigma_a^2 + \sigma_{c01}^2$
13	Differential environment	Male DZ twins	$cov(y_i, y_{i'})_{20}^{T_1 T_2} = 0.5\sigma_a^2 + \sigma_{c20}^2$
13	Differential environment	Female DZ twins	$cov(y_i, y_{i'})_{00}^{T_1 T_2} = 0.5\sigma_a^2 + \sigma_{c00}^2$
13	Differential environment	Mixed DZ twins	$cov(y_i, y_{i'})_{10}^{T_1 T_2} = 0.5\sigma_a^2 + \sigma_{c10}^2$
16	Differential environment	Father son	$cov(y_i, y_{i'})_{22}^{P_j C_j} = 0.5\sigma_a^2 + \gamma_{I2}$
16	Differential environment	Mother daughter	$cov(y_i, y_{i'})_{00}^{P_j C_j} = 0.5\sigma_a^2 + \gamma_{I0}$
16	Differential environment	Father daughter/Mother son	$cov(y_i, y_{i'})_{11}^{P_j C_j} = 0.5\sigma_a^2 + \gamma_{I1}$

(continues on next page)

Appendix Table A1 (Continued). List of Covariance Restrictions.

Equation	Model	Familial Relationship	Extended layout
14	Differential environment	MZ male twin co-twin's son	$cov(y_i, y_{i'})_{21}^{T_j C_k} = 0.5\sigma_a^2 + \gamma_{A21}$
14	Differential environment	MZ female twin co-twin's daughter	$cov(y_i, y_{i'})_{01}^{T_j C_k} = 0.5\sigma_a^2 + \gamma_{A01}$
14	Differential environment	MZ male twin co-twin's daughter/ MZ female twin-co-twin's son	$cov(y_i, y_{i'})_{11}^{T_j C_k} = 0.5\sigma_a^2 + \gamma_{A11}$
14	Differential environment	DZ male twin co-twin's son	$cov(y_i, y_{i'})_{20}^{T_j C_k} = 0.25\sigma_a^2 + \gamma_{A20}$
14	Differential environment	DZ female twin co-twin's daughter	$cov(y_i, y_{i'})_{00}^{T_j C_k} = 0.25\sigma_a^2 + \gamma_{A00}$
14	Differential environment	DZ male twin co-twin's daughter/ MZ female twin-co-twin's son	$cov(y_i, y_{i'})_{10}^{T_j C_k} = 0.25\sigma_a^2 + \gamma_{A10}$
15	Differential environment	Male spouse of MZ twin co-twin's son	$cov(y_i, y_{i'})_{21}^{S_j C_k} = \gamma_{A21}$
15	Differential environment	Female spouse of MZ twin co-twin's daughter	$cov(y_i, y_{i'})_{01}^{S_j C_k} = \gamma_{A01}$
15	Differential environment	Male spouse of MZ twin co-twin's daughter/ Female spouse of MZ twin co-twin's son	$cov(y_i, y_{i'})_{11}^{S_j C_k} = \gamma_{A11}$
15	Differential environment	Male spouse of DZ twin co-twin's son	$cov(y_i, y_{i'})_{20}^{S_j C_k} = \gamma_{A20}$
15	Differential environment	Female spouse of DZ twin co-twin's daughter	$cov(y_i, y_{i'})_{00}^{S_j C_k} = \gamma_{A00}$
15	Differential environment	Male spouse of DZ twin co-twin's daughter/ Female spouse of MZ twin-co-twin's son	$cov(y_i, y_{i'})_{10}^{S_j C_k} = \gamma_{A10}$
17	Differential environment	Male cousins born to MZ twins	$cov(y_i, y_{i'})_{21}^{C_1 C_2} = 0.25\sigma_a^2 + \gamma_{C21}$
17	Differential environment	Female cousins born to MZ twins	$cov(y_i, y_{i'})_{01}^{C_1 C_2} = 0.25\sigma_a^2 + \gamma_{C01}$
17	Differential environment	Mixed cousins born to MZ twins	$cov(y_i, y_{i'})_{11}^{C_1 C_2} = 0.25\sigma_a^2 + \gamma_{C11}$
17	Differential environment	Male cousins born to DZ twins	$cov(y_i, y_{i'})_{20}^{C_1 C_2} = 0.125\sigma_a^2 + \gamma_{C20}$
17	Differential environment	Female cousins born to DZ twins	$cov(y_i, y_{i'})_{00}^{C_1 C_2} = 0.125\sigma_a^2 + \gamma_{C00}$
17	Differential environment	Mixed cousins born to DZ twins	$cov(y_i, y_{i'})_{10}^{C_1 C_2} = 0.125\sigma_a^2 + \gamma_{C01}$
19	Differential environment	Male MZ twin co-twin's spouse/Male spouse of MZ twin- co-twin's spouse	$cov(y_i, y_{i'})_{21}^{T_j S_k} = cov(y_i, y_{i'})_{21}^{S_j S_k} = \gamma_{SL21}$
19	Differential environment	Female MZ twin co-twin's spouse/Female spouse of MZ twin co-twin's spouse	$cov(y_i, y_{i'})_{01}^{T_j S_k} = cov(y_i, y_{i'})_{01}^{S_j S_k} = \gamma_{SL01}$
19	Differential environment	Male DZ twin male co-twin's spouse/Male spouse of DZ twin male co-twin's spouse	$cov(y_i, y_{i'})_{20}^{T_j S_k} = cov(y_i, y_{i'})_{20}^{S_j S_k} = \gamma_{SL20}$
19	Differential environment	Female DZ twin co-twin's spouse/Female spouse of DZ twin co-twin's spouse	$cov(y_i, y_{i'})_{00}^{T_j S_k} = cov(y_i, y_{i'})_{00}^{S_j S_k} = \gamma_{SL00}$
19	Differential environment	Male DZ twin female co-twin's spouse/Male spouse of DZ twin female co-twin's spouse	$cov(y_i, y_{i'})_{10}^{T_j S_k} = cov(y_i, y_{i'})_{10}^{S_j S_k} = \gamma_{SL10}$

Appendix Table A2. Decompositions of Cross-Sectional Inequality and IGE by Gender Based on the Twin Family Design.

	(1)		(2)		(3)		(4)		
	Education		Earnings		Income		Assets		
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	
Men	Cross-Sectional Inequality								
Proportion Additive Genetic Factors	0.08	(0.03)	0.14	(0.04)	0.13	(0.04)	0.12	(0.03)	
Proportion Shared Environment	0.44	(0.02)	0.26	(0.03)	0.28	(0.03)	0.33	(0.03)	
Proportion Idiosyncratic Environment	0.48	(0.02)	0.60	(0.02)	0.59	(0.02)	0.55	(0.02)	
Women									
Proportion Additive Genetic Factors	0.10	(0.03)	0.19	(0.05)	0.15	(0.05)	0.17	(0.05)	
Proportion Shared Environment	0.41	(0.02)	0.27	(0.04)	0.25	(0.04)	0.17	(0.04)	
Proportion Idiosyncratic Environment	0.49	(0.02)	0.54	(0.03)	0.60	(0.02)	0.66	(0.02)	
	IGE								
IGE Men	0.24	(0.01)	0.22	(0.01)	0.22	(0.01)	0.32	(0.01)	
Proportion Shared Environment	0.82	(0.06)	0.68	(0.09)	0.72	(0.09)	0.81	(0.06)	
IGE Women	0.28	(0.01)	0.18	(0.01)	0.17	(0.01)	0.21	(0.01)	
Proportion Shared Environment	0.83	(0.06)	0.47	(0.15)	0.54	(0.15)	0.59	(0.12)	

Note: The table reports the decompositions of cross-sectional inequality and intergenerational elasticities from the twin family design for all outcomes by gender.

Appendix Table A3. Predicted Education Correlations.

	(1)		(2)
	Predicted correlations		Empirical correlations
	Sample A		Sample B
MZ twins (males)	0.60	[0.53,0.67]	0.55
MZ twins (females)	0.60	[0.52,0.68]	0.57
DZ twins (males)	0.47	[0.42,0.52]	0.43
DZ twins (females)	0.41	[0.35,0.46]	0.44
DZ twins (male-female)	0.35	[0.31,0.38]	0.38
Spouses	0.41	[0.40,0.43]	0.41
Father-Son	0.28	[0.26,0.31]	0.31
Mother-Daughter	0.34	[0.31,0.37]	0.33
Father-Daughter/Mother-Son	0.29	[0.27,0.31]	0.28
MZ Twin Uncle-Nephew	0.27	[0.19,0.35]	0.28
MZ Twin Aunt-Niece	0.29	[0.21,0.37]	0.23
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.24	[0.18,0.29]	0.25
DZ Twin Uncle-Nephew	0.16	[0.13,0.20]	0.20
DZ Twin Aunt-Niece	0.20	[0.16,0.24]	0.20
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.18	[0.15,0.20]	0.19
Male Cousins (MZ)	0.23	[0.07,0.38]	0.28
Female Cousins (MZ)	0.27	[0.14,0.41]	0.29
Male-Female Cousins (MZ)	0.21	[0.10,0.31]	0.30
Male Cousins (DZ)	0.15	[0.08,0.22]	0.19
Female Cousins (DZ)	0.14	[0.07,0.22]	0.11
Male-Female Cousins (DZ)	0.11	[0.06,0.16]	0.17
Brothers-in-law (MZ-S)	0.27	[0.19,0.35]	0.28
Sisters-in-law (MZ-B)	0.34	[0.26,0.43]	0.31
Brother-Sister in-law (MZ-BS)	0.36	[0.32,0.40]	0.30
Brothers-in-law (DZ-S)	0.23	[0.19,0.26]	0.27
Sisters-in-law (DZ-B)	0.29	[0.25,0.32]	0.29
Brother-Sister in-law (DZ-BS)	0.28	[0.25,0.30]	0.28

Note: The table reports predicted education correlations (with 95% confidence intervals in brackets) generated after estimating the heterogeneous environment model on a random half sample (column 1) and empirical correlations computed from the remaining cases (column 2). MZ: monozygotic twins; DZ: dizygotic twins; MZ-S=MZ twins are sisters; MZ-B=MZ twins are brothers; MZ-BS=sister (brother) in-law is wife (husband) of MZ brother (sister); DZ-S=DZ twins are sisters or brother-in-law is husband of DZ sister; DZ-B=DZ twins are brothers or sister-in-law is wife of DZ brother; DZ-BS=sister (brother) in-law is wife (husband) of DZ brother (sister).

Appendix Table A4. Additional Sample Descriptive Statistics.

	MZ Families				DZ Families			
	Male		Female		Male		Female	
	Labor Earnings							
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. dev.
Twin	65,154	35,608	48,023	23,323	63,270	35,384	44,942	21,632
Spouse	70,418	46,017	44,898	22,083	67,798	36,924	44,423	21,297
Child	74,349	48,042	56,146	26,744	73,406	43,938	55,298	27,004
	Disposable income							
	Average	Stand. Dev.	Average	Stand. Dev.	Average	Stand. Dev.	Average	Stand. dev.
Twin	35,706	49,219	32,531	14,801	34,869	30,215	30,444	16,505
Spouse	39,451	89,177	30,985	25,213	36,321	42,848	30,363	14,789
Child	46,348	34,734	42,800	56,332	46,497	42,016	41,600	22,566
	Assets							
	Average	Stand. Dev.	Average	Stand. Dev.	Average	Stand. Dev.	Average	Stand. dev.
Twin	145,534	618,016	90,127	190,701	150,102	743,172	74,405	140,169
Spouse	202,950	889,146	81,553	275,400	179,209	735,632	72,916	140,059
Child	164,237	322,665	130,017	215,280	180,483	507,411	125,202	307,889

Note: The table reports means and standard deviations of labor earnings, disposable income, and assets (in 2023 USD) by gender and by zygosity.

