

Clark, Gregory; Hørlyk Kristensen, Martin

Working Paper

The mechanisms of intergenerational status transmission in modern Scandinavia

EHES Working Paper, No. 299

Provided in Cooperation with:

European Historical Economics Society (EHES)

Suggested Citation: Clark, Gregory; Hørlyk Kristensen, Martin (2026) : The mechanisms of intergenerational status transmission in modern Scandinavia, EHES Working Paper, No. 299, European Historical Economics Society (EHES), s.l.

This Version is available at:

<https://hdl.handle.net/10419/339328>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



European
Historical
Economics
Society

EHES Working Paper | No. 299 | March 2026

The Mechanisms of Intergenerational Status Transmission in Modern Scandinavia

Gregory Clark¹

University of Southern Denmark, Danish Institute for Advanced Study, and LSE

Martin Hørlyk Kristensen

University of Southern Denmark

The Mechanisms of Intergenerational Status Transmission in Modern Scandinavia

Gregory Clark¹

University of Southern Denmark, Danish Institute for Advanced Study, and LSE

Martin Hørlyk Kristensen

University of Southern Denmark

Abstract

There is continued debate, with important social consequences, about how social status is transmitted between parents and children. In particular, how much does genetics as opposed to social causation matter? In an ingenious recent article, Collado et al. (2023), using an extended family lineage from Sweden and the outcome of years of education, claim that outcome correlations across relatives are inconsistent with genetic transmission. In contrast, Clark (2023) claims that a simple 3-parameter additive genetic model of social outcomes predicts well correlations as remote as 4th cousins in a stable way for England 1600-2022. In this paper we show that the genetic model of Clark (2023) fits the correlation in educational attainment for extended sets of relatives in modern Denmark and Sweden as well as the 20-parameter social causation model of Collado et al. Further, the parameter values in the genetic model for Denmark and Sweden are similar to those found for a variety of social outcomes in both pre-industrial and modern England.

JEL Codes: J62, J12, D31, I24

Keywords: intergenerational mobility, social mobility, assortative mating, genetic transmission

¹ Correspondance: Gregory Clark (gclark@sam.sdu.dk)

We thank Dolores Collado, Ignacio Ortuno-Ortín, and Jan Stuhler for their generous help in detailing how they implemented their additive genetic transmission model.

Notice

The material presented in the EHES Working Paper Series is property of the author(s) and should be quoted as such. The views expressed in this Paper are those of the author(s) and do not necessarily represent the views of the EHES or its members.

1. Introduction

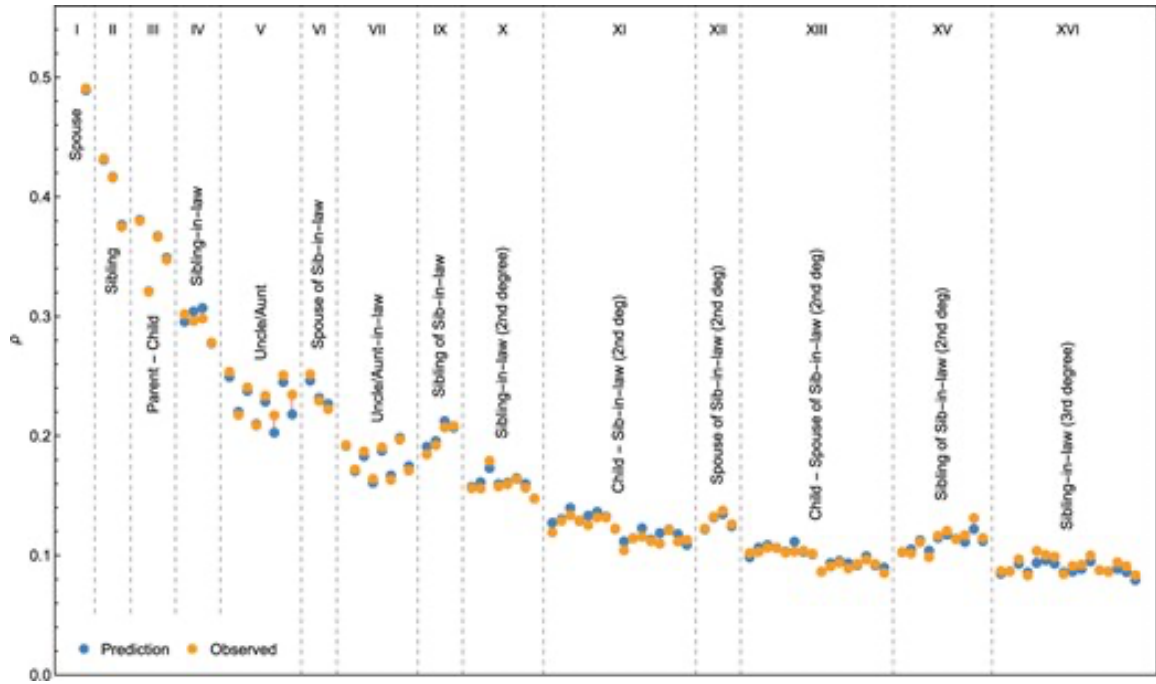
In an ingenious recent article, Collado et al. (2023), set out to model the mechanisms of social persistence and mobility in modern Sweden. To this end, they calculate from register data the correlations in years of education and income of an extended set of 141 types of relatives. These relatives are identified by linking children to each of their parents, parents to grandparents, parents to siblings, siblings to marital partners, and so on through in-laws across as many as five sets of parents. They then fit to 105 of these correlations a 20-parameter social transmission model of status determination which has a combination of latent and observed status, as well as marital matching on both latent and observed status. Their model also allows for different effects by gender.

They use only 105 of the 141 available correlations in fitting the model because for both cousins, and for in-laws related across more than three marriages, the fit of the model is poor. Collado et al. are able to fit well the patterns of correlation with the 105 kinship types. This is shown in Figure 1 where the observed and predicted correlations are displayed. That 20-parameter model, however, while showing strong assortment in parenting, and strong persistence across generations, does not offer much insight into why the explicit and latent variables have the relative strengths observed. But by implication social mobility rates are low, even in a welfare state such as Sweden.

Collado et al. (2023) further claimed to rule out the possibility that genetic transmission can be the major determinant of social status in modern Sweden. They show that an additive genetic model will not fit the observed correlation data. Their additive genetic model is a 5-parameter restricted specification of their 20-parameter model, with either parental assortment on the phenotype, or parental assortment on the genotype. In both cases the fit is poor, with the model showing either too little, or too much, persistence with implied genetic distance. Figure 2 shows how the best fit in each case accords poorly with the actual correlations.

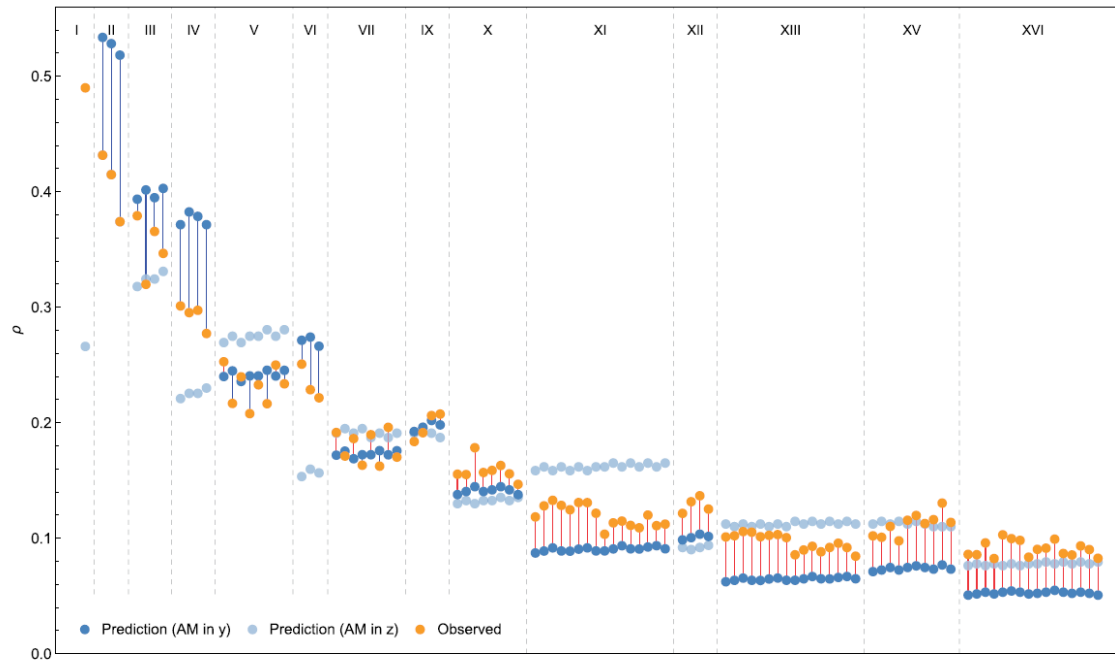
However, as we will show, in specifying the additive genetic model Collado et al. make a crucial assumption, which is that outcomes such as years of education measure without error educational attainment. This, as we will see for the model we estimate where assortment in parenting is on the phenotype, imposes constraints on how high genetic assortment can be. That assumption, however, explains the poor fit of this phenotype-matching genetic model. As can be seen in Figure 2, for the realistic case of parental matching on the phenotype (blue markers), the predicted Collado et al. correlations decline with distance in the family tree too rapidly, as a result of this constraint.

Figure 1: Fit of Collado et al. 20 parameter model.



Source: Collado et al., 2023, Figure 2.

Figure 2: Fit of Standard Genetic Model for Education, 105 correlation pairs



Source: Collado et al. (2023), Figure 6, panel (a).

The genetic correlation in marriage then has to be less than the observed phenotype correlation. With only a noisy measure available to researchers of educational attainment, which is just years of education, then people matching in marriage or parenthood, can be actually matching more closely in the underlying trait of educational attainment than is observed with the poorly measured phenotype. If that matching assumption is relaxed, we show below, then an additive genetic model, in which people do match on a phenotype, will fit the data well.

We further show that our simple additive genetic transmission model describes the data in both societies as well as the social transmission model of Collado et al. (2023), with its 20 free parameters. Also the genetic model for Scandinavia has close to the same parameters as that for England 1600-2022 described in Clark (2023). Additive genetics is still a viable candidate to explain social outcomes, even in Scandinavian societies with extensive state welfare provision.

2. The Mechanisms of Status Determination

To interpret the extensive pattern of familial correlations observed in the Scandinavian register data, we require a framework that links the observed outcomes to the underlying forces generating social status. A natural starting point is a simple model in which individuals possess a latent socioeconomic trait, “social ability”, that is only imperfectly reflected in measurable outcomes such as years of schooling or income and is transmitted across generations through both inheritance and assortative mating.

To make these ideas operational, we draw on a Fisherian framework in which social ability is an additive genetic trait: each individual inherits a random half of each parent’s genetic contribution, and correlation between relatives arises solely from transmission and assortative mating.² This additive structure implies that correlations between relatives decline in a mechanically predictable way with genealogical distance, and that marital sorting amplifies similarity throughout the extended family. Because social ability is imperfectly measured in outcomes such as education or income, all observable correlations are attenuated versions of the underlying genetic relationships. Combining these elements, the model predicts expressions for the expected correlations between any two relatives, whether siblings, cousins, in-laws, or more complex multi-marriage links, allowing us to compare its predictions directly to the rich kinship data found in the Scandinavian registers.

² This is a model of assortative mating initially developed by Fisher (1918), and elaborated by Crow and Felsenstein (1968), Nagylaki (1978), and Falconer (1981).

Table 1 shows the expected correlations between different types of relatives in an extended family tree. The model underlying Table 1 is one where people have an underlying unitary social ability, which is observed imperfectly in such social outcomes as years of schooling. Parents match in this social status with a correlation of r . There is a corresponding correlation between parents in the genetics determining this social ability of m , where m will be related to r through:

$$m = h^2 r$$

and h^2 is the heritability of underlying social ability. m will thus be less than r . Since we only observe this underlying social ability through imperfect measures such as years of schooling, all the correlations in Table 1 will be attenuated by a factor θ . With this specification the genetic correlation between parents can exceed the observed phenotype correlation of measures such as years of education, which is θr .³ Because gender does not matter in this simpler additive genetic model, the 141 distinct correlations in Collado et al. (2023) are described by just 21 formulas in Table 1. The degree of genetic assortment in marriage, m , is the crucial determinant of social mobility rates across multiple generations.⁴

Since our sources on the expected correlations of relatives do not cover affine relatives, we infer what these expected correlations will be ourselves. Our expected affine correlations are mostly close to those derived by Collado et al. (2023). But in some cases we deviate from their expected correlations, though typically in ways modest enough to not significantly affect the fit of the model to the correlation data. As a robustness check we redo all the estimates using the Collado et al. (2023) set of predicted correlations. This makes very little difference to the key estimated parameters. These results are detailed in Appendix B.

³ As noted Collado et al. (2023) assume that the genotype correlation between parents has to be less than the observed phenotype correlation in spouses. The phenotype correlation observed in parents in the Swedish data, for example, is 0.49 for years of education. They assume that the genotype correlation for the genetics that promotes education then has to be significantly below 0.49.

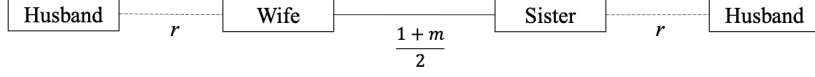
⁴ The indicators of underlying social ability can be different for men compared to women. In this case the correlations between relatives would also potentially be influenced by the gender mix (Espín-Sánchez et al., 2023). In the robustness section, we allow for gendered differences with same-sex indicators in our regressions.

Table 1: Sample of Implied Familial Correlations with Genetic Transmission

Relationship	Definition	Implied Correlation
Spouses	MF	θr
Siblings	M-M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)$
Parent-Child	M $\leftarrow$ F, etc.	$\theta h^2 \left(\frac{1+r}{2} \right)$
Sibling in-law	MF-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)$
Aunt/Uncle	F $\leftarrow$ F-M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)^2$
Spouse of Sib-in-law	MF-FM, etc.	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)$
Aunt/Uncle-in-law	MF $\leftarrow$ M-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)^2$
Cousins	M $\leftarrow$ F-F $\rightarrow$ M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)^3$
Sibling of sib-in-law	F-MF-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)^2$
Sibling-in-law (2 nd degree)	MF-MF-M, etc.	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^2$
Child of sib-in-law (2 nd degree)	M $\leftarrow$ MF-MF-M	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^3$
Spouse of sib-in-law (2 nd degree)	MF-MF-MF	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^2$
Sibling of sib-in-law (2 nd degree)	F-MF-MF-M	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^3$
Child of spouse of sib-in-law (2 nd degree)	M $\leftarrow$ MF-MF-MF	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^3$
Spouse of sib-in-law (2 nd degree) twice rem.	M $\leftarrow$ MF-MF-MF $\rightarrow$ M	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^4$
Sibling-in-law (3 rd degree)	MF-MF-MF-M	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^3$
Spouse of sibling-in-law (3 rd degree)	MF-MF-MF-MF	$\theta h^2 r^4 \left(\frac{1+m}{2} \right)^3$
Sibling-in-law (4 th degree)	MF-MF-MF-MF-M	$\theta h^2 r^4 \left(\frac{1+m}{2} \right)^4$
Spouse of sibling-in-law (4 th degree)	MF-MF-MF-MF-MF	$\theta h^2 r^5 \left(\frac{1+m}{2} \right)^4$
Sibling-in-law (5 th degree)	MF-MF-MF-MF-MF-M	$\theta h^2 r^5 \left(\frac{1+m}{2} \right)^5$
Spouse of sibling-in-law (5 th degree)	MF-MF-MF-MF-MF-MF	$\theta h^2 r^6 \left(\frac{1+m}{2} \right)^5$

Notes: M/F indicates gender, MF or FM denotes spouses, - indicates a sibling link, and $\rightarrow$ indicates a parent-child link. Thus M-MF-MF-M indicates the pair brother-brother-wife-brother-wife-brother.

Figure 3: Example of an expected kinship correlation



Notes: The figure illustrates the kinship relationship MF-FM. The resulting theoretical correlation is $h^2 r^2 \left(\frac{1+m}{2}\right)$.

Figure 3 gives an example of how the theoretical correlation in Table 1 is computed for a specific kinship relationship, husband of sister-in-law (denoted MF-FM). The correlation in an outcome between individuals - assumed to be r for spouses and $\left(\frac{1+m}{2}\right)$ between siblings - is linked to the heritability of the trait (h^2) through a multiplicative relationship, thus yielding a theoretical correlation of $h^2 r^2 \left(\frac{1+m}{2}\right)$ in this example. When we observe this correlation through partial measures such as years of education, however, it will be attenuated by a factor of θ , so it will be $\theta h^2 r^2 \left(\frac{1+m}{2}\right)$.

We exploit the fact that almost all implied correlations in Table 1 can be written in the multiplicative form:⁵

$$\rho_{kln} = \theta h^2 r^k \left(\frac{1+m}{2}\right)^n \quad (1)$$

where k indicates the number of marital links and n the number of genetic steps between two relatives. Taking logs yields a linear estimating equation that maps directly into the observable characteristics of each kinship pair:

$$\ln(\rho_{kln}) = \alpha + \beta_1 k + \beta_2 n + \beta_3 dphen \quad (2)$$

With $dphen = 1$ only for parent-child relations, which include the additional term $\ln\left(\frac{1+r}{1+m}\right)$. In practice, we estimate this model by regressing the logarithm of the empirical correlation on the variables that encode these theoretical components: the number of affine links k , the genetic distance n , and $dphen$. Ordinary Least Squares provides a baseline

⁵ The two exceptions are the parent correlation θr , and the parent-child correlation, which can be written as: $\theta h^2 \left(\frac{1+r}{2}\right) = \theta h^2 \left(\frac{1+r}{1+m}\right) \left(\frac{1+m}{2}\right)$

specification, while Weighted Least Squares, using weights proportional to the structure of heteroskedasticity, gives greater precision in relationships with many observed pairs. We also estimate a fully non-linear specification, imposing the exact Fisherian form specified in Equation (1).

With the log-linear specification the underlying structural components of the model map cleanly into the estimated coefficients. The coefficient on the number of marital links, exponentiated, gives the strength of assortative mating. The coefficient on genetic distance yields the rate at which correlations decays across generations, from which the implied heritability of the latent trait follows directly. The intercept identifies the degree of attenuation from measuring social ability through imperfect proxies such as education or income. This specification allows us to generate predicted correlations for all 141 kinship types.

We exclude spouse correlations from the estimation because they do not involve genetic transmission and therefore do not load on genetic distance, but rather solely on assortative mating and measurement attenuation. Instead, we estimate the model using the remaining 140 relationships and subsequently impute the implied spouse correlations based on the estimated assortment parameter r and attenuation parameter θ .

3. Data

To evaluate the predictive power of our genetic model, we use population-wide correlation matrices that capture how outcomes vary across an extensive set of familial relationships. For Sweden, we rely on the correlations reported by Collado et al. (2023), who document 141 distinct kinship types constructed from the Swedish multigeneration register (see Table 1 for an overview).

For Denmark, we construct an analogous dataset using population registers. This was achieved through a chain-linking process: first, we identified the parents of individuals in our sample, then use shared parent identifiers to identify siblings, and continue merging data to establish connections with more distantly related individuals. This chain-linking procedure yields the same rich set of 141 kinship categories for Denmark, allowing us to compute correlations in educational attainment and income and to compare the Scandinavian correlations in a consistent way.

Table 2 provides an overview of the variables of interest we obtain from the Danish registers. To ensure that educational attainment is fully realized at the time of observation, we restrict the sample to individuals born before 1985. We further limit the sample to individuals with valid identifiers for both parents, which is necessary to construct extended-family links and compute correlations across relatives. Income is measured as average annual income over the ages 35-45 and is expressed in 2015 Danish kroner (DKK).

For each individual, we compute the years of education corresponding to their highest level of attainment. We standardize this variable over time by de-meaning by cohort and gender. For our measure of income, we consider pre-tax income denoted in 2015 DKK and calculate the average income ages 35-45 for each individual. We standardize the values by cohort and gender and censor the income averages at the 1st and 99th percentile to reduce the influence of outliers.

Appendix Table A1 summarizes the number of observed pairs for each kinship type in Sweden and Denmark, along with the correlation in educational attainment. There is a substantial difference between the number of brothers and sisters in the Danish registers. This stems from differing fractions of men and women who have both parents with a personal identifier.⁶ Females are more likely to have a missing value in one of the parent identifiers. We attempt to correct for these differences by also utilizing a sample consisting only of Danish born individuals born post-1960.

Figure 4 shows the Swedish and Danish education correlations by relationship type. There is a striking similarity in the correlations in Denmark and those reported by Collado et al. for Sweden.⁷ The correlation between the Danish and Swedish correlations in Table A1 is indeed 0.98. There are, however, significant differences in the correlations of Danish and Swedish spouses, siblings, and in parent-children relationships. But for the other relationships there are only negligible differences.

⁶ As noted by Mortensen et al. (2006), this difference likely stems from the fact that male children tended to leave the parental household later than female children. When the Danish Civil Registration System was established in 1968, parent-child links were initially constructed based on co-residence at the same address and were deleted once children moved away from home. As a result, parental identifiers, and hence sibling links, were more likely to be retained for males in older cohorts. From 1978 onward, parental links were redefined based on permanent legal relationships, substantially improving the completeness of sibling identification in later birth cohorts.

⁷ Figure A1 illustrates similar patterns of income correlations among relatives in both Denmark and Sweden.

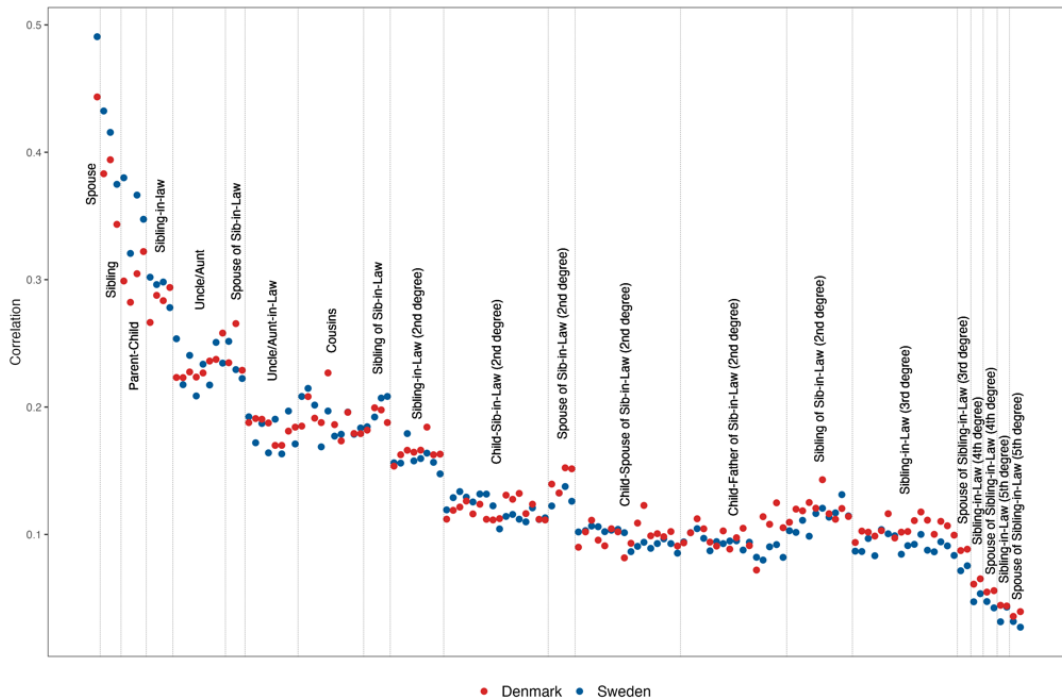
Table 2: Summary statistics for Denmark

Variable	Persons	Mean	Standard Deviation
Years of education	2,331,919	14.3	2.84
Avg. Income	2,246,563	327,778	239,911
Log(income)	2,246,438	11.7	3.15
Birth year	2,364,249	1967	10.6
Female	2,364,249	0.48	0.50
Born Outside Denmark	2,361,148	0.03	0.18

Source: Danish Population Register

Notes: The included group are individuals from the Danish register with both parents' IDs reported. Income and Wealth are in DKK.

Figure 4: Comparison of Swedish and Danish correlations in Education



Notes: This figure presents the sample correlations in educational attainment among Danish and Swedish relatives.

4. Results

This section evaluates a parsimonious additive genetic model using population-wide correlations observed across extended family networks in Denmark and Sweden. Rather than focusing on a single intergenerational relationship, the analysis exploits the full correlation structure spanning siblings, cousins, in-laws, and more distant relatives.

A central feature of our approach is its simplicity. While Collado et al. (2023) fit a richly parameterized social transmission model to the Swedish data, we confront the same extended-family correlations with a three-parameter additive framework capturing just assortative mating, genetic transmission, and measurement attenuation.

We begin with years of education in Sweden and Denmark, then extend the analysis to income. Across outcomes and countries, we estimate the model using both log-linear and fully non-linear specifications and compare fitted correlations directly to the observed data. This approach allows us to assess whether a highly parsimonious genetic model can replicate the detailed empirical patterns documented in Scandinavian registers.

4.1 Years of education

Table 3 reports the fit of the three-parameter additive genetic model to correlations in years of education across extended families in Sweden and Denmark. Across both countries, the model delivers an excellent fit to the data: the log-linear specification explains roughly 95% of the variation in the 140 observed kinship correlations.

The implied parameters are remarkably stable across estimation methods and countries. In particular, the estimated phenotypic correlation in marriage is high ($r \approx 0.80$ - 0.83), indicating strong assortative mating on education, while the implied genetic assortment between spouses is substantial ($m \approx 0.52$ - 0.59). These estimates imply a long-run intergenerational persistence of educational status of roughly 0.76 - 0.80 , even in the context of modern Scandinavian welfare states.

Estimating the model in levels using non-linear least squares (columns 3 and 6) yields very similar parameter values, confirming that the results are not an artifact of the log-linear approximation. Taken together, the estimates in Table 3 show that a parsimonious genetic

transmission model with only three free parameters closely reproduces the full structure of educational correlations observed across extended families in both Sweden and Denmark.

While the baseline genetic model abstracts from gender-specific differences in the transmission of status, it nonetheless captures the overall structure of correlations across relatives remarkably well. Allowing for gender heterogeneity by including categorical indicators for same-gender female pairs, same-gender male pairs, and mixed-gender pairs yields only a small improvement in fit. As shown in Appendix Table A2, introducing these additional indicators leaves the core parameters of the model largely unchanged, while slightly increasing explanatory power, suggesting that gender-specific effects only operate as a second-order refinement to the baseline model.

Figure 5 compares the fit of the simple genetic model, estimated by WLS (column (2) of Table 3), to both the observed correlations in educational attainment and the predictions from the 20-parameter social transmission model of Collado et al. (2023). Overall, the figure shows that the parsimonious genetic model tracks the empirical correlations closely across the full range of kinship types. While the social transmission model fits sibling and sibling-in-law correlations more tightly, the genetic model nonetheless provides a comparably good fit across the remaining kinship categories.

Importantly, despite using far fewer parameters, the genetic model captures the overall level and variation in correlations across relatives in a manner that is very similar to the predictions of the much richer social transmission framework.

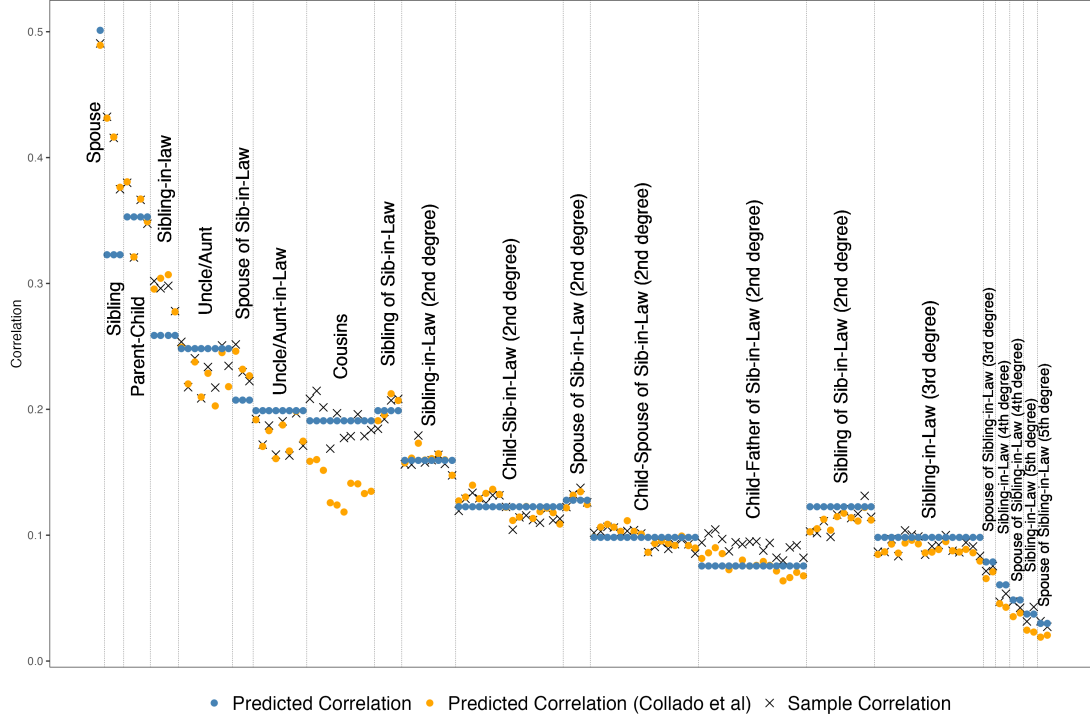
Table 3: Regression fit for genetic model (years of education)

	Sweden			Denmark		
	OLS	WLS	NLS	OLS	WLS	NLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.888*** (0.035)	-0.868*** (0.034)	0.431*** (0.009)	-0.952*** (0.034)	-0.913*** (0.032)	0.381*** (0.008)
Genetic distance	-0.254*** (0.015)	-0.263*** (0.015)		-0.240*** (0.015)	-0.256*** (0.014)	
Phenotypic indicator	0.100 (0.064)	0.089 (0.056)		-0.006 (0.062)	-0.029 (0.049)	
Marital assortment	-0.224*** (0.010)	-0.221*** (0.010)		-0.189*** (0.010)	-0.186*** (0.010)	
r			0.802*** (0.009)			0.823*** (0.009)
m			0.520*** (0.016)			0.587*** (0.018)
r	0.799	0.801	0.802	0.828	0.830	0.823
b	0.776	0.769	0.760	0.787	0.774	0.794
m	0.552	0.538	0.520	0.573	0.549	0.587
h^2	0.691	0.671	0.648	0.692	0.661	0.713
θ	0.596	0.625	0.666	0.557	0.607	0.534
AIC	-192.492	-192.513	-687.827	-202.683	-205.102	-688.101
BIC	-177.784	-177.805	-676.060	-187.975	-190.393	-676.334
R ²	0.947	0.949		0.936	0.942	
Adjusted R ²	0.945	0.948		0.935	0.941	
Observations	140	140	140	140	140	140

Notes: The dependent variable in columns (1)-(3) is the correlation in educational attainment across Swedish relatives as reported in Collado et al. (2023, Table 3). Columns (4)-(6) use the empirical correlation in years of education among Danish relatives.

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Figure 5: Educational Attainment in Sweden - Fit of the Collado et al. (2023) 20-parameter model versus a 3-parameter genetic model



Notes: This figure plots educational attainment correlations in Sweden by relationship, along with the predicted correlation from Collado et al. (2023) versus our genetic model.

For the Danish sample, a potential concern is that the baseline correlations are based on substantially fewer brother pairs than sister pairs, as illustrated in Figure A2. To assess whether this imbalance affects our estimates, we re-estimate the model using a restricted sample consisting only of Danish-born individuals born after 1960, for whom parental identifiers are most complete. This restriction substantially reduces the gender imbalance in observed sibling pairs (see Table A3 for the corresponding sample correlations). The resulting estimates, reported in Table A4 are very similar to those obtained in the baseline specification, continuing to imply strong assortative mating ($r \approx 0.88$) and a high degree of long-run intergenerational mobility ($b \approx 0.75$).

Figure 6 displays how well the additive genetic transmission model fits the data in both Denmark and Sweden. It plots the observed correlations of years of education for each

relative class against the implied genotype share for each kinship type.⁸ In both Denmark and Sweden, the relationship is strikingly close to linear: correlations rise smoothly with the implied share genotype, with little evidence of systematic deviations across kinship categories. This pattern is exactly what the additive genetic model predicts. Accordingly, Figure 6 shows that once the fitted parameters from Table 3 are used to map genealogical distance and marital links into a single measure of genetic relatedness, the rich set of extended-family correlations collapses onto a single, approximately one-dimensional relationship.

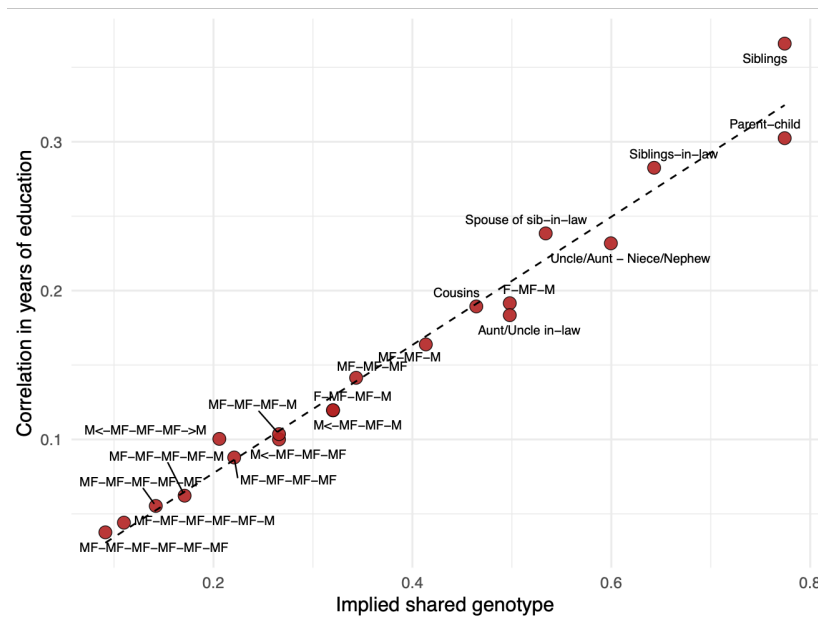
Note that in Figure 6 for both Denmark and Sweden the one relationship which deviates from the predicted pattern in the genetic model is that of siblings, where the actual correlation exceeds the model prediction. This suggests that there may some family effect in terms of years of education. Yet this effect is not transmitted to the children of these siblings, cousins. They correlate in proportion to their predicted shared genotype.

Also of note is that the parameters of the model fitting both Denmark and Sweden are the same as those for both contemporary England, and also England in the nineteenth century. Table 4 shows the estimated parameters r , marital phenotype correlation, m , marital genotype correlation, and b , underlying intergenerational correlation, for each of these cases. The estimated parameters are remarkably similar across these different societies and time periods.

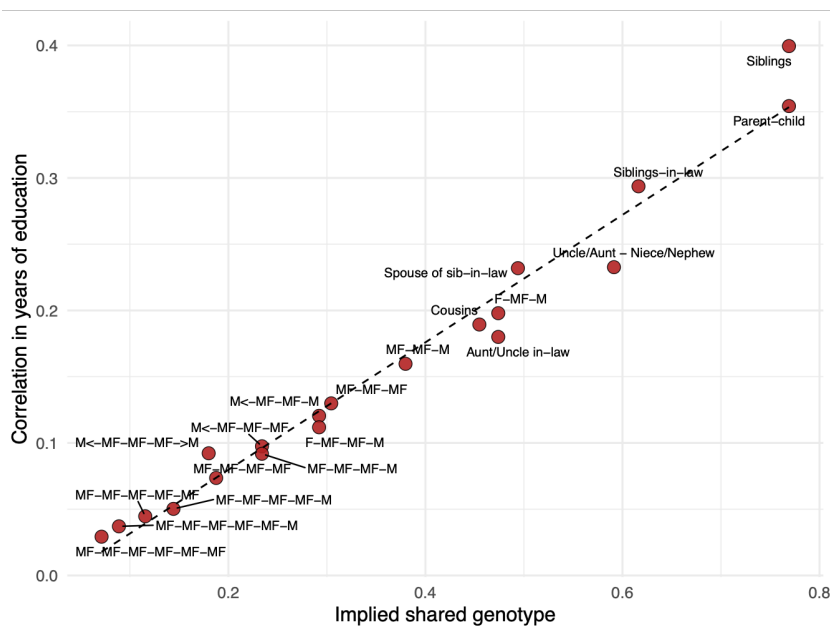
⁸ To construct the figure, we recover the key parameters from the WLS estimates in Table 3 and use them to compute, for each kinship pairing, an implied share-genotype index given by $(\hat{b})^n \cdot (\hat{r})^k$, where $\hat{b} = \exp(\beta_2)$ and $\hat{r} = \exp(\beta_1)$. We collapse the data to one observation per kinship category by taking pair-count-weighted means of the observed correlation (excluding spouses), and plot these along with the best linear fit.

Figure 6: Correlations in educational attainment by implied share genotype

(a) Denmark



(b) Sweden



Notes: This figure plots the observed correlations in years of education among different kinship types against the implied shared genotype predicted by the three-parameter genetic transmission model. Each point corresponds to a distinct kinship category, aggregated from the full set of 141 genealogical types used in the Collado et al. (2023) extended-family design.

Table 4: Comparative Genetic Model Parameters, England versus Modern Nordic

Population	Period	Marital <i>r</i>	Marital <i>m</i>	Parent-Child <i>b</i>
England	19 th c	0.80	0.60	0.80
England	Modern	0.80	0.59	0.79
Denmark	Modern	0.83	0.55	0.77
Sweden	Modern	0.80	0.54	0.77

Notes: The outcomes used in 19th century England were literacy at marriage, occupational status (males), attainment of higher education (males). For contemporary England the outcomes were log house value, the index of multiple deprivation of the address, and company directorships.

Sources: Table 3, Clark (2023), Table A6, Clark and Cummins (2025).

In a recent paper using register data from Norway, Eftedal et al. (2025) report that though genetic distance predominates in determining the correlations in academic test performance at age 12, across an extended lineage of relatives, there is evidence of significant social causation also. Evidence of such social causation is found in that adoptive siblings, cousins, and even second cousins show significant correlations in academic performance. Thus while the correlation in academic scores of full siblings is 0.48, that of adopted siblings is 0.15. This implies that genetics has more than 3 times the influence as does environment, but that environment still plays a role.

It is important, however, that the Eftedal et al. (2025) results relate to the mathematics, reading, and English scores of 12-year-olds. The literature on IQ inheritance, which is a good predictor of school performance, has established clearly that for younger children there are strong influences of the custodial parents on child IQ, even for unrelated adoptive children. But by ages 16-18 family environment has little effect. Scarr and Weinberg (1977) and (1978), for example, found an IQ correlation of 0.33 between unrelated adoptees at age 7, but for another sample, found no correlation between unrelated adoptees aged 16-22. It is also important that the mechanisms by which adoptions were arranged are unknown. If adoption is to some degree selective in the Norwegian data, with children with more educated biological parents matched to adoptive parents with more education, then some of the correlation of adopted children in educational outcomes will come from this parental correlation.

In the outcome used in our study for Denmark and Sweden, completed years of education as an adult, there is more limited evidence from adoptions on whether the outcome shows social transmission from the custodial parents. For Swedish children adopted by Swedish parents, predicting child years of education, the average coefficient on the father's years of education across 3 studies was 0.07, and for the mother's 0.06 (Holmlund, Lindahl and Plug, 2008, Table 1 and Table 5). But for foreign children adopted by Swedish parents the father and mother coefficients were both insignificantly different from 0 (0.014, s.e. 0.009, versus 0.015, s.e. 0.009) (Holmlund, Lindahl and Plug, 2008, Table 5). For foreign children, many of whom in this study coming from Korea, there is often little information on the child that would influence an adoption agency's placement. Thus in cases where adoption placements were most likely random, we do not observe significant social impact on years of education.

Thus the Eftedal et al. (2025) data does not fundamentally challenge the observation for both Denmark and Sweden that a simple Fisherian additive genetic model for years of education, without social impacts on outcomes, describes the correlation patterns well.

4.2 Income

Can the same parsimonious genetic model account for correlations in income across extended family networks? Income is a noisier proxy for latent socioeconomic status than years of education. Thus the correlation between siblings on years of education in Denmark is 0.37, but the correlation in income is only 0.15, only 40% as much. Nevertheless, if the additive genetic framework captures a fundamental component of status transmission, it should also describe the broad structure of income correlations across relatives. Income and years of education, for example, are positively correlated.

For this outcome, we estimate the model for Sweden and Denmark using income averaged over ages 35-45, and report the resulting parameter estimates in Table 5. As before, we estimate the model using OLS, WLS, and a fully non-linear specification to examine whether the implied parameters are consistent across specifications.

The fact that the income correlations are substantially weaker than those for education - across kinship types income correlations are only about one third as large as education correlations - makes the estimates more unstable across methods and less good at predicting outcomes. This is shown in Table 5 where many of the estimates have implausible values for

the error attenuation, θ , which even exceed 1. Also the implied genetic correlation between parents varies from 0.14 to 0.36.

But the most plausible estimate in terms of error attenuation, the non-linear fit for Denmark, does produce an estimate that shows a good fit between the implied genetic correlation of relatives and their income correlation. This is shown in Figure 7, where the R^2 is 0.94.

Note, however, that even in this case the implied genetic correlation between parents is lower than in the case of education, 0.36 versus 0.55, and consequently the income correlation drops more quickly as we move to more distant relatives than in the case of education. The implied phenotypic assortment of parents on the characteristics that determine income is as strong as for education (0.8-0.9). But because the heritability of income is estimated to be much lower, the implied genetic assortment of parents is much lower than for education.

We explore in the appendix whether these different results for income reflect differences in the determination of income by gender, but find such gender effects to be of modest significance. Although income is standardized by gender and cohort, this transformation equalizes levels rather than the extent to which income reflects latent socioeconomic status. We therefore re-estimate the model, restricting attention to male kinship pairings. The resulting estimates, presented in Table A5, continue to exhibit slightly high values of the attenuation parameter θ , which increase slightly relative to the baseline specifications. At the same time, the implied intergenerational persistence and degree of assortative mating remain broadly similar.

Figure A3 presents a comparison of these income estimates with those of Collado et al. (2023) together with the observed correlations. The figure shows that while both models capture the broad decline in income correlations with genealogical distance, the deviations from the empirical correlations are larger and more dispersed, reflecting the weaker signal in income. Relative to the rich social transmission model of Collado et al. (2023), the genetic model delivers a comparable fit despite relying on far fewer parameters. All in all, the figure underscores that income correlations provide a more demanding test of the model but remain broadly consistent with the same underlying transmission structure identified for education.

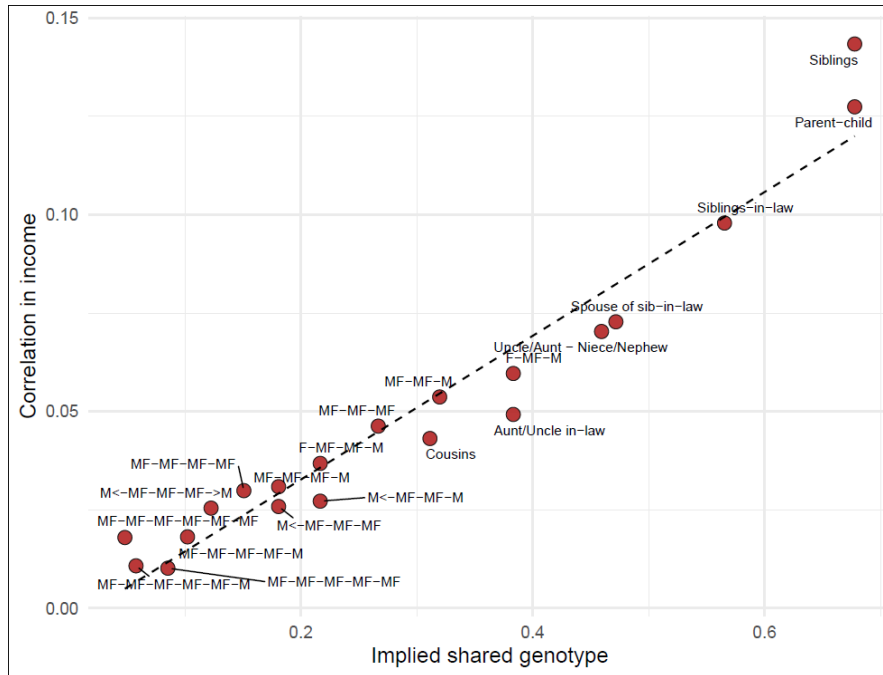
Table 5: Regression fit for genetic model (income)

	Sweden			Denmark		
	OLS	WLS	NLS	OLS	WLS	NLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-1.714*** (0.116)	-1.660*** (0.112)	0.183*** (0.012)	-1.640*** (0.183)	-1.642*** (0.043)	0.158*** (0.006)
Genetic distance	-0.551*** (0.037)	-0.566*** (0.036)		-0.545*** (0.081)	-0.479*** (0.043)	
Phenotypic indicator	0.201 (0.203)	0.163 (0.166)		0.104 (0.335)	0.040 (0.161)	
Marital assortment	-0.148*** (0.049)	-0.161*** (0.049)		-0.109** (0.054)	-0.203*** (0.040)	
r			0.796*** (0.048)			0.834*** (0.022)
m			0.202*** (0.034)			0.356*** (0.032)
r	0.862	0.851	0.796	0.897	0.816	0.834
b	0.577	0.568	0.601	0.580	0.620	0.678
m	0.153	0.136	0.202	0.160	0.239	0.356
h^2	0.178	0.160	0.254	0.178	0.293	0.426
θ	1.014	1.190	0.720	1.088	0.661	0.372
AIC	125.045	124.034	-681.238	270.128	276.823	-790.307
BIC	139.753	138.742	-669.471	284.836	291.531	-778.540
R ²	0.667	0.700		0.500	0.667	
Adjusted R ²	0.660	0.694		0.489	0.660	
Observations	140	140	140	140	140	140

Notes: The dependent variable in columns (1)-(3) is the correlation in income across Swedish relatives as reported in Collado et al. (2023, Table G.2). Columns (4)-(6) use the empirical correlation in income among Danish relatives.

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Figure 7: Correlations in income by implied share genotype, Denmark



5 Conclusions

This paper examines social mobility in modern Scandinavia using population-wide register data and an extended family approach that goes well beyond parent-child comparisons. Exploiting correlations across 141 types of relatives, we document three central findings.

First, long-run persistence in social status is substantially stronger than conventional mobility estimates suggest and is no lower than that observed in pre-industrial England or Sweden. This strong persistence is shown in an alternative way, using just the relative correlations in educational attainment between different types of relatives, in Clark and Kristensen (2026).

Second, marital assortment on socioeconomic status is much stronger than typically assumed, helping to explain the strong persistence across generations. This strong assortment is also shown in an alternative way in Clark and Kristensen (2026).

Third, the observed pattern of correlations observed across relatives in education, income, and test scores can be well described by a parsimonious additive genetic transmission model with only three parameters. This simple framework fits the Danish and Swedish data as well as the 20-parameter social transmission model of Collado et al. (2023). It also yields parameter estimates in the case of education that closely mirror those of England for a range of social outcomes over the period 1600-2022 (Clark, 2023).

These findings have important implications for how social mobility is understood in egalitarian societies. Despite extensive public investment in education and income redistribution, family background remains an important determinant of outcomes. This does not necessarily imply that welfare policies are ineffective, but it does suggest that they have not fundamentally altered the underlying structure of status transmission.

Overall, our findings imply that the Swedish evidence presented by Collado et al. (2023) does not, as claimed, rule out genetic transmission as a determinant of social status. For years of education, a parsimonious genetic model with only three free parameters reproduces the pattern of correlations across 141 types of relative with an explanatory power exceeding 93 percent. Because our analysis is based on observational data, we do not claim to identify genetic causation. What we do show, however, is that if social mechanisms are the dominant force shaping educational attainment, they operate in a manner that closely mirrors the structure implied by additive genetic transmission.

Moreover, the similarity of the estimated parameters in Denmark, Sweden, and historical England suggests that extensive public investment in education and redistribution has not eliminated the role of family background in shaping long-run outcomes. This conclusion aligns with recent evidence emphasizing the persistent importance of family background even in high-capacity welfare states (Heckman and Landersø, 2022).

References

- Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton: Princeton University Press.
- Clark, G. (2023). The inheritance of social status: England, 1600 to 2022. *Proceedings of the National Academy of Sciences*, 120(27), e2300926120.
- Clark, G. and Kristensen, M. H. (2026). The Myth of Nordic Mobility. *Comparative Economic Studies*, 68(1): 6.
- Collado, M. D., Ortuño-Ortín, I., and Stuhler, J. (2023). Estimating Intergenerational and Assortative Processes in Extended Family Data. *The Review of Economic Studies*, 90(3):1195–1227.
- Crow, J. F. and J. Felsenstein. 1968. The effect of assortative mating on the genetic composition of a population, *Eugenics Quarterly*, 15:2, 85-97. DOI: 10.1080/19485565.1968.9987760
- Eftedal, N. H., Eilertsen, E. M., Sunde, H. F., Kleppestø, T. H., Ystrom, E., & Czajkowski, N. O. (2025). Beyond additive genetic effects: Explaining family resemblance in school performance across millions of pairs of Norwegian relatives. *Proceedings of the National Academy of Sciences*, 122(25), e2419627122.
- Espín-Sánchez, J. A., Ferrie, J. P., & Vickers, C. (2023). Women and the Econometrics of Family Trees (No. w31598). National Bureau of Economic Research.
- Falconer, D. S. 1981 *Introduction to Quantitative Genetics*. 2nd ed. London: Longman.
- Fisher, R. A. 1918 The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 52: 399-433 (1918).
- Heckman, James, and Rasmus Landersø. 2022. Lessons for Americans from Denmark about inequality and social mobility. *Labour Economics*, 77.
- Holmlund, Helena, Mikael Lindahl, and Erik Plug. 2008. “The Causal Effect of Parent’s Schooling on Children’s Schooling: A Comparison of Estimation Methods.” IZA DP No. 3630.
- Mortensen, P. B., Gøtzsche, H., Bøcker Pedersen, C., & Østrup Møller, J. (2006). The Danish Civil Registration System: a cohort of eight million persons. *Dan Med Bull*, 53(4): 441-9.
- Nagylaki, T. 1978. The correlation between relatives with assortative mating. *Annals of Human Genetics*, 42: 131.
- Scarr, Sandra and R A Weinberg. 1977. Intellectual similarities within families of both adopted and biological children. *Intelligence*, 1: 170-191.
- Scarr, Sandra and R A Weinberg. 1978. “The influence of “family background” on intellectual attainment” *American Sociological Review*, 43: 674-92.

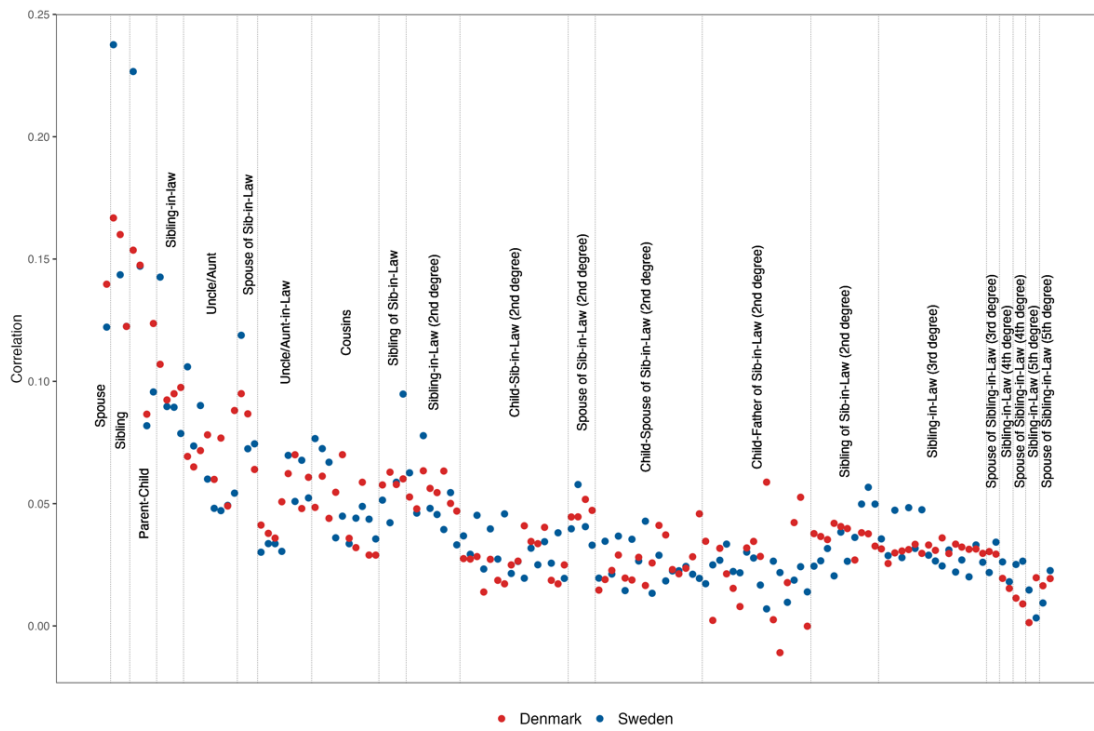
Appendix A: Additional Tables and Figures

Table A1: Sample correlations (educational attainment)

Kinship	Sweden		Denmark	
	Pairs	ρ	Pairs	ρ
Spouses	413,062	0.491	1,869,156	0.443
Brothers	387,028	0.432	820,000	0.383
Sisters	431,698	0.416	686,666	0.394
Brother–Sister	800,127	0.375	1,486,452	0.343
Father–Son	396,304	0.380	1,053,353	0.299
Father–Daughter	376,255	0.321	981,636	0.282
Mother–Son	422,374	0.366	1,144,876	0.305
Mother–Daughter	400,337	0.347	1,061,231	0.322
Brothers in-law (MF-M)	602,262	0.302	670,359	0.266
Brother–Sister in-law	578,269	0.296	662,531	0.288
Brother–Sister in-law	650,127	0.298	615,443	0.284
Sisters in-law (FM-F)	596,540	0.278	597,312	0.294
Nephew–Uncle (BF)	280,067	0.254	97,102	0.223
Niece–Uncle (BF)	266,289	0.218	93,901	0.223
Nephew–Uncle (BM)	312,019	0.241	118,072	0.228
Niece–Uncle (BM)	295,580	0.209	113,566	0.224
Nephew–Aunt (SF)	285,618	0.234	82,155	0.227
Niece–Aunt (SF)	270,325	0.217	78,870	0.236
Nephew–Aunt (SM)	333,141	0.251	101,847	0.237
Niece–Aunt (SM)	316,625	0.234	98,051	0.258
Spouse of Sib-in-law	252,232	0.252	577,509	0.235
Spouse of Sib-in-law	226,795	0.229	568,957	0.266
Spouse of Sib-in-law	464,081	0.222	1,393,198	0.229
Nephew–Aunt in-law	231,767	0.192	109,422	0.188
Niece–Aunt in-law (BF)	221,287	0.172	87,924	0.191
Nephew–Aunt in-law	254,534	0.187	108,942	0.190
Niece–Aunt in-law (BM)	241,873	0.164	104,450	0.188
Nephew–Uncle in-law	227,403	0.190	80,236	0.170
Niece–Uncle in-law (SF)	215,068	0.163	77,126	0.170
Nephew–Uncle in-law	264,524	0.197	100,685	0.181
Niece–Uncle in-law (SM)	251,782	0.171	97,124	0.184
Male Cousins (BF)	70,137	0.208	38,750	0.185
Male Cousins (SM)	82,049	0.215	47,954	0.208
Male Cousins (SF)	156,747	0.202	82,380	0.191
Female Cousins (BF)	63,032	0.169	36,319	0.188
Female Cousins (SM)	73,649	0.197	43,828	0.227
Female Cousins (SF)	140,522	0.177	75,815	0.186
Male–Female Cousins	144,100	0.179	75,817	0.173
Male–Female Cousins	170,577	0.196	91,394	0.196
Male–Female Cousins	148,691	0.179	79,232	0.179
Male–Female Cousins	148,631	0.184	79,234	0.179
F-MF-M	461,883	0.185	389,061	0.182
F-MF-F	500,448	0.192	359,243	0.199
F-FM-M	481,006	0.207	400,901	0.198
M-MF-M	447,263	0.208	433,364	0.188
MF-MF-M	362,409	0.156	730,781	0.154
MF-MF-F	393,579	0.156	662,739	0.163
MF-FM-M	375,442	0.179	379,502	0.166
MF-FM-F	391,389	0.158	339,766	0.165
FM-MF-M	353,470	0.160	376,090	0.166
FM-MF-F	378,720	0.164	346,887	0.184
FM-FM-M	341,316	0.157	363,310	0.163
FM-FM-F	351,350	0.148	324,546	0.163
M←MF-MF-M	202,632	0.119	95,207	0.112
M←MF-MF-F	219,007	0.129	84,992	0.119
M←MF-FM-M	192,819	0.134	77,318	0.121
M←MF-FM-F	199,811	0.129	67,794	0.126
M←FM-MF-M	183,670	0.125	73,490	0.116
M←FM-MF-F	196,631	0.132	65,362	0.124
M←FM-FM-M	160,857	0.132	56,012	0.112
M←FM-FM-F	164,528	0.122	48,624	0.111
F←MF-MF-M	192,818	0.104	91,455	0.112
F←MF-MF-F	208,008	0.114	81,672	0.131
F←MF-FM-M	183,929	0.116	75,296	0.128
F←MF-FM-F	191,177	0.112	65,667	0.132
F←FM-MF-M	175,507	0.110	70,456	0.116
F←FM-MF-F	187,178	0.121	65,362	0.124

F←FM-FM-M	151,606	0.112	56,012	0.112
F←FM-FM-F	155,658	0.113	48,624	0.111
MF-MF-MF	278,938	0.122	626,918	0.140
MF-MF-FM	310,160	0.132	632,376	0.133
MF-FM-MF	299,602	0.138	331,458	0.152
FM-MF-MF	273,809	0.126	318,196	0.152
M←MF-MF-MF	160,726	0.102	84,256	0.090
M←MF-MF-FM	174,261	0.103	84,120	0.102
M←MF-FM-MF	158,401	0.107	70,267	0.111
M←MF-FM-FM	160,105	0.106	66,404	0.096
M←FM-MF-MF	147,949	0.102	65,442	0.091
M←FM-MF-FM	156,876	0.103	65,068	0.104
M←FM-FM-MF	133,588	0.104	51,380	0.102
M←FM-FM-FM	131,756	0.101	47,608	0.082
F←MF-MF-MF	152,751	0.087	80,876	0.093
F←MF-MF-FM	165,828	0.091	80,546	0.109
F←MF-FM-MF	151,100	0.094	68,505	0.123
F←MF-FM-FM	153,065	0.089	64,231	0.099
F←FM-MF-MF	140,585	0.093	62,475	0.101
F←FM-MF-FM	150,162	0.097	62,784	0.098
F←FM-FM-MF	126,129	0.093	49,062	0.102
F←FM-FM-FM	124,968	0.085	45,578	0.091
M←MF-MF-MF→M	84,025	0.094	19,576	0.093
M←MF-MF-FM→M	100,261	0.101	26,158	0.101
M←MF-FM-MF→M	93,237	0.105	21,930	0.112
M←FM-MF-MF→M	80,486	0.097	16,329	0.104
M←MF-MF-MF→F	79,690	0.087	18,768	0.094
M←MF-MF-FM→F	95,733	0.094	25,532	0.091
M←MF-FM-MF→F	89,364	0.093	21,077	0.103
M←MF-FM-FM→F	95,020	0.095	25,188	0.088
M←FM-MF-MF→F	76,514	0.095	15,961	0.097
M←FM-MF-FM→F	89,054	0.088	21,069	0.105
M←FM-FM-MF→F	77,332	0.094	15,788	0.091
M←FM-FM-FM→F	80,067	0.082	18,530	0.072
F←MF-MF-MF→F	76,344	0.080	18,442	0.114
F←MF-MF-FM→F	91,080	0.090	24,350	0.108
F←MF-FM-MF→F	84,736	0.092	20,860	0.125
F←FM-MF-MF→F	72,410	0.082	15,136	0.105
F-MF-MF-M	288,374	0.103	224,800	0.110
F-MF-MF-F	312,703	0.102	207,713	0.120
F-MF-FM-M	311,795	0.111	240,118	0.119
F-MF-FM-F	162,928	0.099	215,407	0.125
F-FM-MF-M	308,163	0.116	242,060	0.121
F-FM-MF-F	166,250	0.121	224,556	0.143
F-FM-FM-M	304,684	0.114	232,562	0.116
M-MF-MF-M	278,416	0.117	253,294	0.112
M-MF-FM-M	149,478	0.131	270,020	0.120
M-FM-MF-M	143,733	0.115	264,726	0.114
MF-MF-MF-M	230,313	0.087	438,905	0.094
MF-MF-MF-F	251,223	0.087	404,376	0.103
MF-MF-FM-M	248,811	0.097	440,396	0.102
MF-MF-FM-F	259,925	0.083	394,051	0.099
MF-FM-MF-M	245,814	0.104	232,180	0.103
MF-FM-MF-F	265,220	0.101	214,945	0.116
MF-FM-FM-M	241,998	0.099	222,513	0.097
MF-FM-FM-F	248,449	0.084	198,708	0.102
FM-MF-MF-M	224,873	0.091	222,460	0.102
FM-MF-MF-F	246,186	0.092	204,172	0.111
FM-MF-FM-M	237,791	0.100	236,327	0.118
FM-MF-FM-F	247,495	0.088	209,865	0.111
FM-FM-MF-M	223,661	0.086	224,907	0.100
FM-FM-MF-F	240,328	0.094	206,010	0.110
FM-FM-FM-M	213,155	0.091	212,238	0.107
FM-FM-FM-F	220,553	0.084	188,585	0.099
MF-MF-MF-MF	176,790	0.071	372,515	0.087
MF-MF-MF-FM	199,041	0.075	389,472	0.088
MF-MF-MF-MF-M	153,057	0.047	267,366	0.061
FM-FM-FM-FM-F	144,976	0.054	112,892	0.065
MF-MF-MF-MF-MF	117,473	0.047	225,449	0.055
MF-MF-MF-MF-FM	135,096	0.042	235,976	0.056
MF-MF-MF-MF-MF-M	106,844	0.031	165,068	0.044
FM-FM-FM-FM-FM-F	100,871	0.043	151,531	0.044
MF-MF-MF-MF-MF-MF	82,523	0.032	138,887	0.036
MF-MF-MF-MF-MF-FM	96,840	0.027	146,291	0.039

Figure A1: Income correlations among relatives



Notes: This figure presents the sample correlations in income among Danish and Swedish relatives.

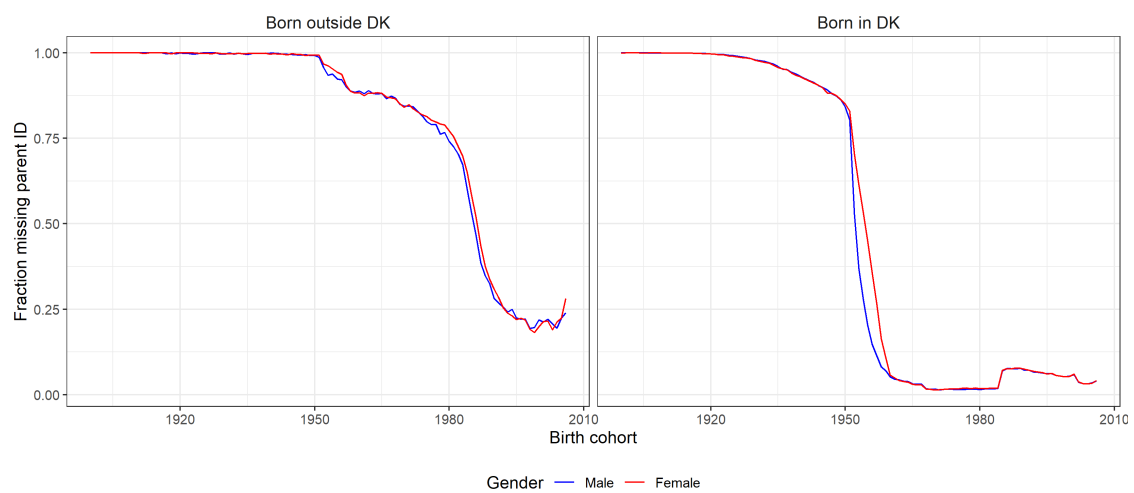
Table A2: Extension of the baseline model (years of education)

	Sweden			Denmark		
	OLS	WLS	NLS	OLS	WLS	NLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.894*** (0.036)	-0.862*** (0.034)	0.430*** (0.009)	-0.972*** (0.034)	-0.929*** (0.032)	0.379*** (0.008)
Genetic distance	-0.254*** (0.015)	-0.270*** (0.015)		-0.241*** (0.014)	-0.259*** (0.014)	
Phenotypic indicator	0.098 (0.063)	0.079 (0.049)		-0.004 (0.060)	-0.030 (0.046)	
Marital assortment	-0.226*** (0.010)	-0.221*** (0.010)		-0.187*** (0.010)	-0.184*** (0.009)	
Same gender, female	-0.021 (0.024)	-0.017 (0.024)	-0.001 (0.004)	0.074*** (0.023)	0.077*** (0.022)	0.012*** (0.004)
Same gender, male	0.054** (0.024)	0.064*** (0.023)	0.012*** (0.004)	-0.001 (0.023)	0.003 (0.022)	0.001 (0.004)
r			0.798*** (0.009)			0.820*** (0.009)
m			0.514*** (0.017)			0.580*** (0.018)
r	0.798	0.802	0.798	0.829	0.832	0.820
h	0.776	0.763	0.757	0.786	0.772	0.790
m	0.552	0.527	0.514	0.571	0.544	0.580
h^2	0.691	0.657	0.644	0.689	0.653	0.707
θ	1.118	1.868	0.668	0.550	0.604	0.535
AIC	-197.285	-198.091	-695.119	-210.903	-214.107	-694.329
BIC	-176.694	-177.500	-677.469	-190.311	-193.515	-676.680
R ²	0.950	0.955		0.941	0.948	
Adjusted R ²	0.948	0.953		0.939	0.946	
Observations	140	140	140	140	140	140

Notes: The dependent variable in columns (1)-(3) is the correlation in educational attainment across Swedish relatives as reported in Collado et al. (2023, Table 3). Columns (4)-(6) use the empirical correlation in years of education among Danish relatives.

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Figure A2: Missing parental ID



Notes: This figure depicts the fraction of individuals for whom at least one parent is missing a personal identifier by gender and place of birth in the Danish register data.

Table A3: Sample correlations (educational attainment, DK-born & born post-1960)

Kinship	ρ	No. pairs	Kinship	ρ	No. pairs
Spouses	0.440	466,900	MF-FM-MF	0.164	90,020
Brothers	0.383	604,680	FM-MF-MF	0.151	87,443
Sisters	0.394	549,196	M<-MF-MF-MF	0.104	15,527
Brother-Sister	0.345	1,136,972	M<-MF-MF-FM	0.103	15,329
Father-Son	0.307	972,640	M<-MF-FM-MF	0.124	15,494
Father-Daughter	0.289	930,380	M<-MF-FM-FM	0.112	13,597
Mother-Son	0.322	1,015,068	M<-FM-MF-MF	0.091	16,264
Mother-Daughter	0.338	969,995	M<-FM-MF-FM	0.107	16,727
Brothers-in-law (MF-M)	0.273	295,075	M<-FM-FM-MF	0.087	14,553
Brother-Sister-in-law (FM-M)	0.287	298,676	M<-FM-FM-FM	0.089	12,863
Brother-Sister-in-law (MF-F)	0.293	285,505	F<-MF-MF-MF	0.110	14,799
Sisters-in-law (FM-F)	0.287	279,033	F<-MF-MF-FM	0.121	14,958
Nephew-Uncle (BF)	0.225	37,270	F<-MF-FM-MF	0.126	14,866
Niece-Uncle (BF)	0.223	35,526	F<-MF-FM-FM	0.092	12,727
Nephew-Uncle (BM)	0.223	64,170	F<-FM-MF-MF	0.091	15,623
Niece-Uncle (BM)	0.215	61,338	F<-FM-MF-FM	0.112	15,970
Nephew-Aunt (SF)	0.204	34,646	F<-FM-FM-MF	0.106	13,567
Niece-Aunt (SF)	0.237	32,967	F<-FM-FM-FM	0.110	12,089
Nephew-Aunt (SM)	0.224	62,360	M<-MF-MF-MF->M	0.104	3,162
Niece-Aunt (SM)	0.248	59,050	M<-MF-MF-FM->M	0.072	3,261
Spouse of Sib-in-law (MF-FM)	0.249	191,927	M<-MF-FM-MF->M	0.139	3,698
Spouse of Sib-in-law (FM-MF)	0.257	202,674	M<-FM-MF-MF->M	0.078	3,346
Spouse of Sib-in-law (MF-MF)	0.234	583,851	M<-MF-MF-MF->F	0.115	3,021
Nephew-Aunt in-Law (BF)	0.178	39,904	M<-MF-MF-FM->F	0.082	3,008
Niece-Aunt in-Law (BF)	0.192	28,569	M<-MF-FM-MF->F	0.098	3,421
Nephew-Aunt in-Law (BM)	0.174	49,797	M<-MF-FM-FM->F	0.118	3,092
Niece-Aunt in-Law (BM)	0.177	47,475	M<-FM-MF-MF->F	0.083	3,109
Nephew-Uncle in-Law (SF)	0.158	23,986	M<-FM-MF-FM->F	0.078	3,527
Niece-Uncle in-Law (SF)	0.172	22,874	M<-FM-FM-MF->F	0.039	3,151
Nephew-Uncle in-Law (SM)	0.175	44,166	M<-FM-FM-FM->F	0.071	2,989
Niece-Uncle in-Law (SM)	0.178	41,957	F<-MF-MF-MF->F	0.141	2,859
Male cousins (BF)	0.142	7,756	F<-MF-MF-FM->F	0.065	3,018
Male cousins (SM)	0.170	21,485	F<-MF-FM-MF->F	0.140	3,374
Male cousins (SF)	0.173	23,543	F<-FM-MF-MF->F	0.085	3,131
Female cousins (BF)	0.168	6,872	F-MF-MF-M	0.122	78,941
Female cousins (SM)	0.195	19,028	F-MF-MF-F	0.126	75,269
Female cousins (SF)	0.174	21,575	F-MF-FM-M	0.128	82,755
Male-Female cousins (BF)	0.176	14,793	F-MF-FM-F	0.126	76,877
Male-Female cousins (SM)	0.176	40,138	F-FM-MF-M	0.126	87,896
Male-Female cousins (BM)	0.159	22,330	F-FM-MF-F	0.152	86,337
Male-Female cousins (SF)	0.159	22,330	F-FM-FM-M	0.120	80,891
F-MF-M	0.186	183,773	M-MF-MF-M	0.116	84,411
F-MF-F	0.202	176,823	M-MF-FM-M	0.135	90,085
F-FM-M	0.200	190,651	M-FM-MF-M	0.120	91,292
M-MF-M	0.191	196,248	MF-MF-MF-M	0.098	113,239
MF-MF-M	0.158	246,247	MF-MF-MF-F	0.103	107,989
MF-MF-F	0.168	232,265	MF-MF-FM-M	0.108	109,139
MF-FM-M	0.176	128,631	MF-MF-FM-F	0.104	102,642
MF-FM-F	0.170	119,389	MF-FM-MF-M	0.110	60,615
FM-MF-M	0.169	134,095	MF-FM-MF-F	0.138	60,091
FM-MF-F	0.189	129,738	MF-FM-FM-M	0.097	55,820
FM-FM-M	0.159	123,216	MF-FM-FM-F	0.107	52,147
FM-FM-F	0.169	114,938	FM-MF-MF-M	0.106	59,709
M<-MF-MF-M	0.106	22,748	FM-MF-MF-F	0.113	57,313
M<-MF-MF-F	0.120	21,316	FM-MF-FM-M	0.127	62,765
M<-MF-FM-M	0.122	20,703	FM-MF-FM-F	0.120	57,614
M<-MF-FM-F	0.113	19,531	FM-FM-MF-M	0.107	59,391
M<-FM-MF-M	0.110	23,438	FM-FM-MF-F	0.114	57,147
M<-FM-MF-F	0.122	22,840	FM-FM-FM-M	0.112	53,376
M<-FM-FM-M	0.117	19,480	FM-FM-FM-F	0.105	50,220
M<-FM-FM-F	0.099	18,316	MF-MF-MF-MF	0.092	73,990
F<-MF-MF-M	0.120	21,651	MF-MF-MF-FM	0.088	75,574
F<-MF-MF-F	0.133	20,584	MF-MF-MF-MF-M	0.077	52,450
F<-MF-FM-M	0.131	19,774	FM-FM-FM-FM-F	0.088	22,803
F<-MF-FM-F	0.119	18,395	MF-MF-MF-MF-MF	0.067	33,733
F<-FM-MF-M	0.116	22,413	MF-MF-MF-MF-FM	0.072	34,692
F<-FM-MF-F	0.122	22,840	MF-MF-MF-MF-MF-M	0.073	24,846
F<-FM-FM-M	0.117	19,480	MF-MF-MF-MF-MF-F	0.057	23,075
F<-FM-FM-F	0.099	18,316	MF-MF-MF-MF-MF-MF	0.069	16,083
MF-MF-MF	0.142	165,760	MF-MF-MF-MF-MF-FM	0.070	16,015
MF-MF-FM	0.139	159,536			

Table A4: Regression fit (educational attainment, Danish-born & post-1960)

	OLS	WLS	NLS
	(1)	(2)	(3)
Intercept	-0.994*** (0.048)	-0.823*** (0.034)	0.385*** (0.009)
Genetic distance	-0.274*** (0.021)	-0.325*** (0.017)	
Phenotypic indicator	0.108 (0.088)	-0.011 (0.033)	
Marital assortment	-0.119*** (0.014)	-0.130*** (0.013)	
r			0.861*** (0.011)
m			0.535*** (0.019)
r	0.887	0.878	0.861
$(1 + m)/2$	0.761	0.722	0.767
m	0.521	0.445	0.535
h^2	0.587	0.506	0.621
θ	0.631	0.867	0.619
AIC	-103.323	-122.177	-658.738
BIC	-88.615	-107.469	-646.972
R ²	0.850	0.928	
Adjusted R ²	0.847	0.927	
Observations	140	140	140

Notes: The dependent variable is the empirical correlation in educational attainment among Danish-born kin born post-1960.

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table A5: Regression fit for 3-parameter genetic model (income, males)

	Sweden			Denmark		
	OLS	WLS	NLS	OLS	WLS	NLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-1.323*** (0.125)	-1.167*** (0.108)	0.299*** (0.018)	-1.450*** (0.297)	-1.472*** (0.083)	0.193*** (0.011)
Genetic distance	-0.554*** (0.039)	-0.595*** (0.036)		-0.579*** (0.132)	-0.572*** (0.037)	
Phenotypic indicator	0.393 (0.245)	0.277** (0.106)		0.156 (0.585)	0.170 (0.377)	
Marital assortment	-0.134** (0.053)	-0.180*** (0.052)		-0.151* (0.086)	-0.156*** (0.036)	
r			0.756*** (0.045)			0.810*** (0.032)
m			0.153*** (0.018)			0.289*** (0.011)
r	0.874	0.835	0.756	0.860	0.856	0.810
b	0.575	0.552	0.577	0.561	0.565	0.644
m	0.149	0.103	0.153	0.121	0.129	0.289
h^2	0.171	0.124	0.203	0.141	0.151	0.357
θ	1.557	2.516	1.475	1.664	1.520	0.541
AIC	0.905	-1.246	-207.173	69.707	37.428	-230.702
BIC	9.223	7.072	-200.518	78.024	45.745	-224.048
R ²	0.877	0.942		0.674	0.881	
Adjusted R ²	0.866	0.937		0.646	0.870	
Observations	39	39	39	39	39	39

Notes: The dependent variable in columns (1)-(3) is the correlation in income across Swedish relatives as reported in Collado et al. (2023, Table G.2). Columns (4)-(6) use the empirical correlation in income among Danish relatives. The sample is restricted to male kinship pairings only.

Significance levels: ***p < 0.01; **p < 0.05; *p < 0.1

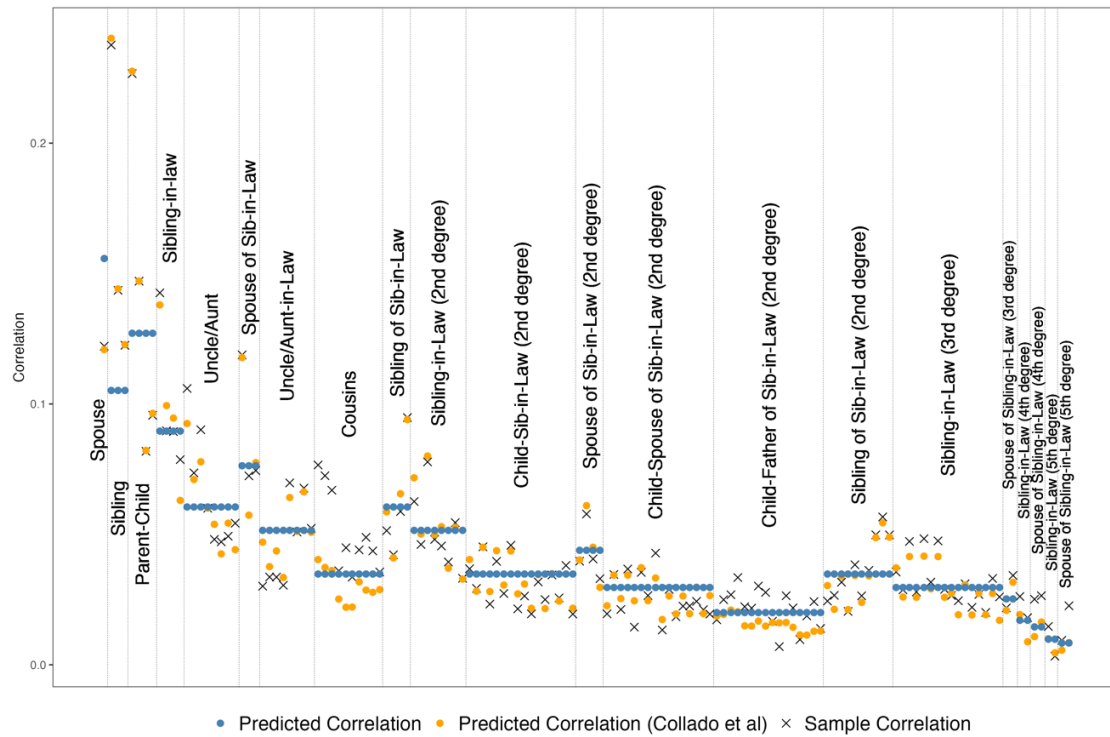
Table A6: Extension of the baseline model (income)

	Sweden			Denmark		
	OLS	WLS	NLS	OLS	WLS	NLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-1.721*** (0.085)	-1.597*** (0.073)	0.182*** (0.010)	-1.644*** (0.192)	-1.685*** (0.031)	0.157*** (0.006)
Genetic distance	-0.567*** (0.026)	-0.610*** (0.023)		-0.545*** (0.082)	-0.483*** (0.031)	
Phenotypic indicator	0.163 (0.143)	0.143** (0.070)		0.104 (0.338)	0.039 (0.116)	
Marital assortment	-0.178*** (0.034)	-0.218*** (0.033)		-0.109** (0.055)	-0.201*** (0.028)	
Same gender, female	-0.218*** (0.054)	-0.196*** (0.055)	-0.008** (0.003)	0.007 (0.129)	0.048*** (0.015)	0.005* (0.003)
Same gender, male	0.471*** (0.053)	0.530*** (0.046)	0.023*** (0.003)	0.007 (0.127)	0.138*** (0.015)	0.004 (0.003)
r			0.774*** (0.043)			0.825*** (0.023)
m			0.162*** (0.033)			0.338*** (0.034)
r	0.837	0.804	0.774	0.897	0.818	0.825
b	0.567	0.544	0.581	0.580	0.617	0.669
m	0.134	0.087	0.162	0.160	0.234	0.338
h^2	0.160	0.108	0.210	0.178	0.286	0.410
θ	1.014	1.190	0.866	1.088	0.661	0.383
AIC	28.170	23.744	-736.071	274.123	185.431	-794.974
BIC	48.762	44.336	-718.422	294.714	206.022	-777.324
R ²	0.838	0.928		0.500	0.818	
Adjusted R ²	0.832	0.925		0.482	0.811	
Observations	140	140	140	140	140	140

Notes: The dependent variable in columns (1)-(3) is the correlation in income across Swedish relatives as reported in Collado et al. (2023, Table 3). Columns (4)-(6) use the empirical correlation in income among Danish relatives.

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Figure A3: Sweden, Income - Fit of the Collado et al. 20 parameter model versus a 3-parameter genetic model (income)



Appendix B: Alternative coding schemes

This appendix documents the robustness of our results to alternative coding schemes for genetic distance and marital links. The baseline specification treats all marital links symmetrically. We consider two alternatives: (i) a coding that distinguishes between internal and terminal marriage links, and (ii) the coding used by Collado et al. (2023), which differs from our baseline in one case.

Table B.1 reports the implied correlations under each coding scheme. The baseline specification corresponds to the expressions shown in Table 1 of the main text. The alternative specifications modify only the treatment of affine links, leaving the genetic distance component unchanged. For most of the kinship types the implied correlations are identical across schemes; differences arise mainly for higher-order chains involving multiple marriages.

A useful way to see how these coding choices enter the empirical exercise is to note that the implied correlation can be expressed in the multiplicative form:

$$\rho_{kln} = \theta h^2 r^k m^l \left(\frac{1+m}{2} \right)^n \quad (B.1)$$

Where n is the number of genetic steps in the chain and k and l summarize the number of marital links. The alternative coding schemes affect only how the affine component $r^k m^l$ by changing how marital links are counted, while leaving the genetic distance term $\left(\frac{1+m}{2} \right)$ unchanged. Taking logs yields a linear estimating equation that we re-estimate under each coding scheme:

$$\ln(\rho_{kln}) = \alpha + \beta_1 k + \beta_2 l + \beta_3 n + \beta_4 dphen \quad (B.2)$$

Where k counts terminal marriages (a marriage that attaches a spouse at the end of the chain), l counts internal marriages (a marriage that connects two genetic segments within the chain), and n is the number of genetic steps in the chain.

Table B.2 reports the estimated parameters obtained when re-estimating the log-linear specification under each coding scheme using WLS. Across Sweden and Denmark, the parameter estimates are relatively stable and the explanatory remains high. These results confirm that the strong fit of the additive genetic framework is not driven by the particular coding of marital links adopted in the baseline specification.

Table B.1: Alternative coding of genetic distance and marital links

Relationship	Definition	Baseline	Two-marriage coding	CIS coding
Spouses	MF	θr	same	same
Siblings	M-M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)$	same	same
Parent-Child	M $\leftarrow$ F, etc.	$\theta h^2 \left(\frac{1+r}{2} \right)$	same	$\theta h^2 \left(\frac{1+m}{2} \right)$
Sibling in-law	MF-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)$	same	same
Aunt/Uncle	F $\leftarrow$ F-M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)^2$	same	same
Spouse of Sib-in-law	MF-FM, etc.	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)$	same	same
Aunt/Uncle-in-law	MF $\leftarrow$ M-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)^2$	same	same
Cousins	M $\leftarrow$ F-F $\rightarrow$ M, etc.	$\theta h^2 \left(\frac{1+m}{2} \right)^3$	same	same
Sibling of sib-in-law	F-MF-M, etc.	$\theta h^2 r \left(\frac{1+m}{2} \right)^2$	$\theta h^2 m \left(\frac{1+m}{2} \right)^2$	$\theta h^2 m \left(\frac{1+m}{2} \right)^2$
Sibling-in-law (2 nd degree)	MF-MF-M, etc.	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^2$	$\theta h^2 r m \left(\frac{1+m}{2} \right)^2$	$\theta h^2 r m \left(\frac{1+m}{2} \right)^2$
Child of sib-in-law (2 nd degree)	M $\leftarrow$ MF-MF-M	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 m^2 \left(\frac{1+m}{2} \right)^3$	$\theta m^2 \left(\frac{1+m}{2} \right)^3$
Spouse of sib-in-law (2 nd degree)	MF-MF-MF	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^2$	$\theta h^2 r^2 m \left(\frac{1+m}{2} \right)^2$	$\theta h^2 r^2 m \left(\frac{1+m}{2} \right)^2$
Sibling of sib-in-law (2 nd degree)	F-MF-MF-M	$\theta h^2 r^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 m^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 m^2 \left(\frac{1+m}{2} \right)^3$
Child of spouse of sib-in-law (2 nd degree)	M $\leftarrow$ MF-MF-MF	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r m^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r^2 m \left(\frac{1+m}{2} \right)^3$
Spouse of sib-in-law (2 nd degree) twice rem.	M $\leftarrow$ MF-MF-MF $\rightarrow$ M	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 m^3 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 r^2 m \left(\frac{1+m}{2} \right)^4$
Sibling-in-law (3 rd degree)	MF-MF-MF-M	$\theta h^2 r^3 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r m^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r m^2 \left(\frac{1+m}{2} \right)^3$
Spouse of sibling-in-law (3 rd degree)	MF-MF-MF-MF	$\theta h^2 r^4 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r^2 m^2 \left(\frac{1+m}{2} \right)^3$	$\theta h^2 r^2 m^2 \left(\frac{1+m}{2} \right)^3$
Sibling-in-law (4 th degree)	MF-MF-MF-MF-M	$\theta h^2 r^4 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 r m^3 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 r m^3 \left(\frac{1+m}{2} \right)^4$
Spouse of sibling-in-law (4 th degree)	MF-MF-MF-MF-MF	$\theta h^2 r^5 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 r^2 m^3 \left(\frac{1+m}{2} \right)^4$	$\theta h^2 r^2 m^3 \left(\frac{1+m}{2} \right)^4$
Sibling-in-law (5 th degree)	MF-MF-MF-MF-MF-M	$\theta h^2 r^5 \left(\frac{1+m}{2} \right)^5$	$\theta h^2 r m^4 \left(\frac{1+m}{2} \right)^5$	$\theta h^2 r m^4 \left(\frac{1+m}{2} \right)^5$
Spouse of sibling-in-law (5 th degree)	MF-MF-MF-MF-MF-MF	$\theta h^2 r^6 \left(\frac{1+m}{2} \right)^5$	$\theta h^2 r^2 m^4 \left(\frac{1+m}{2} \right)^5$	$\theta h^2 r^2 m^4 \left(\frac{1+m}{2} \right)^5$

Notes: M/F indicates gender, MF or FM denotes spouses, - indicates a sibling link, and $\rightarrow$ indicates a parent-child link. Thus M-MF-MF-M indicates the pair brother-brother-wife-brother-wife-brother. Cells with implied correlation “same” corresponds to the implied correlation reported in the “Baseline” column.

Table B.2: Estimated parameters under alternative coding schemes (years of education)

(a) Denmark

	Baseline	Two-marriage	CIS
r	0.830	0.816	0.882
m	0.549	0.515	0.630
b	0.774	0.757	0.815
θ	0.607	0.664	0.504
R^2	0.942	0.943	0.956

(b) Sweden

	Baseline	Two-marriage	CIS
r	0.801	0.768	0.841
m	0.538	0.453	0.608
b	0.769	0.726	0.804
θ	0.671	0.796	0.537
R^2	0.949	0.953	0.959

Notes: The table reports parameter estimates from the log-linear specification described in Section 2, re-estimated under alternative coding schemes for marital links. The baseline specification treats all marital links symmetrically. The two-marriage coding distinguishes between internal and terminal marriage links, while the Collado coding adopts the correlation structure used in Collado et al. (2023). All regressions are estimated using WLS.

EHES Working Paper Series

Recent EHES Working Papers

- EHES 298** Automated historical census digitization using image augmentation and transformer-based methods
Leonardo Costa Ribeiro, Jonatan Andersson, William Skoglund, Jakob Molinder, Martin Önnersfors
- EHES 297** The Jamaica Life Expectancy Paradox and Hookworm Eradication
Elisabeth Preyer, Eric Strobl
- EHES 296** Historical social tables: advantages, methodology, and problems
Philipp Erfurth, María Gómez León, Giacomo Gabbuti, Branko Milanović
- EHES 295** Physical Memories of the Past and Support for the Far-Right: Evidence from Inter-War Denmark
Lasse Aaskoven, Christian Vedel
- EHES 294** The Early Evolution of Segregation and Neighborhood Disadvantage: Evidence from Stockholm During Industrialization, 1878–1950
Jakob Molinder, Martin Önnersfors
- EHES 293** „Vinum regnum, rex vinorum” Royal and Cameral Vineyard Possession in the Tokaj Estate in the 18th Century (Reforms, Viticulture, Wine Treatment, and Their Implications)
Attila Ulrich
- EHES 292** Unions and the Great Leveling: Evidence from Across the Atlantic
William Skoglund, Jakob Molinder,
- EHES 291** The Great Inner Divergence: TFP and Manufacturing Dualism in Industrializing Empires before WWI
Guillem Blasco-Piles
- EHES 290** Mitigating the impact of extreme temperatures: the role of public health interventions in Germany, 1888-1913
Daniel Gallardo-Albarrán

All papers may be downloaded free of charge from: <https://www.ehes.org/>

The European Historical Economics Society is concerned with advancing education in European economic history through study of European economies and economic history. The society is registered with the Charity Commissioners of England and Wales number: 1052680.