On the origins of socio-economic inequalities: Evidence from a "children of twins" design

Paul Bingley — VIVE Copenhagen

Lorenzo Cappellari — Università Cattolica Milano

Konstantinos Tatsiramos — University of Luxembourg and LISER

September 2018

Abstract

We decompose sibling correlations and intergenerational elasticities of life-cycle earnings and education into pre- and post-birth factors within a unified framework that nests previous models. Using data on the Danish population of twins and their children we find that post-birth factors can explain a higher share of siblings' similarity in income and education. This share is higher than in previous studies because we allow for heterogeneity in the environmental influences across sibling types. We also show that pre-birth factors explain a higher share of the similarity in income and education between fathers and sons. Finally, we find that pre-birth factors matter more for income than for education, no matter which measure we consider.

Keywords: Sibling correlations; Intergenerational transmission; Children of twins

JEL Codes: D31, I24; J62;

I. Introduction

There is a large body of literature in economics on the origins of observed differences in socio-economic outcomes such as labour income and education. Some of these differences reflect idiosyncratic ability differences or different individual-specific life experiences. Others may be due to factors individuals share with members of their family or community and are more worrying from an equal opportunity perspective. Pre-birth factors through genetic differences and post-birth factors in the form of parenting or schooling differences are important examples of shared determinants of inequalities (for a review see Sacerdote 2011).

To shed light on the importance of pre- vs. post-birth factors the literature has used two approaches which rely on differences in the genetic connectedness between pairs of individuals. The first approach decomposes the sibling correlation (SC) of socio-economic outcomes into pre and post-birth factors using various sibling types, such as twins and adoptees. The sibling correlation is an omnibus measure of the many influences that sibling share both through the family and through environments external to the family, such as schools or communities. The second approach focuses on the parent-child transmission of outcomes, measured by the intergenerational elasticity (IGE), using adoptees. The two measures are linked with recent evidence showing that parent-child transmission is the main driver of sibling correlations, implying that the resemblance of siblings in socio-economic outcomes stems mainly from factors that parents transmit to their children (Bingley and Cappellari, 2018).

What is the relative weight of pre- and post-birth factors in shaping the correlation of sibling outcomes and parent-child transmission? Existing evidence derived from separate models of sibling or parent-child correlations suggests that the resemblance of sibling earnings stems largely from a significant role of *pre-birth factors* (Björklund, Jäntti and Solon, 2005), while *post-birth* environment is more important in explaining the similarity between fathers and sons (Bjorklund, Lindahl and Plug, 2006). This is somehow surprising, because by design the sibling correlation is a broader measure of shared influences and should reflect to a greater extent the role of post-birth environments compared with the IGE. In this paper we develop a more general approach encompassing sibling correlations and intergenerational links, and we show that the existing evidence substantially overestimates the importance of pre-birth factors in explaining sibling similarity, while it underestimates it for understanding the resemblance between fathers and sons.

Our contribution is to provide a unified framework for analysing the origins of inequality of socio-economic outcomes, which is based on a Children of Twins design (CoT) using data on the universe of twins from Denmark and their children. There are three key advantages of a CoT design. First, it combines the two existing approaches under a single framework because we observe not only information on outcomes within a generation (twin siblings), but also between generations (the children of twins). Therefore, we can decompose the sibling correlation of socio-economic outcomes and the intergenerational elasticity into pre- and post-birth effects under the same set of assumptions. Second, the CoT design allows relaxing some of the maintained assumptions of canonical twin studies. In particular, we (i) relax the equal environment assumption of monozygotic (MZ) and dizygotic (DZ) twins, (ii) allow for parental assortative mating on genes, and (iii) explore the importance of gene-environment interactions. Finally, the CoT design allows decomposing the sibling correlation into pre and post-birth factors using only siblings reared together and the intergenerational elasticity using only fathers and their biological children. Previous studies have relied on non-intact families either comparing siblings reared together with those reared apart or using adoptees.

We find that *post-birth factors* can explain a higher share of *sibling* similarity in income and education than pre-birth factors. We show that allowing for heterogeneity in the environmental influences across MZ and DZ twins reduces the importance of pre-birth factors in explaining differences in these outcomes. Pre-birth factors explain a higher share of the similarity in income and education between fathers and sons, but they explain less than half of sibling resemblance because siblings are influenced not only by their parents but also by factors outside the family.

Finally, we find that pre-birth factors matter more for income than for education, no matter which measure (SC or IGE) we consider. This suggests that when outcomes are produced inside institutions (e.g. in schools), rather than in the market, the impact of differences in genetic endowments can be mitigated making environmental heterogeneity predominant in the variance decomposition.

II. Overview of related literature

The literature has used two approaches to shed light on the importance of pre- vs. post-birth factors in explaining variation of socio-economic outcomes. The first approach focuses on the resemblance of twins and other sibling types to explain the sibling correlation of outcomes, while the second approach considers the resemblance of adopted children with their biological and adoptive parents looking at the intergenerational elasticity.

A. Decomposition of sibling correlation

Twin studies rely on the fact that MZ (identical) twins have the same set of genes while DZ (fraternal) twins have, on average, a half-identical set. A basic assumption in these studies is that each pair of twins is affected by their environment to the same degree. This means that the environment experienced by MZ twins does not make them similar to a greater degree than the environment experienced by DZ twins makes them to be similar. The equal environment assumption implies that if MZ twins are more similar than DZ twins in some outcome this is driven by their greater genetic similarity. However, this assumption might be strong because there are factors that make the environments of MZ twins more similar that the environments of DZ twins. For example, MZ twins may perceive themselves or being treated by their parents or their teachers as if they are more alike. In addition, MZ twins are same-sex, while one-third of DZ twins are opposite-sex. If the equal environment assumption is not valid then the role of pre-birth factors will be overstated.

Existing evidence based on the decomposition of sibling correlations using twins and other sibling types suggests a very small role of family environment in explaining the variation in income and education. Pre-birth effects (genes) explain 88 per cent of the variation in education (Behrman and Taubman, 1989) and 28 per cent of the variation in earnings (Björklund, Jäntti and Solon, 2005), which accounts for about 80 per cent of the earnings correlations for twin brothers and non-twin full brothers. The importance of pre-birth factors on explaining the variation in earnings reduces to 19.9 per cent after allowing for different environmental influences among sibling pairs, but still explains about 60 per cent of the resemblance among siblings.

This approach to the analysis and decomposition of sibling correlations derives from the more general framework of behavioural genetics, which was developed outside economics to study phenotype correlations across extended groups for whom the expected degree of genetic resemblance is known a priori.[1] Recent applications in economics to inequality in outcomes different from income or education still make use of the canonical model based on twins or siblings (see e.g. Cesarini et al. 2009 and Fagereng, Mogstad and Rønning, 2018). Indeed, ours

[1] There is also a growing number of papers in economics using the *molecular genetics* approach, which takes advantage of the availability of measured DNA sequences. One difference between approaches is that while molecular genetics explains variation in outcomes due to genes using the observed variability in both genes and outcomes, behavioural genetics makes assumptions that enable capturing all genetic determinants of outcomes inequality, both observable and unobservable.

is the first paper that uses a CoT design to apply the behavioural genetic approach on the economic outcomes of extended family groups, and we exploit this novel research design to relax some assumptions of the canonical model (see Section 4). To the best of our knowledge, so far CoT designs have been used in economics to integrate out family fixed effects in intergenerational regressions (e.g. Behrman and Rosenzweig, 2002).

B. Decomposition of intergenerational elasticity

Adoption studies used in economics compare the intergenerational elasticity of socio-economic outcomes of adopted children with their biological and adoptive parents. If the transmission coefficient of adoptive parents is lower than that of biological parents this is evidence for some genetic influences. Instead, when adoptees resemble their adoptive parents more than their biological parents this suggests some environmental influences.

Existing evidence suggests environmental factors transmitted from adoptive fathers are more important than pre-birth factors transmitted from biological fathers in explaining adoptees' earnings (Björklund, Lindhal and Plug, 2006). For education, pre-birth and post-birth factors have overall similar influences (Björklund, Lindhal and Plug, 2006, Sacerdote, 2007).

As reported in Sacerdote (2011), there is often a disconnect between the twin and adoption literatures about the importance of family environment. For example, in the case of studies focusing on earnings summarized above, twin studies find that genes explain a higher proportion of variance in earnings, while adoption studies find that family environment explains a larger proportion of the variation in earnings. Adoptees studies may exaggerate the role of the environment because they estimate it using the environments provided by adoptive parents, who may put an extra-effort to counterbalance the lack of genetic transmission, so they are not representative of the full population of parents. With our research design based on children of twins we reconcile this disconnect by combining both approaches under a single framework, which in addition does not require separation of children from their biological parents either through separation or adoption.

III. Data

We use data from administrative registers of the Danish population and from the Danish Twins Registry. The civil registration system was established in 1968 and everyone resident in Denmark then and since has been registered with a unique personal identification number which has subsequently been used in all national registers enabling accurate linkage.

The Twins Registry has identified more than 65,000 twin pairs born since 1870 through parish and hospital records (Skytthe et al. 2002). Zygosity is established for same-sex twins on the basis of responses to four survey questions about twin similarity; a method validated to an overall accuracy of 96 percent (Christiansen et al. 2003). We sample all MZ and DZ twin individuals known to the Twins Registry. Links from children to legal parents originate from municipal and parish records and are complete for births from 1955 onwards. Using this linkage, we find all children of these twins; then we find co-parents of these children.

Our aim is to construct a twin family dataset centred around a single twin pair spanning two generations. In the cases where both parents are twins, we randomly select the focal twin pair (alternatively we could keep all twin pairs and re-use twins that appear twice). In the cases where the children are also twins and have children, we drop the third generation (alternatively we could keep the third generation and re-use generation 2-3 as additional generation 1-2 observations). Finally, we keep only men in our twin family dataset.

We observe annual pre-tax labour earnings obtained from income tax returns. Each January employers report earnings for each employee for the previous year to the tax authorities, and in March the tax authorities send these returns to the employees themselves for verification. We use the sum of earnings from all employments during the year for the period

1980-2011 over which it is available in the Statistics Denmark Income Statistics Register; see Baadsgaard and Quitzau (2011) for a detailed description of Danish income registers.

Schooling information is reported by educational institutions to the Ministry of Education; see Jensen and Rasmussen (2011). Highest level of education is calculated by Statistics Denmark on the basis of information from the Ministry about prerequisites and normed times for completing each qualification. Highest level of education is defined as the qualification that would take the longest time for an individual to obtain by the shortest possible route.

We select the working sample with the primary goal of obtaining estimates of permanent incomes. Because the twin population is small, we cannot afford estimating permanent incomes via a fully-fledged multi-person model of life-cycle earnings dynamics that enables handling measurement errors coming from transitory income shocks and life cycle biases (see e.g. Bingley, Cappellari and Tatsiramos, 2017). We therefore resort to a simpler model with constant permanent incomes to separate permanent from transitory shocks, and deal with life cycle bias by limiting the observation of current incomes in the 30-45 age range, where life-cycle biases are considered minimized (see Nybom and Stuhler, 2016). Specifically, we consider individuals that potentially can be observed at least 5 times in the 30-45 age window, which, coupled with the availability of income data for 1980-2011, means that we sample persons born between 1939 (aged 41 in 1980) and 1977 (aged 30 in 2007).

We keep one son (if present) per father and exclude second born children of twins. We also exclude families where the father-son age spacing is below 18 years or above 50, or where the cousins' age spacing is above 15. Finally, we exclude twins and twin spouses that are not in the birth cohort range of CoTs' fathers. We do this by excluding families where the oldest twin or twin spouse is younger than the youngest father of a CoT. In this way we limit mixing childlessness with censored fertility or with the fact that children are still younger than 30 and cannot contribute data to the permanent income process.

We describe the resulting working sample in Table 1. Panel A reports sample sizes. Data refer to about 10000 twin families, and more than 80 percent are the families of DZ twins. Of the overall number of families, the majority (6140) has 2 fathers (a father being either a twin or a twin spouse), and slightly less than half (4645) has no children that are observed in the 30-45 age range, while about another third (3631) has one child and the remainder two children. There are about 1000 families with two fathers and two children. Remaining figures in Panel A provide the distribution of observations by person type and zygosity.

In Panel B of Table 1 we provide some descriptive statistics on the outcomes of interest by reporting their raw correlations within the family. In the left column we consider current incomes which are a mixture of permanent incomes and transitory shocks, therefore we expect their correlations between family members to underestimate correlations in permanent incomes. We distinguish correlations by zygosity of the twin couple in the family whenever the link between relatives goes through the twins, which is true in every case with the exception of father-son correlations. Comparing MZ and DZ twins, we see that the latter correlation is about half the former, which is consistent with the idea that MZ twins share a larger fraction of pre-birth endowments compared to DZ twins but may also reflect heterogeneous environmental similarities between MZ and DZ twins. Correlations of current incomes between fathers and son are of a similar size as the ones for DZ twins, and also similar to the ones for uncle-nephews, cousins or brothers in law belonging to the families of MZ twins, while for the relatives of DZ twins all these correlations are generally lower, pointing again to the relevance of pre-birth factors in shaping correlations in outcomes.

In the right column of Table 1, Panel B reports family associations in years of education. Two facts are worth noting. First, these correlations are generally much larger (more than five times) compared to current incomes, which may reflect the fact that in the case of education the degree of measurement error on the relevant outcome is limited compared to the case of current incomes which are a poor proxy of permanent incomes. Secondly, and remarkably, there is no much variation in correlations comparing members of MZ and DZ families, which points to a secondary role of pre-birth factors in shaping educational inequalities.

IV. Empirical Approach and Identification

A. The Canonical Model

Past research has been using data on twins with information on their zygosity to decompose the variance of various outcomes into pre- and post- birth components with the so-called *ACE model*. Let y_i denote the long-term outcome of person *i* (in our case either permanent incomes or years of education) in deviations from the population mean and consider the following factorisation:

$$y_i = a_{f(i)} + c_{f(i)} + e_i\,, \qquad (1)$$

where *f* is the family of the individual, $a_{f(i)}$ is an *additive genetic factor* shared by the members of family *f*, $c_{f(i)}$ is a *common environmental factor* shared by the members of family *f*, while e_i is a *unique environmental factor* capturing idiosyncratic deviations within the family. Factors are drawn from zero mean distributions with variances σ_a^2, σ_c^2 and σ_e^2 and are assumed orthogonal. Orthogonality between *e* and each of *a* and *c* follows from the definition of idiosyncratic deviation. Orthogonality between *a* and *c* is an assumption of the model, which excludes the possibility that certain genes may select into certain environments. To operationalize the model, three further assumptions are usually introduced. First, that there is *no assortative mating on genes* between spouses, implying that DZ twins (or indeed non-twin siblings) share on average half of their genes. Second, that there is *equal environment* for MZ and DZ twins, excluding for example the possibility that parents treat MZ twins differently (and perhaps more identically) than what they would treat their DZ twins. Third, that the three components enter the model in an additive linear manner, excluding the possibility of *gene-environment interactions*, i.e. of any variation in genes' expression with environmental exposure.

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

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

The covariance of outcomes between MZ twins *i* and i' is:

$$cov(y_i, y_{i'})_{MZ} = \sigma_a^2 + \sigma_c^2 \ . \qquad (3)$$

Finally, the covariance of outcomes between DZ twins *i* and i' is:

$$cov(y_i, y_{i'})_{DZ} = 0.5\sigma_a^2 + \sigma_c^2 \qquad (4)$$

because they only share half of their genes under no assortative mating. Note also that the environmental variance component σ_c^2 is the same for MZ and DZ twins due to the equal environment assumption.

Equations (2) – (4) identify the three variance components of the model and enable decomposition of the variance of outcome y_i into pre-birth, post-birth and idiosyncratic factors, providing insights on the degree of heritability. Past research has been using estimated variance components to decompose the *sibling correlation* between pre- and post-birth factors, see for example Björklund, Jäntti and Solon (2005).

B. Extending the Canonical Model with a Children of Twins (CoT) design

We contribute to the literature with the introduction of a CoT design for the analysis of heritability in permanent incomes and education. Besides information on twins' outcomes and their types (MZ or DZ), a CoT design also exploits information on the outcomes of children of twins, which enables considering family ties not only within the twins' generation, but also between generations. Moreover, it exploits *avuncular relationships*, that is, relationships between individuals belonging to different *nuclear families* within the same *extended family*; for example, between cousins or between uncles and nephews. The advantage of a CoT design over the canonical twin model is twofold. First, the larger amount of information contained in CoT data can be used to relax the assumptions of the canonical model. Second, the CoT design allows estimating the impact of heritability on intergenerational transmission, thus providing a direct answer to the long-standing question about the extent to which parent-child transmission depends on factors already determined at birth. Such a direct decomposition is not feasible with the canonical twin model that only considers associations within a generation. Indeed, such a possibility is precluded also to research designs exploiting varieties of sibling types, such as in Björklund, Jäntti and Solon (2005).

Using the CoT design allows observing a multiplicity of family links that are functions of the genetic component, which, in turn, enables relaxing the equal environment assumption, allowing shared environmental factors to be drawn from different distributions depending on the type of family link. For the same reasons, we can allow for parental assortative mating on genes by letting the genetic similarity of DZ twins (or between parents and children) to be shifted by an unknown parameter to be freely estimated (a parameterisation used by Bowles and Gintis 2002). However, the presence of multiple family relationships is not helpful in identifying a parameter for gene-environment correlation, because that parameter would turn out to be ubiquitous throughout the equations of the extended model and thus not separable from the genetic component. We, therefore, conduct the analysis with the CoT under the assumption of no genes-environment interaction. Similar remarks about lack of identification apply to interaction terms between the genetic and environmental component. To explore the relevance of gene-environment interactions, in Section V.D we will consider changes in the estimates of the genetic factor comparing families who are exposed to different environments.

Because one outcome of interest is permanent income, we focus our attention on men to limit issues of endogenous female participation in the age range when we source income data (30-45). We use information on twins, their sons, and in the case of female twins on their spouses. We relax the equal environment assumption and allow environmental sharing to depend on the type of family link which is analysed.

For MZ twins, the covariance of outcomes becomes

$$cov(y_i, y_{i'})_{MZ} = \sigma_a^2 + \sigma_{cMZ}^2 \; , \qquad (3')$$

where σ_{cMZ}^2 is the MZ-specific variance of shared environmental factors. For DZ twins, the covariance becomes

$$cov(y_i, y_{i'})_{DZ} = (0.5 + \delta)\sigma_a^2 + \sigma_{cDZ}^2 \; , \qquad (4')$$

where, besides including a DZ-specific variance of shared environments, we also allow for parental assortative mating on genes, such as the genetic resemblance of DZ twins may differ from one-half by an unknown factor δ to be estimated. Positive values of δ would be consistent with the idea of positive assortative mating, implying that DZ twins share more than a half of their genes.

CoT designs generate restrictions also for intergenerational family links such as between fathers and sons or between uncles and nephews. For fathers and sons, the covariance of long-term outcomes is

$$cov(y_i, y_{i'})_{FS} = (0.5 + \delta)\sigma_a^2 + \sigma_{cI}^2 \; , \qquad (5)$$

which allows environmental sharing to be specific to the father-son relationship through the parameter σ_{cI}^2. In the presence of parental assortative mating on genes, the genetic resemblance of fathers and sons will differ from one-half by the assortative mating factor δ. For uncles and nephews, the covariance is

$$cov(y_i, y_{i'})_{UN} = \delta^{I(S)}(0.5 + \delta)^{(1+I(DZ))}\sigma_a^2 + \sigma_{cUN}^2 \; , \qquad (6)$$

where $I(S)$ is a dummy which is equal to one for the male spouses of twins, $I(DZ)$ is a dummy for DZ twins, while σ_{cUN}^2 parameterises environmental sharing in avuncular relationships between generations. Expression (6) states that if uncle i is the MZ twin of $i's$ father (and therefore $I(S) = 0$ and $I(DZ) = 0$), then the uncle-nephew genetic correlation is the same as the father-son one (in equation 5), while the environmental sharing will be specific to the uncle-nephew relationship. Instead, if uncle i is the DZ twin of $i's$ father (and therefore $I(S) = 0$ and $I(DZ) = 1$), then the uncle-nephew genetic resemblance will be lower because in that case DZ twins share only a fraction of their genes equal to $(0.5 + \delta)$, so the genetic resemblance in the next generation will be equal to $(0.5 + \delta)^2$. Finally, if uncle i is the spouse of the MZ or DZ twin of $i's$ parent (in which case $I(S) = 1$) all the above uncle-nephew genetic relationships will be further mediated by the assortative mating parameter (δ); if there was no assortative mating ($\delta = 0$), then in that case there would be no genetic association between nephews and uncles.

CoT designs also provide information on avuncular relationships within generations such as between cousins and between brothers-in-law. The covariance of long-term outcomes between cousins is

$$cov(y_i, y_{i'})_{CC} = (0.5 + \delta)^{(2+I(DZ))}\sigma_a^2 + \sigma_{cCC}^2 \; , \qquad (7)$$

where environmental sharing is allowed to depend on the cousin-specific parameter σ_{cCC}^2 and their genetic resemblance depends on whether their parents are MZ or DZ twins. The covariance between brothers-in-law is

$$cov(y_i, y_{i'})_{BB} = (\delta)^{(I(S)+I(S)')}(0.5 + \delta)^{(I(DZ))}\sigma_a^2 + \sigma_{cBB}^2 \; , \quad (8)$$

where, again, genetic resemblance emerges through assortative mating whenever twin spouses are involved, with the assortative mating coefficient δ raised to the power of 1 or 2 depending on whether only one or both brothers-in-law are twin spouses. The shared environment for brothers-in-law is captured by a specific parameter σ_{cBB}^2.

In sum, identification of the various variance components is achieved by exploiting expected genetic resemblance of the various couples constituting the CoT extended family and thanks to the possibility of having couple-specific environmental factors. Using these parameters, we can compute the sibling correlation of permanent incomes as

$$\rho_{MZ} = \frac{\sigma_a^2+\sigma_{cMZ}^2}{\sigma_a^2+\sigma_{cMZ}^2+\sigma_e^2} \qquad (9)$$

$$\rho_{DZ} = \frac{(0.5+\delta)\sigma_a^2+\sigma_{cDZ}^2}{\sigma_a^2+\sigma_{cDZ}^2+\sigma_e^2}, \qquad (9')$$

where each expression lends itself to be additively decomposed into genetic and environmental factors. We can also compute and decompose the intergenerational elasticity (IGE) given by

$$\beta = \frac{(0.5+\delta)\sigma_a^2+\sigma_{cI}^2}{\sigma_a^2+.34\sigma_{cMZ}^2+.66\sigma_{cDZ}^2+\sigma_e^2}, \qquad (10)$$

where we have used the fact that about a third of twin couples are MZ.

Years of education are observable in the data, but permanent incomes are not and we only observe annual incomes, that are a mixture of permanent incomes and transitory income shocks. To identify permanent incomes separately from transitory income shocks we use a simple income dynamics model similar to the analysis in Björklund et al. (2009) of sibling correlations, which postulates a time invariant permanent component with AR(1) transitory shocks. Differently from them, we allow for non-stationarity of the AR(1) and for aggregate time-shifters on income components. In sum, our income dynamics model is the following:

$$w_{it} = \mu_t + \pi_t y_i + \tau_t v_{it}, \qquad (11)$$

where w_{it} is person i's year t current income in logs, μ_t its period-specific mean, π and τ are period-specific shifters on the permanent and the transitory component. Transitory shocks are specified as non-stationary AR(1)

$$v_{it} = \lambda v_{it-1} + \varepsilon_{it} \ ,$$

where ε_{it} is a White Noise innovation with variance σ_{ε}^2 and the process initial condition $v_{i\tilde{t}(i)}$ is individual-specific with variance $\sigma_{\tilde{v}}^2$.

Empirically, we remove the period-specific mean from current incomes via a first stage regression run by person type (whether a twin, a twin spouse or a child of a twin), and impose the restrictions implied by the model on the empirical covariance structure of log residuals using the Equally Weighted Minimum Distance (EWMD) estimator.

V. Results

A. Decomposition of permanent earnings

We report parameter estimates for the model of permanent incomes variance components in Colum (1) of Table 2 Panel A. Most of income dispersion is purely idiosyncratic, as shown by the estimate of σ_e^2 being larger than those of all other variance components of the model. Particularly it is roughly three times larger than the genetic (pre-birth) component. Estimates of shared environmental factors show considerable heterogeneity which is in line with what could be expected a priori. That is, environmental resemblance is larger for MZ twins than between DZ twins, and it is also larger for brothers than for other pairs in the model. However, formal tests of hypothesis never reject the equal environment assumption, either globally (the $\chi^2(5)$ statistic for the hypothesis that all environmental coefficients are the same is equal to 4.7 with a p-value of 0.45) or for sub-groups (for example the $\chi^2(1)$ statistic of equal environment for MZ and DZ twins is equal to 0.82 with a p-value of 0.36). Finally, the estimated assortative mating parameter is sizeable ($\delta = 0.17$) but rather imprecisely estimated.

In Panel B, we use these estimates to decompose relevant statics of permanent income inequality in into pre- and post-birth shares. For MZ twins, the sibling correlation of incomes is 0.43, implying that 43 percent of income inequality is due to factors shared by MZ twins, while 57 percent is due to within-twins idiosyncratic variation. The correlation is lower for DZ twins and equals 0.28 (close to estimates reported in Bingley et al. 2017 for the full population of Danish brothers). This lower value not only depends on the fact that pre-birth factors have full impact among MZ couples while they only partially affect (through the fraction $(0.5 + \delta)$) the correlation for DZ ones, but it also comes from the fact that environmental commonalities are larger for MZ twins that for DZ ones. In each case, pre-birth factors account for less than half of the sibling correlation (38 percent for MZ and 45 for DZ), while most of sibling resemblance is explained by shared environments post-birth.

The other relevant statistic in Panel B of Table 2 is the intergenerational elasticity (IGE) which measures parent-child transmission of outcomes. Thanks to the use of the CoT design, our is the first paper allowing a comparison of pre- and post-birth decompositions between and within generations within a unified framework. We estimate an IGE of 0.18 of which roughly two-thirds can be interpreted as an expression of pre-birth factors.[2] This result is in stark contrast with the evidence for within-generation inequality obtained for the sibling correlations. The reason is that the environmental sharing of fathers and sons is much lower than the one of twins (it is about a half of that of DZ twins and a quarter of that of MZ twins), which highlights the importance of relaxing the common environment assumption. Had we imposed common environment, the pre-birth share of the IGE would have – by construction – coincided with the one of DZ twins. The differential importance of shared environments within and between generations is consistent with the idea that while within a generation siblings (twins) are exposed to a common context that is partly external to the family, between generations there is

[2] This estimate is somewhat larger than other estimates for Denmark. For example, Björklund and Jäntti (2009) report an IGE of 0.12 but this estimate refers to the full population of fathers and their sons, while in our case fathers belong to the sub-population of twins.

less scope for environmental influences that are orthogonal to parental endowments; this makes the purely environmental sharing of fathers and sons less relevant than the ones of brothers.

Our results show that pre-birth factors are more relevant in accounting for family associations in permanent incomes between than within generations, and the intuition is that within a generation environmental factors operating post-birth have more substantial impacts. It is interesting to compare our results with findings in the literature. Within generations, Björklund et al. (2005) use varieties of sibling types to decompose the sibling correlation relaxing the equal environment assumption and find that 55 (60) percent of the sibling correlation for MZ (DZ) twins reared together originates pre-birth. Between generations, Björklund et al. (2006) use data on biological and adopted parents and find that most (67 percent) of intergenerational transmission in earnings originates post-birth. There is tension between these findings because one would expect environmental factors to be more relevant within than between generations, which typically motivates the sibling correlation as omnibus measure of sibling similarities that encompasses intergenerational associations. This tension is not present in our findings, which indicate that the larger share of post-birth environments operates within generations. One reason why studies that use varieties of sibling types may overstate pre-birth components is because they only exploit information on within-generation relationships which may be influenced by common environmental factors operating outside the family (e.g. schools or youth communities) that are not captured by the post-birth parameters of the model. On the other hand, studies contrasting biological and adoptive intergenerational links assume that the environmental influences of adoptive parents are representative for the whole population, which might not hold if adoptive parents put extra parental effort to compensate for lack of pre-birth endowments.

Our estimates suggest that parental assortative mating on genes is not a relevant mechanisms of income associations within the family. Also, in some cases the assumption of common environment does not seem restrictive, particularly for blood-sharing avuncular relationship (uncles-nephews and cousins).[3] In Column (2) of Table 2, we therefore estimate a restricted version of the model which imposes the two restrictions i.e. $\delta = 0$ and $\sigma^2_{cUN} = \sigma^2_{cCC}$. While estimates' precision is generally increased in this restricted version, results still point to the same main evidence of Column (1), namely the importance of allowing for differential environments and the different importance of pre-birth factors that emerges if one considers the intergenerational perspective rather than looking at sibling pairs.

B. Decomposition of years of education

To benchmark our results on family-related variance components of the income process, in Table 3 we present parameter estimates obtained by applying the model to decomposing statistics on educational attainment. In particular, the outcome variable in the table are years of education in logs, where the log transformation ensures that we can interpret the regression coefficient of sons' outcomes on fathers' as an elasticity (results obtained using years of education in levels are very similar to the ones presented). Parameter estimates in Panel A Column (1) show that also in the case of education, the estimates of parental assortative mating are positive but rather imprecise; moreover, the environmental parameter specific of cousins appears to suffer from some empirical identification issues resulting in a point estimate that is 2 orders of magnitude lower that the one of other model parameters. Both these considerations

[3] When the uncle is the spouse of the twin he will not be linked with the nephew by blood; however, we keep referring to uncles-nephews links as blood-sharing in general, in the sense that this group includes some couples that are related by blood, as opposed to brothers-in-laws who represent a group of avuncular links in which no couple in the group is blood-related.

suggest the possibility of applying the restricted model with no assortative mating and unique environment for blood-sharing avuncular relationships, reported in Column (2).

Findings about educational variance decompositions in Panel B of Table 3 tell a different story compared with income. Both sibling correlations and the IGE for education are higher compared with permanent incomes, something that has already been documented in other countries (see Björklund and Jannti, 2014). The estimated share of these statistics that can be ascribed to pre-birth factors is lower compared with the case of permanent incomes, no matter which statistic or version of the model is considered. The low prevalence of pre-birth factors (and symmetrically a higher shared for environmental ones between brothers) suggests that, even in a context characterised by a homogeneous educational system like Denmark, families can to a great extent choose environmental characteristics that have an impact on children's educational achievement. For example, residential choices may be such as to favour enrolment in schools with better peers. Another relevant finding is that the common environment assumption seems to hold for MZ and DZ twins, suggesting that social interactions with peers in schools may not vary much by zygosity. The contrast with results for permanent income suggests that for outcomes generated through market interactions (i.e. permanent incomes) rather than within an institution (like education), factors passed on by parents and – at least in part – already determined at birth (such as risk aversion or competitive attitudes) have a greater expression.

C. Comparison with the canonical model

Our CoT design nests the canonical twin model used in previous studies. In Table 4 we exploit this property to present estimates obtained from the canonical model, which can be helpful in shedding some light on the additional insights that can be gained thanks to the CoT design. The model is estimated on twin data only, therefore we focus the attention on decomposing the sibling correlation and we do not consider intergenerational links. Compared to the full CoT model, results of the canonical model tend to overstate the importance of pre-birth factors in generating income variation and to underestimate their role in educational inequalities.[4] These apparently contradicting findings are both symptoms of the fact that the canonical design identifies the role of pre-birth factors only out of differences between twin types (MZ vs. DZ) without further anchoring the estimates using other couples of family members, assuming common environment. In turn, this leads to attribute most of shared inequality to environmental factors where environmental influences are similar between twin types (i.e. years of education) and to overstate pre-birth variation when environmental influences are different between twin types (i.e. permanent incomes). In the latter case, the additional environmental commonality between MZ and DZ twins that is assumed away by the canonical model is loaded by the pre-birth component.

D. Gene-environment interactions

In Table 5, we explore how the estimates change if we allow for gene-environment interactions. We address gene-environment interactions by investigating differences in the relevance of pre-birth factors across families characterised by presumably different environments. In particular, we consider differences in the gender mix of the twin couple. There are several papers arguing that gender composition of siblings may impact on the development of fundamental traits such

[4] For example, the share of sibling correlation in income explained by pre-birth factors increases from 0.37 (for MZ) and 0.45 (for DZ) in the CoT model to 0.88 and 0.79, respectively, in the canonical model. For education, the share of pre-birth factors in the sibling correlation decline from 0.24 (MZ) and 0.11(DZ) in the CoT model to 0.11 and 0.06, respectively, in the canonical model.

as risk aversion or competitiveness (e.g. Butcher, K. F., & Case, A., 1994). To the extent that these traits depend – at least partly – on pre-birth environments, we should observe the impact of pre-birth factors on outcome associations within the family to vary with the gender composition of the twin pair.

We proceed by splitting the data using the gender composition of the twins (whether all men, all women or mixed) and by re-estimating the empirical second moments for these split samples. Next, we pool all empirical moments from the split samples in estimation of the model, and we allow the pre-birth parameter to shift by gender composition of the twins. Because we use only data on men, data for twins will not effectively contribute to estimating the environmental shift of the pre-birth parameter, since there is no variation in gender composition for twins in the estimating sample. All other family links do not necessarily imply that the twins are two men and will therefore feature the environmental shift of the pre-birth parameter. One implication is that among the statistics of interest (the sibling correlation and the IGE), the only one that can meaningfully be allowed to change because of the gene-environment interaction is the IGE.

Results in Table 5 show some heterogeneity in the relevance of pre-birth components across families that are characterised by presumably different types of environment, even though differences across family types are not statistically significant. The general pattern emerging from this exercise is that in environments characterised by larger heterogeneity (i.e. in families where the twins are of mixed gender) the pre-birth component has the larger impact on intergenerational transmission. The literature on gender interactions in the family has shown that there may be a reduction of gender-stereotyped behaviours in mixed-gender siblings (e.g. Peter, Lundborg, and Webbink, 2015). Our results suggest that in such contexts pre-birth factors increase their relevance for explaining intergenerational transmission.

VI. Conclusion

Family background is an important determinant of inequality of socio-economic outcomes later in life but whether this influence is pre-determined at birth or whether investments from parents play a role remain open questions. The two most common research designs based on twin data and data on adoptees have reached different conclusions. In this research we provide a unified framework based on a Children of Twins design, which nests previous models and allows to decompose the sibling correlation of socio-economic outcomes and the intergenerational elasticity into pre- and post-birth effects under the same set of (weaker) assumptions using only twin parents and their biological children.

We find that post-birth factors explain a higher share of siblings' similarity in income and education. This share is higher than in previous studies because we allow for heterogeneity in the environmental influences across sibling types. Pre-birth factors explain a higher share of the similarity in income and education between fathers and sons, but they explain less than half of sibling resemblance because siblings are influenced not only by their parents but also by factors outside the family. Finally, we find that pre-birth factors matter more for income than for education both when looking at siblings or parent-child correlations.

These findings suggest that not all of the influence from family background is already determined pre-birth but rather that there is an important influence from parenting and the environment produced in the family. Furthermore, when outcomes are produced inside institutions, such as schools, rather than in the market, the impact of differences in endowments can be mitigated. These findings point to the role of policies targeted on parents and institutions who create those environments.

REFERENCES (Incomplete)

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.

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.

Bingley, P. and L. Cappellari (2018). Correlation of Brothers' Earnings and Intergenerational Transmission. *Review of Economics and Statistics*.

Bingley P., L. Cappellari and K. Tatsiramos (2017). Family, Community and Life-Cycle Earnings: Evidence from Siblings and Youth Peers. CESifo Working Paper No. 6743.

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. *The Quarterly Journal of Economics*, 121(3), 999-1028.

Bowles, S. and H. Gintis (2002). The Inheritance of Inequality. *Journal of Economic Perspectives*, 16(3), 3-30.

Butcher, K. F. and A. Case (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, 109(3), 531-563.

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

Fagereng, A., M. Mogstad and M. Rønning (2018). Why Do Wealthy Parents Have Wealthy Children? Working paper.

Peter, N., P. Lundborg and D.H. Webbink (2015). The Effect of Sibling's Gender on Earnings, Education and Family Formation. IZA Discussion Paper No. 9128.

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.

Sacerdote, B. (2007). How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees. *The 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.

Table 1: Descriptive statistics

a) Number of observations

	Families	Twins	Twins' Spouses	Children of Twins
MZ	1760	1642	1195	1353
DZ	8202	7993	5808	5534
Total	9962	9235	7003	6867
Cohort 1940		5438	4300	0
Cohort 1950		3530	2332	55
Cohort 1960		259	164	2767
Cohort 1970		8	11	4181

b) Raw outcome correlations within twins' families

	Current Income		Education	
	MZ	DZ	MZ	DZ
Brothers	0.12	0.07	0.53	0.50
Father-Son	0.05		0.28	
Uncle-Nephew	0.05	0.03	0.25	0.21
Cousins	0.06	0.02	0.18	0.17
Brothers in Law	0.05	0.03	0.22	0.22

Table 2: Parameter estimates and decompositions for permanent incomes

	(1) Full model		(2) Restricted model	
	Coeff.	S.E.	Coeff.	S.E.
A) Variance components				
Unique Environment (σ_e^2)	0.0519	0.0086	0.0521	0.0086
Additive Genetic (σ_a^2)	0.0147	0.0109	0.0172	0.0097
Shared Environment				
MZ Twins (σ_{cMZ}^2)	0.0242	0.0122	0.0219	0.0112
DZ Twins (σ_{cDZ}^2)	0.0119	0.0081	0.0132	0.0064
Fathers and Sons (σ_{cI}^2)	0.0053	0.0072	0.0065	0.0052
Uncles and Nephews	0.0039	0.0075		
(σ_{cUN}^2)			0.0056	0.0024
Cousins (σ_{cCC}^2)	0.0034	0.0107		
Brothers in Law (σ_{cBB}^2)	0.0098	0.0056	0.0109	0.0040
Assortative mating (δ)	0.1667	0.6279		
B) Decompositions				
Sibling correlation MZ (ρ_{MZ})	0.4286	0.0577	0.4285	0.0577
Share pre-birth	0.3778	0.2807	0.4398	0.2520
Sibling correlation DZ (ρ_{DZ})	0.2766	0.0601	0.2643	0.0433
Share pre-birth	0.4514	0.3274	0.3940	0.2312
IGE (β)	0.1829	0.0328	0.1776	0.0274
Share pre-birth	0.6483	0.4633	0.5661	0.3248

Table 3: Parameter estimates and decompositions for years of education

	(1) Full model		(2) Restricted model	
	Coeff.	S.E.	Coeff.	S.E.
	A) Variance components			
Unique Environment (σ_e^2)	0.0198	0.0032	0.0198	0.0032
Additive Genetic (σ_a^2)	0.0082	0.0038	0.0091	0.0052
Shared Environment				
MZ Twins (σ_{cMZ}^2)	0.0344	0.0046	0.0335	0.0055
DZ Twins (σ_{cDZ}^2)	0.0332	0.0046	0.0357	0.0035
Fathers and Sons (σ_{cI}^2)	0.0076	0.0042	0.0101	0.0027
Uncles and Nephews (σ_{cUN}^2)	0.0056	0.0046	0.0075	0.0012
Cousins (σ_{cCC}^2)	0.00004	0.0062		
Brothers in Law (σ_{cBB}^2)	0.0147	0.0034	0.0165	0.0014
Assortative mating (δ)	0.3623	0.3476		
	B) Decompositions			
Sibling correlation MZ (ρ_{MZ})	0.6825	0.0525	0.6825	0.0525
Share pre-birth	0.1927	0.0892	0.2141	0.1201
Sibling correlation DZ (ρ_{DZ})	0.6577	0.0448	0.6227	0.0379
Share pre-birth	0.1759	0.1023	0.1133	0.0651
IGE (β)	0.2387	0.0156	0.2302	0.0160
Share pre-birth	0.4814	0.2810	0.3101	0.1774

Table 4: Parameter estimates and decompositions in the canonical twin model

	(1) Permanent income		(2) Years of education	
	Coeff.	S.E.	Coeff.	S.E.*
A) Variance components				
Unique Environment (σ_e^2)	0.0544	0.0131	0.0379	0.0003
Additive Genetic (σ_a^2)	0.0331	0.0139	0.0047	0.0007
Shared Environment (σ_c^2)	0.0043	0.0090	0.0381	0.0005
B) Decompositions				
Sibling correlation MZ (ρ_{MZ})	0.4080	0.0586	0.5295	0.0033
Share pre-birth	0.8848	0.2485	0.1105	0.0163
Sibling correlation DZ (ρ_{DZ})	0.2275	0.0449	0.5002	0.0033
Share pre-birth	0.7934	0.0399	0.0584	0.0090

*unadjusted

Table 5: Parameter estimates and decompositions with environmental shifts of genetic component

	(1) Permanent income		(2) Years of education	
	Coeff.	S.E.	Coeff.	S.E.
	A) Variance components			
Unique Environment (σ_e^2)	0.0498	0.0085	0.0130	0.0031
Additive Genetic (σ_a^2) - All men	0.0162	0.0109	0.0095	0.0049
Additive Genetic (σ_a^2) - All women	0.0246	0.0129	0.0062	0.0047
Additive Genetic (σ_a^2) - Mixed	0.0301	0.0125	0.0158	0.0048
Shared Environment				
MZ Twins (σ_{cMZ}^2)	0.0271	0.0124	0.0331	0.0053
DZ Twins (σ_{cDZ}^2)	0.0148	0.0070	0.0356	0.0033
Fathers and Sons (σ_{cI}^2)	0.0043	0.0060	0.0089	0.0025
Uncles and Nephews (σ_{cUN}^2)	0.0056	0.0027	0.0067	0.0011
Cousins (σ_{cCC}^2)				
Brothers in Law (σ_{cBB}^2)	0.0122	0.0042	0.0165	0.0014
	B) Decompositions			
Sibling correlation MZ (ρ_{MZ})	0.4686	0.0645	0.7657	0.0573
Share pre-birth	0.3743	0.2537	0.2231	0.1133
Sibling correlation DZ (ρ_{DZ})	0.2864	0.0506	0.6935	0.0452
Share pre-birth	0.3529	0.2442	0.1181	0.0614
IGE (β) – All men	0.1471	0.0313	0.2389	0.0197
Share pre-birth	0.6532	0.4542	0.3477	0.1792
IGE (β) – All women	0.1790	0.0314	0.2229	0.0211
Share pre-birth	0.7406	0.3560	0.2584	0.1963
IGE (β) – Mixed	0.1970	0.0270	0.2648	0.0191
Share pre-birth	0.7776	0.3059	0.4703	0.1440