I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16520

On the Origins of Socio-Economic Inequalities: Evidence from Twin Families

Paul Bingley
Lorenzo Cappellari
Konstantinos Tatsiramos

OCTOBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16520

On the Origins of Socio-Economic Inequalities: Evidence from Twin Families

Paul Bingley
VIVE Copenhagen

Lorenzo Cappellari
Università Cattolica Milano, LISER and IZA

Konstantinos Tatsiramos
University of Luxembourg, LISER and IZA

OCTOBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

On the Origins of Socio-Economic Inequalities: Evidence from Twin Families*

We propose a twin family model linking twins with their spouses and children to quantify the relative importance of genetic and environmental factors in explaining the variance of socio-economic outcomes. Using data from the Danish Twins Registry and population registers, we test and relax the assumptions of the standard behavioral genetics model most frequently applied in economics using twins or adoptees. Exploiting an education reform differentially affecting parents, we find no evidence of gene-environment interactions. While we find some assortative mating based on genetic factors, differentially shared environments are key: they explain half of the variance in years of schooling, whereas genetic factors explain only nine percent. We find similar percentages for earnings, income, and wealth. Decomposing intergenerational elasticities reveals that shared environments explain 50% for earnings, 60% for income, 70% for wealth, and 80% for schooling. Family environments are more important than previously understood.

JEL Classification: D31, D63, E21, E24, I24

Keywords: nature, nurture, family background, genes, environment, inequality

Corresponding author:
Konstantinos Tatsiramos
University of Luxembourg, LISER
Maison des Sciences Humaines
11, Porte des Sciences
L-4366 Esch-sur-Alzette / Belval
Luxembourg
E-mail: konstantinos.tatsiramos@uni.lu

* This research was financed by the Luxembourg National Research Fund (FNR), grant reference [C21/SC/15771764]. We thank Kaare Christensen, Petter Lundborg, Erik Plug, and Jan Stuhler for their discussions on this project. All errors and omissions are the responsibility of the authors.

1. Introduction

While it is widely acknowledged that both inherited endowments and family background play pivotal roles in shaping the determinants of individual success, such as human capital, abilities, and skills (Becker and Tomes, 1979), understanding the inequalities in opportunities for success in life requires tracing individuals back to their social origins. Extensive literature in economics and the broader social sciences has explored this issue, employing two approaches to study the origins of observed disparities in socioeconomic outcomes. These approaches have yielded contrasting conclusions regarding the relative importance of endowments at birth versus the influence of family environments (see Sacerdote, 2011, and Mogstad and Torsvik, 2023, for reviews).

The first approach, grounded in the standard model of behavioral genetics known as the ACE model, decomposes the variance in outcomes into three components: (1) the proportion explained by *additive* genetic variation (heritability), (2) the proportion explained by *common* (or shared) *environment* and (3) the proportion explained by *idiosyncratic* (non-shared) *environment.* Empirical applications of the ACE model using twins (Taubman, 1976; Behrman and Taubman, 1989; Björklund, Jäntti and Solon, 2005; Cesarini et al. 2009a,b; Barnea et al. 2010; Cronqvist and Siegel, 2015) and adoptees (Sacerdote, 2007; Fagereng, Mogstad and Rønning, 2021), consistently find that genetic factors explain a much higher proportion of the variance in socioeconomic outcomes compared to shared environmental factors. In contrast, the second approach, which employs regression-based methods on adoptees' data and focuses on the intergenerational elasticity (IGE) of economic outcomes, concludes that shared family environments play a significant role in shaping children's outcomes (Sacerdote, 2002; Bjorklund, Lindahl and Plug, 2006; Lundborg, Plug and Rasmussen, 2021).[1]

In this paper, we introduce a different approach to studying the origins of economic inequality—one based on a *twin family design* that links twins with their spouses and children.[2] We can reconstruct the twins' families by combining data from the Danish Twin Registry (which identifies twins and their zygosity) with population registers.[3] Our twin family design offers two main advantages over the ACE model applied to twin pairs, the classic twin design (CTD). First, we can decompose cross-sectional inequality without imposing the restrictive assumptions of the CTD, which might lead to biased estimates of the importance of genetic and common environmental effects (Taubman, 1976; Mogstad and Torsvik, 2023). Second, we can estimate the influence of genetic factors on inequality not only within the twins' generation but also on intergenerational transmission, thereby directly answering the long-standing question about the extent to which parent-child transmission depends on factors determined at birth. As the CTD only considers within-generation associations, it does not allow for directly decomposing intergenerational effects. Indeed, whereas only adoption studies have previously decomposed intergenerational elasticities into genetic and environmental shares, our approach identifies this decomposition without requiring children's separation from their birth parents, which may raise concerns about external validity.[4]

[1] In a novel extension of the adoption study design, Lundborg, Plug and Rasmussen (2021) study children conceived through fertility treatment using donor sperm or eggs, which they refer to as embryo adoptions.
[2] 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.
[3] Nance and Corey (1976) introduced the twin family model. Behrman and Rosenzweig (2002) used schooling differences between twin parents and between their children to estimate the intergenerational transmission of schooling.
[4] For applications that extend the CTD outside economics, see, e.g., Eley, et.al. (2015). Collado, Ortuño-Ortín and Stuhler (2023) combine information on relatives both within and between generations to quantify the extent of intergenerational transmission of latent genetic factors.

The CTD requires the imposition of five assumptions to identify the three variance components (genetics, shared environment, and non-shared environment). The first assumption 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 bias down the heritability estimates from the CTD, inflating the estimate of the common environmental factor. The second assumption asserts that genetic variation has an additive effect, with no genetic dominance or gene-gene interactions. Genetic dominance would bias up the heritability estimates from the CTD because it would make dizygotic (DZ or fraternal) twins less similar for outcomes influenced by dominance. The third assumption states that genes and environment do not interact, implying that the effect of genetic factors on outcomes does not vary with the contextual environment. The direction of the CTD model bias depends on the type of gene-environment interaction.

The fourth assumption posits that pairs of differing genetic similarities share the environment to the same extent, meaning that the influence of shared environmental factors is the same for monozygotic (MZ or identical) and DZ twins. Any deviation from this assumption, where MZ and DZ twins do not share environmental influences equally, would bias CTD estimates by overstating the heritability of traits, attributing differences in outcomes between MZ and DZ twins to genetic differences. The fifth and final assumption asserts that there is no correlation between genetic factors and the environment, i.e., genes are not sorted according to environment. Given that studies of socioeconomic outcomes do not find evidence of gene-environment correlations (Björklund, Jäntti, and Solon, 2005; Fagereng, Mogstad, and Rønning, 2021; Collado, Ortuño-Ortín, and Stuhler, 2023), we will maintain this fifth assumption while relaxing and testing the first four.

We begin the analysis by applying the CTD to educational correlations between MZ and DZ twins. We find that genetic factors account for about one-third of the variation in years of schooling. At the same time, shared environments explain only about a quarter of it, with the remaining inequality attributable to idiosyncratic determinants of education. Subsequently, we relax each of the first four assumptions of the CTD. First, we complement the CTD with information on the twins' spouses, allowing us to test for assortative mating on genetic factors. Second, we complement the CTD with information on the children of twins, allowing us to test for genetic dominance. Third, we exploit an educational reform that affected the parent generation, allowing us to test for gene-environment interactions. Our findings reveal no compelling evidence of genetic dominance or gene-environment interactions. However, we do find a modest positive correlation between the genotypes related to education for the twins and their spouses (0.16), resulting in a larger percentage of inequality accounted for by genetic variation compared to the standard model (41% vs. 34%) and a smaller percentage accounted for by shared environments (18% vs. 24%).

Finally, we complement the CTD with information on the twins' spouses *and* children. This extension allows us to test the equally shared environments assumption that the data rejects. We find that MZ twins share their environments far more than DZ twins. Imposing the equally shared environments assumption tends to load any differential correlation in years of schooling between MZ and DZ twins onto genetic differences. Allowing for differential environments between twin pairs with different zygosity and relative-specific environmental sharing for other relatives within the family, we find that shared family background explains half of the variance in years of schooling. In contrast, only nine percent of the variance is due to heritability, while assortative mating on genetic factors has no additional impact on these inequality decompositions.

We reach similar conclusions about the relative importance of genetic variation and shared environments for earnings, income, and wealth as we do for education. Genetic factors explain about 15% of the total variance for these outcomes, while shared family environments account for about 35%. The remaining variance is attributed to individual-specific factors not shared within the family.

Relying on the parameter estimates of the twin-family design, we also decompose the intergenerational elasticities of earnings, income, wealth, and education. We find that the shared environment explains a large percentage of the correlation of outcomes between parents and children, equal to about 50% for earnings, 60% for income, 70% for wealth, and 80% for years of education.

Our findings contribute to the literature that applies the ACE model to socioeconomic outcomes by highlighting the critical role played by the equally shared environments assumption. Furthermore, our research contributes to the literature that uses adoptees to decompose the intergenerational transmission of socioeconomic outcomes. We find that family environments are more important than previously understood for education, earnings, income and wealth. The limited role of genetic variation in accounting for education inequalities corroborates recent evidence from a calibrated extended family model based on horizontally distant relatives (Collado, Ortuño-Ortín, and Stuhler, 2023). We propose an estimable model that requires only a few relatives—up to six family members—of known genetic relatedness, and we derive the additional moment restrictions for testing each of the assumptions of the CTD, which we apply to a broad set of socioeconomic outcomes.

The rest of the paper is structured as follows. Section 2 describes the data. Section 3 presents the standard ACE model, the empirical correlations of education, and its variance decomposition into genetic variation, shared environment, and non-shared environment. Section 4 extends the standard ACE model to allow for assortative mating on genetic factors, genetic dominance, gene-environment interactions, and differentially shared environments between twins with different genetic similarities. Section 5 reports results for earnings, income, and wealth, as well as for decomposing the intergenerational elasticities. 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, everyone resident in Denmark has been registered with a unique personal identification number that is now 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 all 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. As record incompleteness also implies that, for parents born before 1935, the first registered child may not be the firstborn, we exclude families in which twins or spouses are born before 1935.[5] To include a family, we require that twins, twin co-parents, and children meet these birth cohort criteria.[6]

To observe individuals of labor market age, we exclude children born after 1984 and parents born after 1966. Table 1 presents the sample that we will 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. 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).

3. The Classic Twin Design

[5] Individuals need to be resident in Denmark at some point in 1980-2019.
[6] We also drop the very small number of families in which both parents are twins.

The classic twin design (CTD)—also known as the ACE model—is the method that is most often used for decomposing the variation of traits into additive genetic and shared environmental factors. Researchers have applied it to psychological characteristics such as IQ and personality, psychiatric disorders such as schizophrenia and bipolar disorder, and socioeconomic outcomes such as education, earnings, and wealth.[7]

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

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

where $f(i)$ is the family of person i, while $h(i)$ denotes the person's role within the family; 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 a shared environmental factor (shared by the members 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.

We must impose several assumptions to identify the three components of the CTD. The first assumption is that there is *no assortative mating based on genetic factors*, i.e., mating is random, implying that fraternal twins share, on average, half of their segregating genetic endowments. Positive assortative mating on genetic factors suggests that DZ twins share more than half of their

[7] Polderman et. al. (2015) perform a meta-analysis of twin studies on a very wide variety of traits. Applications of the ACE model are not limited to twin data; see Sacerdote, 2007, and Fagereng, Mogstad and Rønning, 2021, for economic applications using adoptees.

segregating genetic endowment. The second assumption is *no genetic dominance*, that gene variants do not interact but add up to affect outcomes.[8]

The third assumption is that the three components of Equation (1) enter the model linearly, excluding the possibility of *gene-environment interactions* due to gene expression variation with environmental exposure. The fourth assumption is that MZ and DZ twins share environments to the same extent, irrespective of their genetic similarity. This "*equally shared environments*" assumption can be very restrictive because it excludes, for example, the possibility that parents (or schoolmates or neighbors) treat MZ twins more similarly than they treat DZ twins. The fifth assumption is that there is no selection of genetic similarity over environments. 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 such a 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.[9] Drawing on this evidence, we develop our model maintaining no gene-environment correlation and using our data's features to relax the first four assumptions.

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

$$var(y_i) = \sigma_a^2 + \sigma_c^2 + \sigma_e^2 \ . \qquad (2)$$

[8] While assortative mating affects genetic relatedness of any relationship type, not only of DZ twins, dominance impacts primarily DZ twins and would spread to other relationship types only if combined with assortative mating (see Keller et al., 2009).

[9] A negligible correlation between genetic factors and the environment within educational inequality is also reported by Collado, Ortuño-Ortín and Stuhler (2023).

MZ twins share all of their genes at conception, while, in the absence of assortative mating and genetic dominance, DZ twins share only half of their segregating genes on average. Let $z \in \{0, 1\}$ be an indicator for MZ twin pairs. We summarize the covariance of outcomes between twins as follows

$$cov(y_i, y_{i'})_z^{T_1T_2} = 0.5^{(1-z)}\sigma_a^2 + \sigma_c^2 \, . \qquad (3)$$

Equation (3) encompasses two moment restrictions depending on the value of z, one for MZ twins ($z = 1$) and one for DZ twins ($z = 0$). Therefore, equations (2) and (3) identify the three variance components of the CTD model exactly, enabling us to decompose the variance of outcome y_i into genetic, shared environmental, and idiosyncratic factors and to quantify the degree of heritability, measuring heritability by the proportion of the genetic component within the total outcome variance, 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). Specifically, we use an Equally Weighted Minimum Distance estimator in which we weight the minimization by the identity matrix and adjust standard errors using the empirical matrix of fourth moments.

In Table 2, we report the empirical correlations of years of education for twins by the zygosity of the pair. The first two rows report education correlations for MZ twins distinguished by the gender of the twins; for both women and men, the correlation is large, approaching 0.6. Subsequent rows refer to DZ twin pairs. The correlations are lower than for MZ twins, showing a distinctive contrast between same-gender and mixed-gender pairs.

3.2 Results

Column (1) of Table 3 reports the parameter estimates for education, which we obtain by matching the moment restrictions of the CTD of Section 3.1 to all empirical moments of Table 2. In Panel A, the variance components suggest that genetic factors are more important than the shared environment in accounting for education inequality. Panel B decomposes cross-sectional inequality, showing that heritability accounts for 34% of the total variance. In contrast, the shared environment accounts for 24%, and we attribute the remaining variance to the idiosyncratic environmental factor that twins do not share. These findings are in line with results in the literature that, on average, the heritability of education is around 40% (Mogstad and Torsvik, 2023).

Using the parameter estimates from the CTD model, we can predict the correlations between MZ twins (0.58) and DZ twins (0.41). These predicted correlations are very close to the empirical correlations in Table 2. We can also predict the percentages of the variance attributable to shared environments for MZ twins (41%) and DZ twins (59%). These percentages are statistically significantly different.

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). With only three moment restrictions, if we relax any of the CTD assumptions, we can no longer identify the model. By including information on other family members, 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.

Before discussing how we can relax, in turn, the assumptions of the CTD, we present in Appendix Table A1 the empirical correlations of education for the various pairs of relatives that we match within the twin families. We distinguish relatives by gender and by zygosity for relationships connected through the twin pair (uncles/aunts-nephews/nieces, cousins, siblings-in-law). Correlations in education between spouses and between parents and their children are large (0.41 and 0.30, respectively). Between mothers and daughters, correlations are 0.34. Correlations in education between nuclear families are generally larger when they are connected by MZ twins. Relatives of the same gender have larger correlations than mixed-gender pairs.

4.1 Assortative Mating

Assortative mating is the tendency for individuals to mate with partners who are similar to themselves. Researchers have found evidence of assortative mating on education (Eike, Mogstad, and Zafar, 2019; Collado, Ortuño-Ortín and Stuhler, 2023), earnings (Gonalons-Pons, 2017), income (Greenwood et al., 2014), and wealth (Fagerang, Guiso, and Pistaferri, 2022). In contrast to these studies of assortative mating on socio-economic outcomes, we are interested in the related but distinct phenomenon of assortative mating based on genetic factors.

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 genotypes among spouses (see Bowles and Gintis, 2002).[10] With assortative mating, the spousal covariance becomes

$$cov(y_i, y_{i'})^{T_j S_j} = \delta\sigma_a^2 + \sigma_{cS}^2 , \qquad j = 1{,}2 \qquad (4)$$

[10] 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 σ_{cS}^2 parameterizes environmental sharing among spouses. Correlation of genotypes between spouses does not increase genetic similarity among MZ twins. However, assortative mating increases the genetic similarity of DZ twins by a factor of 0.5δ. The twin covariance becomes

$$cov(y_i, y_{i'})_z^{T_1T_2} = (0.5(1+\delta))^{(1-z)}\sigma_a^2 + \sigma_{cT}^2 \,, \qquad (5)$$

where σ_{cT}^2 denotes the shared environmental variance component for the twins. Equation (5) extends equation (3) of the CTD to incorporate assortative mating based on genetic factors. We cannot identify the model resulting from Equations (2), (4), and (5) because, relative to the CTD, this model adds one equation and two parameters. To identify this extended model, we use information on the covariances between siblings-in-law.[11] The following two equations give the moment restrictions for siblings-in-law

$$cov(y_i, y_{i'})_z^{T_jS_k} = \delta(0.5(1+\delta))^{(1-z)}\sigma_a^2 + \sigma_{cSL}^2 \quad j,k = 1,2 \qquad j \neq k \qquad (6)$$

$$cov(y_i, y_{i'})_z^{S_1S_2} = \delta^2(0.5(1+\delta))^{(1-z)}\sigma_a^2 + \sigma_{cSL}^2, \qquad (7)$$

where σ_{cSL}^2 parameterizes environmental sharing between siblings-in-law. Equation (6) defines the covariance between twins and the spouses of the co-twins. Equation (7) defines the covariance between the spouses of the twins. We identify this extended CTD by introducing information on siblings-in-law, thereby adding four equations but only one environmental parameter.

4.1.2 Results

Extending the CTD by including spouses and siblings-in-law, we allow for assortative mating based on genetic factors and re-estimate the variance components and the decomposition of education inequality. Column (2) of Table 3 shows the correlation in genotypes related to education

[11] We use the term "siblings-in-law" to refer to spouse's co-twin, or co-twin's spouse, regardless of marital status.

between spouses is 0.16. This estimate implies that genetic similarity accounts for about 16% of the spousal correlation in education, computed as $\delta\sigma_a^2/(\delta\sigma_a^2 + \sigma_{cS}^2)$. We reach this conclusion while maintaining assumptions 2-5 of the CTD, a conclusion to which we return in Section 4.4.

Comparing Panels A and B in Figure 1, we see that allowing for assortative mating leads to an insignificantly higher percentage of heritability compared with the CTD (41% and 34%, respectively), and a lower percentage of educational inequality attributed to shared environments compared with the CTD (18% and 24%, respectively).[12]

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. A specific type of gene-gene interaction, dominance is a relationship that is 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 the same parents. In contrast, other relatives exhibit much less dominance because they inherit genetic combinations in a less stable way.

4.2.1 Model and Identification

Now we consider extending the CTD to allow for genetic dominance. To focus on the implications of the dominance assumption for the CTD, we once again assume no assortative mating based on

[12] For the decomposition of cross-sectional education inequality, we use equation (2), replacing the common environmental component σ_c^2 with the environmental parameter of twins σ_{cT}^2.

genetic factors. 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

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

Cesarini et al. (2009b) call this extension to the ACE model the ACDE. Assuming genetic dominance, we introduce a new fourth element to the CTD variance decomposition

$$var(y_i) = \sigma_a^2 + \sigma_{cT}^2 + \sigma_d^2 + \sigma_e^2 \,. \qquad (9)$$

Because MZ twins share all of their genes, the dominance factor enters their covariance function directly. However, because DZ twins share only half of their segregating genes, the dominance factor enters their covariance function with a weight of only 0.25. This weight is the probability that parents transmit the same (dominant) genes because each parent has a 50% chance of passing on a dominant allele to their offspring, and there is a 50% chance that both parents pass on the same dominant allele. Therefore, the moment restriction for twins becomes

$$cov(y_i, y_{i'})_z^{T_1T_2} = 0.5^{(1-z)}\sigma_a^2 + \sigma_{cT}^2 + 0.25^{(1-z)}\sigma_d^2 \,. \qquad (10)$$

Extending the CTD to allow for dominance means that the model is no longer identified, because now we have four parameters and only three moment restrictions. To identify this extended CTD, we need to supplement the model with moment restrictions featuring two distinct characteristics: First, they should depend on shared environmental and additive genetic factors, not dominant ones. Second, given the common environment, they should allow additive genetic factors to vary according to the type of relationship. 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 exhibit these characteristics, because the covariances do not depend on the zygosity of the twin parent. Parent-child covariances add only one moment restriction and one common environment parameter.

Moment restrictions that satisfy the criteria required for identifying the CTD with dominance are those between uncles or aunts and nephews or nieces (avuncular relationships), or between cousins. Because we assume random mating, these moment restrictions do not depend on genetic dominance. Covariances between these pairs depend only on shared environmental and additive genetic factors, in which the strength of the genetic connection depends on the twins' zygosity. These pairs include one or two family members from the offspring generation, and an individual in this extended model can now be either a twin or a twin's child, i.e., $h(i) \in \{T_1, T_2, C_1, C_2\}$.

Moment restrictions for avuncular relations are

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

where σ_{cA}^2 parameterizes environmental sharing in avuncular relationships. For cousins, the moment restrictions are:

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

where the parameter σ_{cC}^2 captures environmental sharing among cousins. We can now identify this extended CTD model because each of equations (11) and (12) adds two moment restrictions—one for each zygosity—and one environmental parameter.

4.2.2 Results

To extend the CTD model to encompass dominance, we apply the factorization of equation (8). Table 3, Panel A, Column 3 shows that the variance component of the dominance factor σ_d^2 is relatively small and statistically insignificant. While the estimate of the additive genetic component σ_a^2 is slightly smaller than the CTD estimate, this difference is insignificant. Furthermore, Table 3, Panel B shows that when we extend the CTD model, we find very little change in the percentage of educational variance attributable to genetic factors (32% with dominance versus 34% in the CTD). These three insignificant differences support the CTD assumption of no genetic dominance.

4.3 Gene-Environment Interactions

To allow for gene-environment interactions—for the environment to mediate genetic expression—we exploit a change to the social environment for the twin pairs in our sample. Specifically, we focus on a 1937 Danish school reform requiring urban municipalities to provide eighth- and ninth-grade teaching.[13] Municipalities expanded teaching capacity only gradually, so that several urban municipalities had yet to offer eighth and ninth grades in 1958, when a second reform required all municipalities to offer these grades.

To measure distance to the nearest school that teaches eighth and ninth grades, we use the twins' year of birth and parish of birth registration.[14] We split twin families into two groups according to whether the distance to school was above or below the median at age 14, the normal age for enrolling in eighth grade.[15] By interacting this binary partition of the sample with the variance components of the model, we can test for gene-environment interactions, i.e., for whether estimated coefficients are equal across educational regimes.

---

[13] Municipalities are the local government units responsible for primary and lower secondary schools. Schooling laws distinguish between municipalities that have a designated market town (an urban municipality) and those that do not. 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. Thus the 1958 birth cohort had the first children legally required to receive instruction in eighth and ninth grades.

[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.

[15] Parish of residence is not registered until 1970. For calculating distance to school, we assume that individuals reside in their registered parish of birth at age 14.

4.3.1 Results

In Column (3) of Table 3, we report estimates from the extended CTD model, allowing for gene-environment interactions with the educational reform. The information necessary for calculating twins' distance to school is available for 75,651 of 80,204 individuals in our main sample.[16] Using this information, we assign 8,453 (34,526) families (individuals) to the group most exposed to the environmental shift induced by the educational reform—those living closest to school—corresponding to 52% (45%) of the observations. If gene-environment interactions matter, we should observe the genetic component varying with differences in exposure to the schooling environment induced by the reform.

In Panel A of Table 3, we report estimates of the genetic, shared environmental, and idiosyncratic variance components for the treated—those exposed to the educational reform—and the non-treated. For the treated, who live closest to school, the environment (both shared and idiosyncratic) accounts for a smaller percentage of the variance in education than it does for the non-treated. However, the percentage of the variance accounted for by genetic factors is not statistically different between the two groups [$\chi^2(1)$=0.33, p-value=0.566]. Therefore, because the genetic variance component does not differ between institutionally different environments, we find, at least in our setting, no evidence of gene-environment interactions.

4.4 Differential Shared Environments

4.4.1 Model and Identification

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

To relax the equally shared environments assumption of the CTD, we introduce a twin family design, that uses information on twins, their spouses, and children, i.e., $h(i) \in \{T_1, T_2, S_1, S_2, C_1, C_2\}$. As in the models with assortative mating or dominance, the twin family model uses relationships between nuclear families, for example, between cousins or siblings in law (horizontal relationships, within generations) or between twins (or twin spouses) and their "niblings" (nephews/nieces— vertical relationships, between generations).

Two aspects of the relationships between members of the twin family can cause them to share environments to a different degree. The first aspect is the zygosity of the twin pair. Zygosity may affect not only the extent of environmental sharing for the twins (thus violating the equally shared environments assumption of the CTD) but also the shared environmental component for other relatives. For example, cousins with MZ twin parents may see one another more often than cousins born to DZ twins. The second aspect is the gender composition of the relationship. For example, relatives of the same gender may share the environment to a greater extent than mixed-gender pairs, and environmental sharing may differ by gender for same-gender pairs.

For incorporating gender-related environmental variation into the model, let $g \in \{0, 1, 2\}$ indicate the number of males in a pair of relatives. With differential environments, the covariance of education for twins becomes

$$cov(y_i, y_{i'})_{gz}^{T_1 T_2} = 0.5^{(1-z)}\sigma_a^2 + \sigma_{cTgz}^2 \ . \qquad (13)$$

Equation (13) encompasses five moment restrictions: two for MZ twin pairs (two males or two females) and three for DZ twin pairs (same-gender or mixed-gender pairs).

The model defined by equations (2) and (13) 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 σ_{cTgz}^2) but only six moment restrictions. To identify this extended CTD, we use additional moment restrictions from between generations, i.e., covariances between parents and children and between twins (or co-parents) and their niblings.

For parents and children, the covariance is

$$cov(y_i, y_{i'})_g^{P_j C_j} = 0.5\sigma_a^2 + \sigma_{cIg}^2 \,, \qquad P = S,T \qquad j = 1,2 \qquad (14)$$

which contains three moment restrictions, each depending on the gender composition of the parent-child pair (both men, both women or mixed–gender), and adds three environmental sharing factors to the parameter set. Given that parents always pass on average 50% of their segregating genes to their children, the moment restrictions and the environmental sharing factors do not vary with the twins' zygosity.

For twins and their niblings, the covariance is

$$cov(y_i, y_{i'})_{gz}^{T_j C_k} = 0.5^{(1-z)} 0.5\sigma_a^2 + \sigma_{cAgz}^2 \,, \qquad j,k = 1,2 \qquad j \neq k \qquad (15)$$

where σ_{cAgz}^2 parameterizes differential environmental sharing in avuncular relationships. The children of MZ twins share genetics with their parents to the same extent ($0.5\sigma_a^2$), as they do with their consanguineous uncles and aunts (who are MZ co-twins). In contrast, for the children of DZ twins, the genetic sharing factor with their consanguineous uncles and aunts (DZ co-twins) is halved ($0.25\sigma_a^2$). Equation (15) adds six moment restrictions (each combination of zygosity and gender composition of the avuncular pair) and six parameters (the environmental variance components σ_{cAgz}^2) to the model. If, instead of considering twins, we consider the twins' spouses, who are unrelated by blood to the co-twins' children (and have no genetic similarity, absent assortative mating on genetic factors), covariances between twins' spouses and their niblings depend only on environmental sharing

$$cov(y_i, y_{i'})_{gz}^{S_j C_k} = \sigma_{cAgz}^2 \,. \qquad j,k = 1,2 \qquad j \neq k \qquad (16)$$

We can now identify this extended CTD model because it has fewer parameters than moment restrictions: 16 parameters (idiosyncratic variance, genetic variance, and 14 shared environmental components) and 21 moment restrictions (equation 2 and equations 13-16). Equation (16) provides sufficient additional information for identifying the model because it adds six moment restrictions without adding new parameters. To avoid adding parameters, we assume that the degree of environmental sharing between twins and their niblings (Equation 15) is the same as that between twins' spouses and niblings (Equation 16). Our twin family design assumes that avuncular relationships share their environments equally. We argue that this assumption is much weaker than the CTD assumption that MZ and DZ twins share their environments equally. To identify the model parameters, we take the difference between Equations 15 and 16, resulting in a function that depends only on the genetic variance component.

Because in the rest of the model we add at least as many equations as parameters, these additions do not affect identification. The degree of environmental sharing between cousins depends on their gender composition and the zygosity of their twin parents, resulting in the following moment restrictions

$$cov(y_i, y_{i'})_{gz}^{C_1C_2} = 0.5^{(1-z)}0.25\sigma_a^2 + \sigma_{cCgz}^2 , \qquad (17)$$

where the six parameters σ_{cCgz}^2 (one for each combination of z and g) capture environmental sharing in avuncular relationships within generations.

The twin family design also encompasses relationships between spouses and between siblings-in-law. When we assume no assortative mating on genetic factors, these other relationships do not share genetic factors. For twins and their spouses, the education covariance is as follows

$$cov(y_i, y_{i'})^{T_jS_j} = \sigma_{cS}^2 . \qquad j = 1,2 \qquad (18)$$

This relationship does not depend on the zygosity of the twins. For siblings-in-law, i.e., relationships between a twin and the co-twin's spouse or between a twin's spouse and the co-twin's spouse, the outcome covariance is given by

$$\left.\begin{matrix} cov(y_i, y_{i'})_{gz}^{T_j S_k} \\ \\ cov(y_i, y_{i'})_{gz}^{S_1 S_2} \end{matrix}\right\} = \sigma_{cSLgz}^2 . \quad j,k = 1,2 \quad j \neq k \tag{19}$$

4.4.2 Results

Table 4 presents parameter estimates from the twin family model, in which we extend the CTD by relaxing the equal environments assumption. In Column (1), we assume no assortative mating based on genetic factors. The estimates show substantial heterogeneity in the degree of environmental sharing between different relatives. We reject the equally shared environments assumption for MZ and DZ twins (with a p-value equal to 0.011 for men and 0.006 for women). In Column (2), we report the variance components when we calibrate the model with 0.16 correlation between spouses in genetic factors related to education—the value of assortative mating based on genetic factors reported in Table 3.[17] The variance components in both columns of Table 4 are very similar.

Table 5 presents the decomposition of education inequality from the twin family model.[18] When we allow shared environments to vary by zygosity, we find that heritability accounts for 9% of the total variance and that shared environments account for 50%. We attribute the remaining

[17] We calibrate assortative mating in the model with fully flexible environmental variation, because in that context the siblings-in-law moment restrictions (which identify assortative mating in the baseline design) carry four distinct environmental parameters (one for each zygosity-gender composition combination), implying that we cannot identify the assortative mating parameter.

[18] For the decomposition of cross-sectional education inequality, we use equation (2) substituting the common environmental component σ_c^2 with the average environmental parameter of MZ twins across genders. In Appendix Table A.4 we also report gender-specific inequality decompositions.

variance to non-shared environmental factors. Assortative mating does not affect the inequality decomposition. In Figure 2, we compare the decomposition of education inequality between the CTD and the twin family design, demonstrating that the CTD overestimates the role of genetic factors. Shared environments become the most important component of the total variance, increasing from 24% in the CTD to 50% in the twin family design. In contrast, the percentage of the variance that genetic factors account for declines from 34% to 9%. These findings demonstrate that the equal environments assumption is too restrictive and biases the variance decomposition. We find that shared environments are much more important in explaining educational inequality than previously thought.

A unique feature of our twin family design is that we not only can decompose cross-sectional education inequality—a within-generation measure—but also can decompose the intergenerational elasticity (IGE) of education into environmental and genetic factors. The IGE is given by

$$\beta = \frac{0.5\sigma_a^2 + \overline{\sigma_{cI}^2}}{\sigma_a^2 + \overline{\sigma_{cT1}^2} + \sigma_e^2} = \underbrace{\frac{0.5\sigma_a^2}{\sigma_a^2 + \overline{\sigma_{cT1}^2} + \sigma_e^2}}_{genetic\ component} + \underbrace{\frac{\overline{\sigma_{cI}^2}}{\sigma_a^2 + \overline{\sigma_{cT1}^2} + \sigma_e^2}}_{environmental\ component}, \tag{20}$$

where overbars denote averages of parameters over the index g. Table 5 shows that the IGE of education is equal to 0.24, with environmental factors accounting for 81%.[19]

We present an overview of our results for all relatives in Table 6. The first column shows that predicted educational correlations match the empirical moments of Table 2 very closely. Given the large number of relative-specific environmental sharing parameters, we should expect to find

[19] 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 sibling correlation as follows

$$corr(y_i, y_{i'})_{gz}^{T_1T_2} = \frac{0.5^{(1-z)}\sigma_a^2 + \sigma_{cTgz}^2}{\sigma_a^2 + \sigma_{cTg1}^2 + \sigma_e^2} = \underbrace{\frac{0.5^{(1-z)}\sigma_a^2}{\sigma_a^2 + \sigma_{cTg1}^2 + \sigma_e^2}}_{genetic\ component} + \underbrace{\frac{\sigma_{cTgz}^2}{\sigma_a^2 + \sigma_{cTg1}^2 + \sigma_e^2}}_{environmental\ component}$$

a close match.[20] Columns 2-4 of Table 6 report percentages of the total variance attributed to shared environments, where the columns calibrate the model with different degrees of assortative mating. In the absence of assortative mating (δ=0), shared environments account for at least 80% of the total variance.[21]

Calibrating assortative mating based on genetic factors with the 0.16 correlation in education-related genotypes (reported in Table 3), we find that the percentage of the variance attributed to environmental factors changes only slightly. This calibration exercise confirms the conclusion we reached when decomposing cross-sectional inequality from the twin family model, with and without assortative mating. In the model with flexible shared environments, genetic variation accounts for only 4% of spousal correlation in years of education (compared with 16% in the CTD model with assortative mating). We find that decomposing the correlation for cousins yields the largest effect from introducing assortative mating. Column 4 of Table 6 shows that to obtain a substantive impact on correlations for other relatives, we need to calibrate assortative mating based on genetic factors with the value 0.30. Because this correlation implies that spouses have more genetic factors in common than cousins who have MZ parents, we consider 0.30 to be an upper bound.

[20] 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 A2, show that the 95% confidence interval of the predicted correlations from the first subsample always contains the empirical correlations estimated from the second subsample.

[21] The environmental percentages are stable if we compare DZ twins with other relationships that depend on zygosity, i.e., cousins or uncle/aunt-nephew/niece. Larger environmental percentages for DZ twins could be a symptom of (omitted) dominance.

5. Additional Results – Earnings, Income, and Wealth

This section presents estimation results for the additional outcomes of earnings, income, and wealth.[22] We observe annual pre-tax labor earnings from income tax returns. Employers report earnings for each employee to the tax authorities, who in turn 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 employments during a year for the period 1980-2018.

Because the twin population is small, we do not have the number of observations necessary for estimating a model of permanent earnings with multi-person life-cycle dynamics (see, e.g., Bingley, Cappellari, and Tatsiramos, 2021). Therefore, rather than estimating permanent earnings, we follow the simpler empirical strategy of measuring permanent earnings as the average earnings percentile between ages 30 and 50.[23] An intuitive way of separating permanent earnings from transitory shocks, averaging individual earnings over this age range helps 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). Using earnings percentiles instead of earnings levels, we also avoid the biases associated with modeling fluctuations of earnings levels in the cross-section and over the life cycle.

For permanent income and wealth, we also use average percentiles between ages 30 and 50.[24] We calculate disposable income by subtracting taxes from personal gross income and adding

[22] Appendix Table A2 presents descriptive statistics for earnings, income, and wealth.
[23] To calculate permanent earnings, we require at least five annual earnings observations. Percentiles are first estimated by year, and then recentered by gender and person type (whether a twin, child or spouse).
[24] Requiring at least five data points in the 30-50 age range and excluding cases with missing information marginally reduces sample size for income and wealth (each having about 79,200 individuals) with respect to the sample used for analyzing educational inequality. For earnings, we reduce the sample somewhat more (about 72,500 individuals for analysis), with the loss concentrated among females in the parental generation.

transfers. We measure wealth by using taxable assets. The wealth register contains information on end-of-year financial assets, non-financial assets, and liabilities.[25] 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.[26] As registration of pension wealth from defined contribution plans began in 2012, for consistency of our measures, we exclude these plans from our wealth measures.

To decompose cross-sectional inequality for earnings, income, and assets, we use estimates from the CTD model in Panel A of Table 7. Genetic variation accounts for 60% of cross-sectional inequality for earnings, 54% for disposable income, and 42% for assets. Similar to the results from the CTD for education (Panel B of Table 3), genetic factors explain a much larger percentage of cross-sectional inequality than the percentage explained by shared environments.

In Panel B of Table 7, we report the cross-sectional decomposition from the twin family design. Relaxing the equally shared environments assumption reduces the percentage of the variance explained by genetic factors from 60% to 17% for earnings, from 54% to 13% for disposable income, and from 42% to 14% for assets. As with education, we find that the CTD assumption of equally shared environments substantially overstates the role that genetic factors play in explaining inequality of earnings, income, and assets. Our decompositions show that shared environments are more important than genetic factors for explaining inequality of socioeconomic outcomes, with shared environments accounting for almost 40% of the total variance for earnings and income, and 35% for assets.

[25] 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).
[26] See Boserup et al. (2018) for the importance of wealth and assets over the life cycle in Denmark.

We summarize this evidence in Figure 3, which shows that heritability accounts for 13% to 17% of cross-sectional inequality across economic outcomes. Genetic factors matter more for these other outcomes than for education. We find the heritability of earnings to be almost double that of education. Conversely, shared environmental factors explain slightly more than a third of earnings dispersion, while accounting for half of education inequality.

In Panel B of Table 7, we decompose the IGE of earnings, income, and assets. Using the estimates from the twin family model, we calculate the IGE of earnings and income to be 0.18, and the IGE of assets to be 0.25. We find that shared environments account for 52% of the intergenerational correlation of earnings, 62% of income, and 71% of assets. Figure 4 summarizes this evidence, showing that shared environmental factors explain more than half of the intergenerational correlation of socioeconomic outcomes, with shared environment being most important for education (81% from Table 5).[27] Table 8 illustrates the evidence for different relatives, reporting both predicted correlations and predicted environmental shares. Correlations tend to be lower for earnings, income, and assets than for education, especially for non-twin pairs.[28]

In general, genetic factors tend to matter less for education than for these other outcomes, despite our measuring economic outcomes later in life.[29] One explanation for this difference is that education takes place in an environment that is egalitarian, at least in principle, with schools and teachers moderating the effects of genetic factors in shaping inequalities.[30] This moderating effect

[27] In Appendix Table A.4 we report decompositions of cross-sectional inequality and IGEs by gender.
[28] Correlations between spouses are lower for disposable income than for earnings and for assets. This difference may reflect disposable income including transfers that are effectively gender-specific, e.g., parental leave, taken mostly by women in the parent cohorts in our study.
[29] In a longitudinal study of IQ correlations between adopted offspring and their rearing parents, Willoughby, et. al. (2021) find the shared environment explains a falling percentage of the variance with age.
[30] In Denmark, education is tuition-free at all levels. From age 18, students are entitled to six years of grants for further education.

from schools does not operate to the same extent for outcomes determined in the market. Among the economic outcomes, genetic factors are most important for earnings. Compensatory effects of the welfare state may explain why pre-birth factors are less important for income and wealth.

6. Conclusions

We quantify the relative importance of genetic and shared environmental factors for explaining the variance of socioeconomic outcomes. Using data from the Danish Twins Registry and population registers, we propose a twin family model linking twins with their children and co-parents. This model allows us to relax and test the assumptions of the classic twin design (CTD), which, if violated, could lead to biased estimates of heritability (the proportion of inequality accounted for by genetic factors). Our twin family model allows us to decompose cross-sectional inequality and intergenerational elasticities.

We reject the equally shared environments assumption—that MZ and DZ twins share environments to the same extent. Maintaining this assumption leads to upward-biased estimates of heritability. Allowing for differentially shared environments within the twin family design, we find that shared environmental factors explain 50% of the variance in years of schooling. In contrast, genetic factors explain only 9%. We find similar percentages for earnings, income, and wealth. Decomposing intergenerational elasticities, shared environmental factors explain 50% for earnings, 60% for income, 70% for wealth, and 80% for years of education.

We derive the additional moment restrictions for testing each of the assumptions of the CTD. Linking twin pairs with their spouses is necessary for allowing for assortative mating. Connecting twin pairs with their children is essential for allowing for genetic dominance. Finally, connecting twins with their spouses and children is necessary for relaxing the equally shared environments assumption. We find some evidence of assortative mating based on genetic factors. When we maintain the assumption of no assortative mating in the CTD, heritability estimates are downward biased. In contrast, in our twin family design, we find that assortative mating based on genetic factors does not alter the main findings. Moreover, we find no evidence for genetic dominance, nor do we find evidence for gene-environment interactions, which we test by using an educational expansion that varies school proximity in the parent generation differentially between twin families.

Our findings from the twin family design suggest that family environment is more important than previously thought in twin studies that decompose cross-sectional inequality and in those of adoptee studies that decompose intergenerational transmission. We corroborate recent evidence for education from a calibrated extended family model based on horizontally distant relatives (Collado, Ortuño-Ortín, and Stuhler, 2023). The twin family design we propose has the advantage of being estimable and of requiring only twins, spouses, and their children, combined with information on genetic relatedness for decomposing cross-sectional inequality. The twin family design also allows us to decompose intergenerational elasticities and answer the long-standing question about the extent to which parent-child transmission depends on factors pre-determined at birth, without having to rely on adoptees.

Our model demonstrates the utility of twins for studying the origins of socioeconomic outcomes. By connecting twins' spouses and children, we avoid the biases of the CTD. We find that relaxing the equally shared environments assumption reveals a greater role for shared environmental factors in explaining socioeconomic inequality. However, these findings come from Denmark, with its relatively generous college student support and high rates of income taxation and redistribution. Future studies applying our model to countries with more inequality may reveal additional CTD biases.

While we find no evidence of gene-environment interactions in education using a reform to partition the variance in years of schooling for our twin families, our test clearly depends on outcome and context. Nonetheless, we cannot rule out gene-environment interactions for other outcomes and in other contexts. Future work could apply our model to groups facing different institutional settings, for example, different wage bargaining regimes affecting earnings variance, or different tax regimes affecting income and wealth. Evidence of such gene-environment interactions would deepen scholarly understanding of the behavioral mechanisms that drive responses to policy reforms.

With the widening availability of multi-generational surveys and administrative datasets in general and genetically informed datasets in particular, our modeling approach can help social scientists understand the origins of inequality without imposing the strong assumptions of the CTD. Extending our model to include other relatives—non-twin siblings, half siblings, step- siblings and grandparents—would provide a more nuanced understanding of the role of shared environments in explaining inequality. These extensions provide fertile ground for future research.

REFERENCES

Baadsgaard, M., and J. Quitzau (2011). Danish Registers on Personal Income and Transfer Payments. *Scandinavian Journal of Public Health,* vol. 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.

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 (1989). "Mostly in the Genes"? Nature-Nurture Decomposition Using Data on Relatives. *Journal of Political Economy*, vol. 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*, vol. 92(1), 323–334.

Bingley, P., and L. Cappellari 2019. Correlation of brothers' earnings and intergenerational transmission. *Review of Economics and Statistics*, vol. 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.

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, pp. 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.

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.

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.

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

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

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.

Eika, L., M. Mogstad and B. Zafar (2019). Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy*, vol. 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.

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.

Gonalons-Pons, P., and C.R. Schwartz (2017). Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage? *Demography*, vol. 54(3), 985-1005.

Greenwood, J., N. Guner, G. Kocharkov and C. Santos 2014. Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, vol. 104(5), 348–53.

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.

Karlson, K., and R. Landersø (2012) The Making and Unmaking of Opportunity: Educational Mobility in 20th Century-Denmark. *IZA Discussion Paper* 14135.

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., 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.

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.

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*, vol. 83(4), 811–826.

Nybom, M. and J. Stuhler (2016). Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation. *Journal of Human Resources*, 51(1): 239-268.

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.

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 Benhahib, 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.

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)

Notes: Standard deviations are in parentheses.

Table 2. Empirical Education Correlations among Twins

MZ twins (males)	0.58
MZ twins (females)	0.59
DZ twins (males)	0.45
DZ twins (females)	0.42
DZ twins (male-female)	0.36

Notes: The table reports empirical education correlations for different types of twin pairs.

Table 3. Parameter Estimates and Decomposition of Education Cross-Sectional Inequality Based on the Classic Twin Design and Extensions.

Panel A. Variance Components

	(1) Classic Twin Design (CTD)		(2) CTD with Assortative Mating		(3) CTD with Genetic Dominance		(4) CTD with Gene-Environment Interactions	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Genetic Factors	419.97	51.09	495.66	59.45	343.06	89.24		
Genetic Factors - Treated							449.44	57.16
Genetic Factors - Not Treated							511.84	92.48
Assortative Mating on G.F.			0.16	0.04				
Genetic Dominance					51.28	58.06		
Shared Environment								
Twins	297.68	33.27	221.54	42.07.	323.32	38.28		
Spouses			426.65	21.92				
Parent-Child					131.56	44.15		
Uncle/Aunt-Nephew/Niece					118.41	28.10		
Cousins					101.21	17.20		
Siblings-in-law			311.55	15.22				
Treated Twins							125.51	37.21
Non-Treated Twins							382.40	58.57
Idiosyncratic Environment	516.98	20.60	504.17	21.24	516.98	20.60		
Idiosync. Env. - Treated							365.29	22.40
Idiosync. Env. - Not Treated							630.55	38.62

(Continues on next page)

Panel B. Cross-Sectional Inequality Decomposition

	(1) Classic Twin Design (CTD)		(2) CTD with assortative mating		(3) CTD with genetic dominance		(4) CTD with Gene-Environment Interactions	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Total Variance	1234.63	8.36	1221.37	7.03	1234.63	8.36		
Proportion Genetic Factors	0.34	0.04	0.41	0.05	0.32	0.04		
Proportion Shared Env.	0.24	0.03	0.18	0.03	0.26	0.03		
Total Variance - Treated							940.24	10.79
Proportion Genetic Factors							0.48	0.06
Proportion Shared Env.							0.13	0.04
Total Variance – Not Treated							1524.80	12.67
Proportion Genetic Factors							0.34	0.06
Proportion Shared Env.							0.25	0.04

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 4. Parameter Estimates for Twin Family Design.

	(1) Twin Family Design		(2) Twin Family Design with Assortative Mating	
	Coeff.	s.e.	Coeff.	s.e.
Genetic Factors	110.81	37.81	116.77	36.41
Shared Environment				
MZ twins (males)	643.42	46.61	637.46	45.66
MZ twins (females)	570.27	46.08	564.30	45.08
DZ twins (males)	533.75	30.23	521.43	31.56
DZ twins (females)	428.68	29.20	416.35	30.6
DZ twins (male-female)	394.36	24.82	382.04	26.27
Spouses	503.86	8.69	485.17	10.49
Father-Son	255.93	22.16	243.61	24.04
Mother-Daughter	262.26	21.22	249.93	23.13
Father-Daughter/Mother-Son	227.98	20.31	215.65	22.35
MZ Twin Uncle-Nephew	264.37	30.92	252.79	32.15
MZ Twin Aunt-Niece	217.09	27.45	205.51	28.45
MZ Twin Uncle (Aunt)-Niece (Nephew)	218.73	21.52	207.15	22.92
DZ Twin Uncle-Nephew	174.09	14.93	165.16	15.94
DZ Twin Aunt-Niece	177.43	13.48	168.50	14.47
DZ Twin Uncle (Aunt)-Niece (Nephew)	166.29	10.45	157.36	11.76
Male Cousins (MZ)	189.25	42.71	177.67	43.57
Female Cousins (MZ)	187.90	38.96	176.32	39.75
Male-Female Cousins (MZ)	174.37	31.72	162.79	32.79
Male Cousins (DZ)	129.45	23.36	120.52	23.97
Female Cousins (DZ)	83.49	20.63	74.56	21.32
Male-Female Cousins (DZ)	104.08	15.53	95.15	16.46
Brothers-in-law (MZ-S)	354.59	38.86	351.60	39.03
Sisters-in-law (MZ-B)	376.07	34.62	373.08	34.74
Brother-Sister in-law (MZ-BS)	405.88	19.12	387.20	19.9
Brothers-in-law (DZ-S)	323.92	16.73	316.12	16.96
Sisters-in-law (DZ-B)	329.93	15.25	322.13	15.48
Brother-Sister in-law (DZ-BS)	337.73	11.50	329.93	11.82
Idiosyncratic				
Male parents	538.63	31.61	538.63	31.61
Female parents	468.80	29.35	468.80	29.35
Male children	83.00	34.16	83.00	34.16
Female children	88.51	32.19	88.51	32.19

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 Genetic Factors	0.09	0.03	0.10	0.03
Proportion Shared Environment	0.50	0.03	0.49	0.03
Intergenerational Elasticity (IGE)	0.24	0.01	0.24	0.01
Proportion Shared Environment	0.81	0.06	0.80	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.

	Predicted correlation	Predicted Environmental Proportion		
		δ=0	δ =0.16	δ =0.3
MZ twins (males)	0.58	0.85	0.85	0.84
MZ twins (females)	0.59	0.84	0.83	0.82
DZ twins (males)	0.46	0.91	0.87	0.84
DZ twins (females)	0.42	0.89	0.84	0.80
DZ twins (male-female)	0.37	0.88	0.83	0.79
Spouses	0.41		0.96	0.93
Father-Son	0.30	0.82	0.76	0.70
Mother-Daughter	0.34	0.83	0.76	0.70
Father-Daughter/Mother-Son	0.29	0.80	0.74	0.67
MZ Twin Uncle-Nephew	0.28	0.91	0.85	0.79
MZ Twin Aunt-Niece	0.26	0.89	0.82	0.75
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.25	0.89	0.82	0.75
DZ Twin Uncle-Nephew	0.18	0.93	0.85	0.75
DZ Twin Aunt-Niece	0.20	0.93	0.85	0.75
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.18	0.92	0.84	0.74
Male Cousins (MZ)	0.26	0.87	0.78	0.67
Female Cousins (MZ)	0.28	0.87	0.78	0.67
Male-Female Cousins (MZ)	0.25	0.86	0.76	0.65
Male Cousins (DZ)	0.17	0.90	0.78	0.63
Female Cousins (DZ)	0.13	0.86	0.69	0.50
Male-Female Cousins (DZ)	0.15	0.88	0.74	0.57
Brothers-in-law (MZ-S)	0.27		0.99	0.97
Sisters-in-law (MZ-B)	0.33		0.99	0.97
Brother-Sister in-law (MZ-BS)	0.33		0.99	0.97
Brothers-in-law (DZ-S)	0.25		0.97	0.93
Sisters-in-law (DZ-B)	0.29		0.97	0.93
Brother-Sister in-law (DZ-BS)	0.28		0.97	0.93

Note: The table reports 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) Earnings		(2) Income		(3) Assets	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Panel A. Classic Twin Design						
Cross-Sectional Inequality						
Proportion Genetic Factors	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
Panel B. Twin Family Design						
Cross-Sectional Inequality						
Proportion Genetic Factors	0.17	0.04	0.13	0.04	0.14	0.04
Proportion Shared Environment	0.37	0.05	0.38	0.05	0.34	0.04
Intergenerational Elasticity (IGE)	0.18	0.01	0.18	0.01	0.25	0.01
Proportion Shared Environment	0.52	0.13	0.62	0.13	0.71	0.08

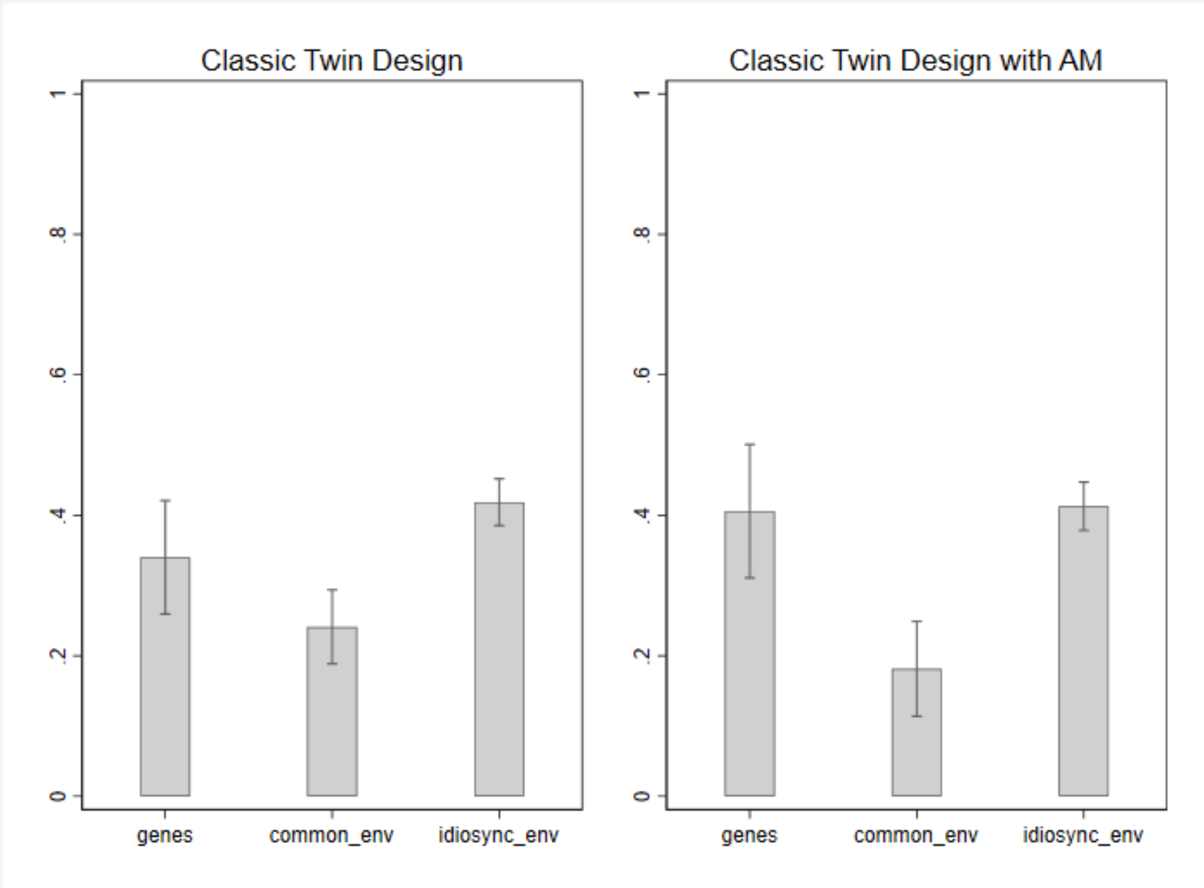
Note: The table reports the decompositions of cross-sectional inequality for earnings (Column 1), disposable income (Column 2), and assets (Column 3) from the CTD in Panel A, and the decompositions for these outcomes of cross-sectional inequality and intergenerational elasticities from the twin family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

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

	(1) Earnings		(2) Income		(3) Assets	
	Predicted	Env. Proportion	Predicted	Env. Proportion	Predicted	Env. Proportion
MZ twins (males)	0.50	0.72	0.52	0.76	0.56	0.78
MZ twins (females)	0.62	0.69	0.51	0.69	0.38	0.56
DZ twins (males)	0.28	0.75	0.28	0.77	0.35	0.82
DZ twins (females)	0.30	0.68	0.28	0.72	0.24	0.64
DZ twins (male-female)	0.15	0.46	0.14	0.50	0.20	0.64
Spouses	0.18		0.06		0.22	
Father-Son	0.21	0.68	0.21	0.72	0.31	0.81
Mother-Daughter	0.17	0.47	0.17	0.54	0.18	0.59
Father-Daughter/Mother-Son	0.15	0.49	0.15	0.55	0.21	0.69
MZ Twin Uncle-Nephew	0.16	0.78	0.16	0.81	0.24	0.87
MZ Twin Aunt-Niece	0.14	0.67	0.15	0.75	0.11	0.65
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.13	0.70	0.13	0.74	0.15	0.77
DZ Twin Uncle-Nephew	0.08	0.78	0.08	0.81	0.14	0.89
DZ Twin Aunt-Niece	0.09	0.76	0.09	0.77	0.10	0.81
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.08	0.75	0.08	0.78	0.09	0.81
Male Cousins (MZ)	0.22	0.85	0.25	0.89	0.14	0.80
Female Cousins (MZ)	0.08^	0.50^^	0.13	0.75	0.18	0.83
Male-Female Cousins (MZ)	0.18	0.75	0.19	0.80	0.20	0.83
Male Cousins (DZ)	0.09	0.82	0.11	0.87	0.08	0.82
Female Cousins (DZ)	0.09	0.76	0.08	0.77	0.11	0.85
Male-Female Cousins (DZ)	0.08	0.77	0.09	0.82	0.07	0.78
Brothers-in-law (MZ-S)	0.20		0.23		0.23	
Sisters-in-law (MZ-B)	0.20		0.20		0.23	
Brother-Sister in-law (MZ-BS)	0.16		0.15		0.18	
Brothers-in-law (DZ-S)	0.15		0.14		0.19	
Sisters-in-law (DZ-B)	0.15		0.14		0.15	
Brother-Sister in-law (DZ-BS)	0.11		0.10		0.12	

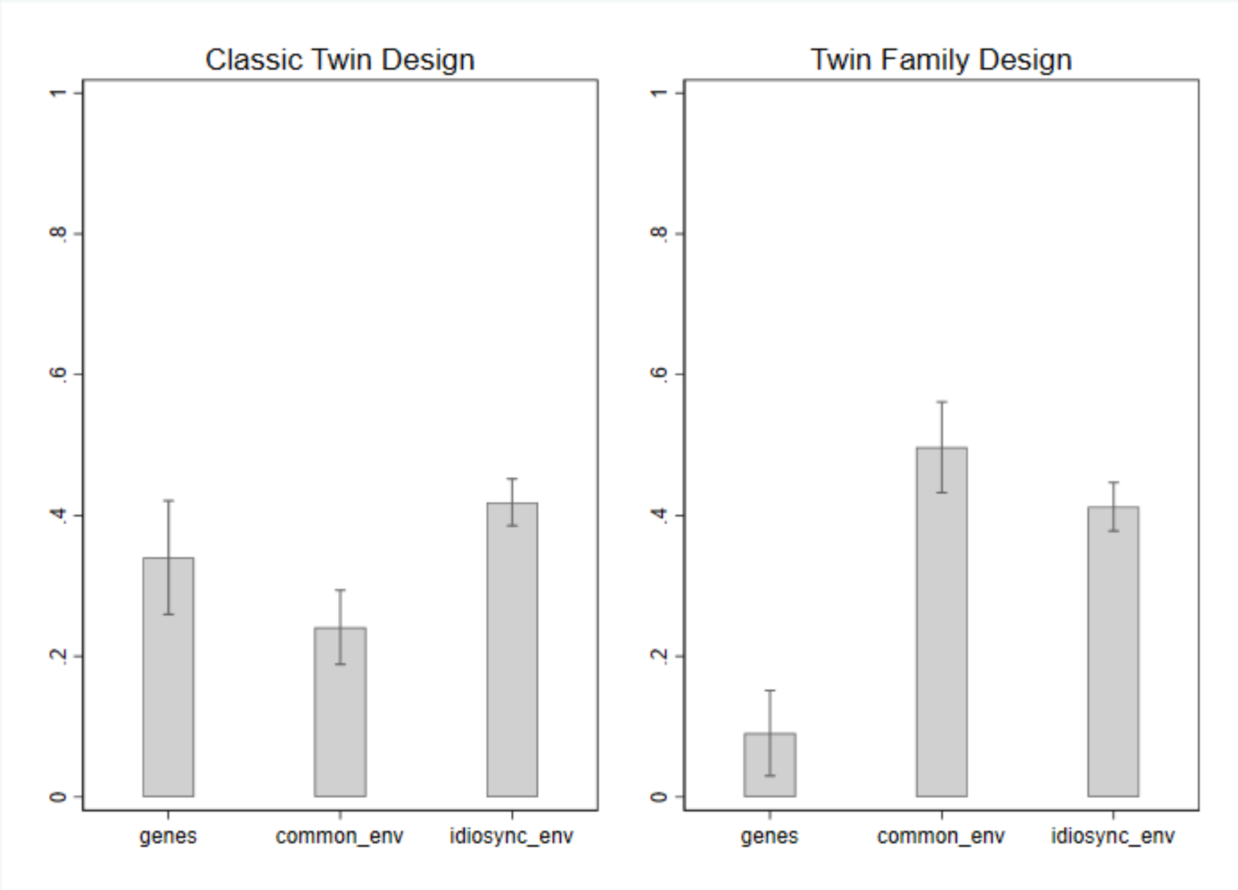
Note: The table reports predictions from the twin family design, with environmental proportions computed under different assumptions for assortative mating based on genetic factors. All predictions are statistically significant with a p-value<0.005, except ^ (p-value=0.075) and ^^ (p-value=0.111). Empty 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 Education Inequality.



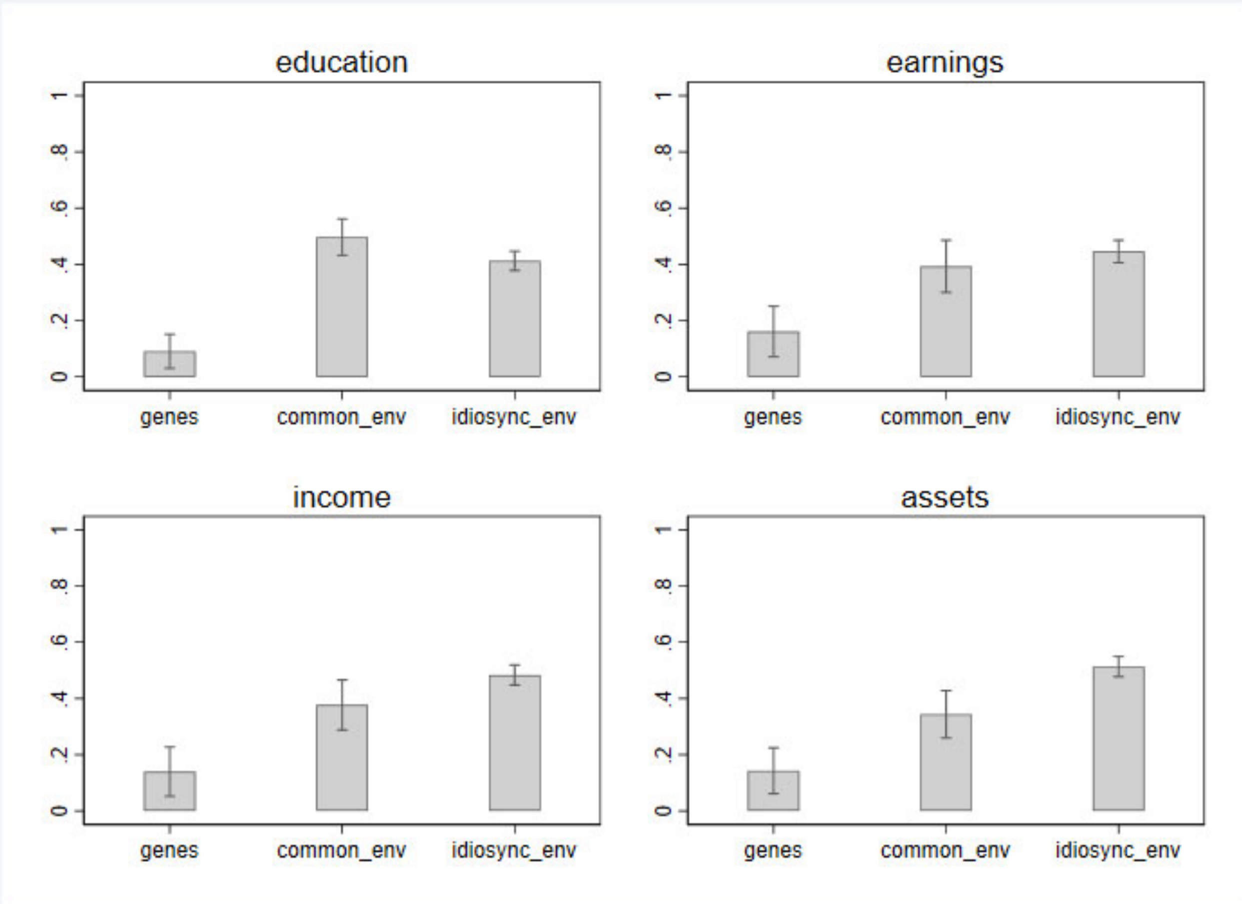
Note: The figure reports the decomposition of education inequality into genetic factors, common environment, and idiosyncratic (non-shared) environment. The estimates in the left panel are based on the CTD. In contrast, the estimates in the right panel are based on the extended version of the CTD, where we include twins and their spouses and allow for assortative mating based on genetic factors.

Figure 2.
Decomposition of Cross-Sectional Education Inequality.



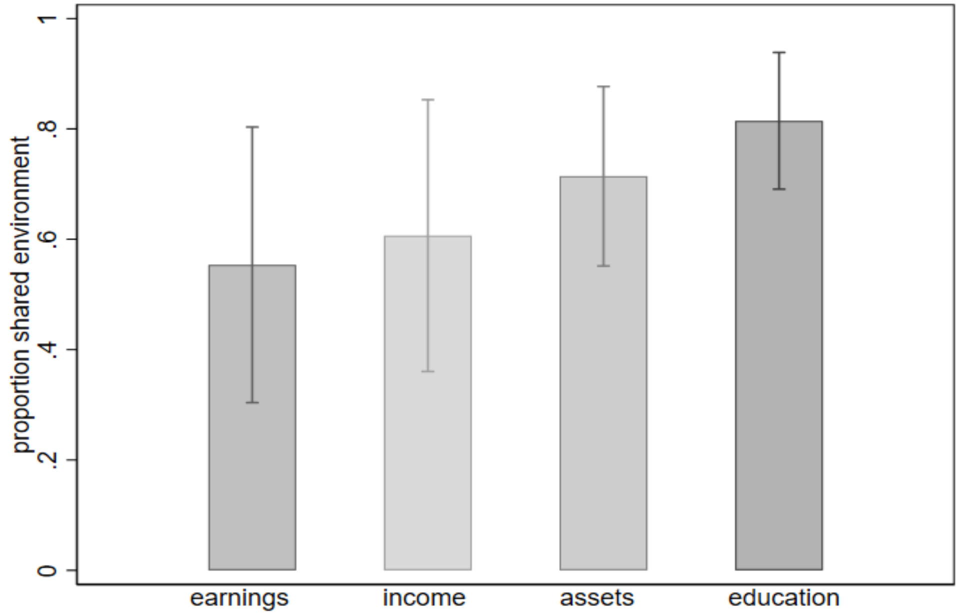
Note: The figure reports the decomposition of education inequality into genetic factors, common environment, and idiosyncratic (non-shared) environment. The estimates in the left panel are based on the classic twin design. In contrast, the estimates in the right panel are based on the twin family design, where we include twin spouses and their children and allow for differential shared environments and assortative mating based on genetic factors.

Figure 3.
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 4.
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

Table A1.
Education Correlations for Twin Families.

Spouses	0.41
Father-Son	0.30
Mother-Daughter	0.34
Father-Daughter/Mother-Son	0.29
MZ Twin Uncle-Nephew	0.28
MZ Twin Aunt-Niece	0.26
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.25
DZ Twin Uncle-Nephew	0.18
DZ Twin Aunt-Niece	0.20
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.18
Male Cousins (MZ)	0.26
Female Cousins (MZ)	0.28
Male-Female Cousins (MZ)	0.25
Male Cousins (DZ)	0.17
Female Cousins (DZ)	0.13
Male-Female Cousins (DZ)	0.15
Brothers-in-law (MZ-S)	0.28
Sisters-in-law (MZ-B)	0.33
Brother-Sister in-law (MZ-BS)	0.33
Brothers-in-law (DZ-S)	0.25
Sisters-in-law (DZ-B)	0.29
Brother-Sister in-law (DZ-BS)	0.28

Note: The table reports empirical correlations for different family relations. 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 A2.
Predicted Education Correlations.

	(1)		(2)
	Predictions sample A		Empirical 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) obtained from a random half of the sample described in Section 2 ("Half-sample A") and empirical correlations obtained from the remaining cases ("Half-sample B"). 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 A3. Additional Sample Descriptive Statistics (figures in Danish Kroner, 7 kroner = $1).

	MZ Families				DZ Families			
	Men		Women		Men		Women	
	Labor Earnings							
	Average	St. Dev.	Average	St. Dev	Average	St. Dev.	Average	St. dev.
Twin	390,145	213,221	287,563	139,657	378,865	211,881	269,115	129,532
Spouse	421,665	275,552	268,853	132,232	405,976	221,101	266,003	127,527
Child	445,206	287,678	336,204	160,143	439,555	263,103	331,125	161,701
	Disposable income							
	Average	Stand. Dev.	Average	Stand. Dev	Average	Stand. Dev.	Average	Stand. dev.
Twin	213,811	294,727	194,798	88,631	208,795	180,928	182,298	98,831
Spouse	236,232	533,995	185,539	150,977	217,489	256,575	181,813	88,557
Child	277,532	207,986	254,49	337,319	278,426	251,594	246,34	135,124
	Assets							
	Average	Stand. Dev.	Average	Stand. Dev	Average	Stand. Dev.	Average	Stand. dev.
Twin	871,462	3,700,697	539,683	1,141,923	898,814	4,450,132	445,541	839,337
Spouse	1215,27	5,324,230	488,341	1,649,102	1,073,106	4,404,984	436,620	838,674
Child	983,453	1,932,127	778,547	1,289,100	1,080,736	3,038,390	749,712	1,843,646

Note: The table reports means and standard deviations of labor earnings, disposable income, and assets (in Danish Kroner) separately for men and women and by zygosity.

Appendix Table A4. 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 Genetic Factors	0.09	(0.03)	0.15	(0.04)	0.12	(0.04)	0.13	(0.04)
Proportion Shared Environment	0.50	(0.04)	0.35	(0.04)	0.40	(0.04)	0.44	(0.04)
Women								
Proportion Genetic Factors	0.10	(0.03)	0.21	(0.05)	0.15	(0.05)	0.17	(0.05)
Proportion Shared Environment	0.49	(0.04)	0.41	(0.06)	0.36	(0.06)	0.22	(0.05)
				IGE				
IGE Men	0.24	(0.01)	0.21	(0.01)	0.22	(0.01)	0.31	(0.01)
Proportion Shared Environment	0.82	(0.06)	0.65	(0.09)	0.73	(0.09)	0.80	(0.06)
IGE Women	0.28	(0.01)	0.18	(0.01)	0.17	(0.01)	0.21	(0.01)
Proportion Shared Environment	0.82	(0.06)	0.41	(0.16)	0.56	(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.