Ethnic Capital or Dynastic Human Capital? Evidence from Multigenerational Data*

Adrian Adermon† Mikael Lindahl‡ Mårten Palme§ Jonatan Riberth¶

April 9, 2026

Abstract

Intergenerational convergence across immigrant groups is often found to be slower than traditional parent–child models predict. A common explanation is that ethnic capital, measured as the average human capital of the origin group, independently shapes individual outcomes. This paper shows that such estimates can be biased when extended-family influences are omitted, as standard specifications may mechanically attribute family transmission to group effects. Using Swedish administrative registers with complete multigenerational family networks, we decompose persistence in educational outcomes into parental, extended-family, and country-of-origin components. Once extended-family human capital is accounted for, country-of-origin effects become negligible. Parental and extended family factors explain nearly all observed persistence, implying a reinterpretation of ethnic-capital models.

*We are grateful to Martin Nybom and Dinand Webbink as well as seminar audiences at the Department of Economics at Stockholm University, Uppsala University, University of Gothenburg, IFAU in Uppsala, VATT in Helsinki, the 2023 FEW summer meeting in Uppsala, the 2024 Uppsala Immigration Lab/Urban Lab workshop at Rånäs, the 2024 ESPE conference in Rotterdam, the 2024 Workshop on Wealth Inequality, Social Mobility, and Equality of Opportunity in Vienna, the 2025 Workshop on the Intergenerational Mobility of Immigrants in Princeton, University of Luxembourg, University of Duisburg-Essen, TU Dortmund; University of Potsdam; Workshop on Intergenerational Persistence and Inequality, EUI, Florence, and the BI Norwegian Business School in Oslo for comments and discussions. Adermon gratefully acknowledges financial support from the Swedish Research Council for Health, Working Life and Welfare.

†IFAU, Uppsala. adrian.adermon@ifau.uu.se.

‡Department of Economics, University of Gothenburg. mikael.lindahl@economics.gu.se.

§Department of Economics, Stockholm University. marten.palme@su.se.

¶IIES, Stockholm University. jonatan.riberth@iies.su.se.

1 Introduction

Immigration has reshaped the demographic and economic landscape of advanced Western economies. As of 2020, 281 million people—about 3.6 percent of the global population—were international migrants, and migratory pressures are likely to intensify as geopolitical conflict and climate change continue. Disparities in education and labor-market outcomes across immigrant-origin groups and their descendants therefore remain a central concern for policymakers and researchers (OECD, various; OECD/European Union, 2018), particularly given mounting evidence that such gaps can persist across multiple generations.

A large empirical literature following Borjas (1992, 1995) interprets these patterns as evidence that ethnic capital, the human capital of the ethnic or immigrant group, exerts an independent influence on intergenerational mobility and long-run assimilation.[1] This conclusion typically relies on regressions that include both parental characteristics and group averages of characteristics based on the ethnic or country-of-origin of the parents or the parents' ancestors,[2] and find significant coefficients on each. We show that such estimates can arise mechanically when extended-family influences are omitted. Whenever human capital is correlated within dynasties beyond the nuclear family, specifications that condition only on parental and group-level variables can attribute part of the family effect to the ethnic group. In this case, standard estimates need not identify group effects at all.

We revisit the question posed by Borjas (1992) of why convergence across immigrant groups is slower than expected from individual-level Markovian models. To do so, we extend the dynastic human capital framework of Adermon et al. (2021) to incorporate ethnic capital, thereby integrating two influential strands of research: the ethnic-capital literature and recent work modeling intergenerational mobility through latent human-capital factors using extended-family data (Adermon et al., 2021; Braun and Stuhler, 2018; Collado et al., 2023). This framework allows us to jointly estimate parental, extended-family, and group-level components of persistence, correcting standard ethnic-capital estimates for bias arising from measurement error in human capital and from omitted extended-family influences.

Our empirical analysis uses population-wide administrative data covering all individuals born in Sweden between 1972 and 2006. We link each individual to parents,

[1] See e.g. Aydemir et al., 2009, Sweetman and Dicks, 1999, Bauer and Riphahn, 2007a, Nielsen et al., 2003, Ward, 2020, among others.

[2] From here on, we mostly use the term *country of origin* instead of *ethnic* when referring to ethnic capital.

grandparents, and extended family members, defining country of origin by grandparents' country of birth. Our main outcome is the Grade Point Average (GPA) at the end of compulsory schooling—a standard proxy for human capital for the child generation—and years of schooling for the parental generation (e.g., Adermon et al., 2021).

Sweden provides a particularly suitable setting to study intergenerational human-capital assimilation—both between immigrants and natives and across immigrant groups of different origins—for at least two reasons. First, because Sweden remained neutral during World War II, large-scale immigration began shortly after the war, making it possible to observe third-generation outcomes today. Second, Sweden's administrative registers, notably the *Multi-Generation Register* linking the entire population across generations, combined with detailed education data, permit a comprehensive empirical analysis of intergenerational assimilation.

Figure 1 illustrates the stakes by showing the slow intergenerational convergence in schooling between immigrants and natives, and across immigrant groups. First-generation gaps are large—up to 4.5 years between East Asian and Middle Eastern/North African migrants. Convergence occurs across generations but is incomplete: nontrivial differences remain in the third generation, even among individuals fully educated in Sweden, which contradicts Becker and Tomes' (1986) idea that all human capital differences disappear in three generations in "open" societies (see also, e.g., Gielen and Webbink, 2025; Solon, 2018).

[Figure 1 about here.]

Our results indicate that dynastic human capital is the primary driver of third-generation outcomes. In our preferred specification, parental education accounts for roughly two-thirds of observed group-level persistence, extended-family factors account for between one-quarter and one-third, while ethnic capital explains approximately five percent. These patterns are robust to alternative specifications, including adjustments for measurement error and the inclusion of neighborhood fixed effects. Thus, what has often been interpreted as evidence of group-level ethnic capital appears largely to reflect intergenerational transmission within families.

We further examine mechanisms that could amplify persistence among immigrants. Limited institutional access may increase reliance on intra-family transmission; initial settlement patterns may generate persistent neighborhood effects; and our measure of extended-family networks may be a proxy for a a poorly measured ethnic capital variable. We assess these channels by estimating the model separately for different immigrant-origin groups as well as for natives, incorporating neighborhood fixed effects, and decomposing dynastic capital by country of origin.

Taken together, the results suggest that persistent disparities in immigrant integration are driven primarily by dynastic channels operating through the nuclear and extended family. While group heterogeneity and neighborhood influences are present, they do not materially alter this conclusion. A policy implication of our findings is that welfare-state policies aimed at equalizing outcomes across immigrant groups may be insufficient to speed long-run assimilation if dynastic transmission remains the dominant channel.

This paper reinterprets the ethnic-capital framework of Borjas (1992) by showing that conventional estimates of group effects may partly reflect dynastic transmission within extended families rather than group-level influences. Borjas noted already in his 1992 article that errors in the measure of fathers' human capital could bias the estimated ethnic-capital coefficient upward. As emphasized by Katz (2024), he also showed that implausibly large measurement error would be required to fully account for the ethnic-capital component estimated in his paper.[3] Ward (2020) examine related sources of spurious correlation, emphasizing measurement error in grandparental characteristics and persistence in settlement patterns. In contrast, we embed the ethnic-capital component in a framework that jointly models transmission operating through parents, extended families, and ethnic groups.

This perspective connects three literatures that have largely developed separately. Evidence on ethnic capital across countries—including Canada, Switzerland, and Denmark—has been mixed (Aydemir et al., 2009; Bauer and Riphahn, 2007b; Nielsen et al., 2003; Sweetman and Dicks, 1999). A related literature documents heterogeneous patterns in the evolution of native–immigrant and racial gaps, particularly in the limited set of studies using three-generation data (Abramitzky et al., 2021; Boustan et al., 2025; Chetty et al., 2020; Gielen and Webbink, 2025; Hammarstedt and Palme, 2012; Prokic-Breuer et al., forthcoming; Ward, 2020; Zhao and Drouhot, 2024; Zorlu and van Gent, 2024). Finally, recent work using extended-family data has developed methods to estimate latent intergenerational transmission parameters (Adermon et al., 2021; Braun and Stuhler, 2018; Collado et al., 2023). By integrating these approaches, we provide a framework that allows ethnic-capital effects to be estimated while accounting for dynastic transmission.

The remainder of the paper is organized as follows. Section 2 presents the empirical

[3]Subsequent papers, e.g., Card et al. (2000), among others, have raised related concerns in the broader literature on intergenerational mobility among immigrants.

framework. Section 3 describes the data and institutional context. Section 4 reports the main findings, and Section 5 concludes.

2 Empirical Framework

Our aim is to measure the rate of intergenerational human capital convergence between immigrant groups with different national origins and decompose this parameter into its underlying components. The rate of intergenerational convergence between country-of-origin groups is captured by the parameter α_1 in the model

$$\bar{y}^c_e = \alpha_0 + \alpha_1 \bar{y}^p_e + u_e, \tag{1}$$

where $\bar{y}^c_e$ and $\bar{y}^p_e$ are country-of-origin group averages of outcome y in the child and parent generations, respectively. A value of $0 \leq \alpha_1 < 1$ indicates convergence between groups, while $\alpha_1 > 1$ indicates divergence—i.e., groups drifting apart over time. Figure 1 shows evidence of convergence among descendants of immigrants to Sweden. In the following, we discuss how we estimate α_1, and how we can learn about its underlying components.[4]

2.1 Extending the Ethnic Capital Model of Borjas

We build on Borjas (1992), who estimated the model

$$y^c_i = \gamma_0 + \gamma_1 \bar{y}^p_{f(i)} + \gamma_3 \bar{y}^p_{e(i)} + \varepsilon_i, \tag{2}$$

where y^c_i is (a measure of) human capital for individual i; $\bar{y}^p_{f(i)}$ is average human capital of individual i's parents, indexed as family f; and $\bar{y}^p_{e(i)}$ is average human capital in individual i's country-of-origin group, $e(i)$.[5] Taking ethnic group averages of equation (2), we see that $\alpha_1 = \gamma_1 + \gamma_3$, allowing for the role of ethnic capital in explaining the slow convergence between immigrant groups.

There are several reasons why OLS estimates of γ_1 and γ_3 may be biased. These include selective patterns in the timing and location of immigration, as well as imperfect measurement of parental human capital (discussed below). A less explored concern is an overly narrow definition of the family. In the standard Borjas framework, factors

[4]Since our framework takes the individual as the unit of observation, and since country-of-origin groups e vary in size, we work with a weighted version of equation (1)—or equivalently, an individual-level regression where we replace $\bar{y}^c_e$ by individual outcomes y^c_i.

[5]The functions $f(i)$ and $e(i)$ assign to each individual i a family f and a country-of-origin group e, respectively.

operating through the extended family may be incorrectly attributed either to parents or to the ethnic group, thereby biasing the estimated coefficients.[6]

Leveraging our ability to identify extended-family links for all individuals in the sample, we adopt an approach similar to Adermon et al. (2021) to explicitly model extended-family human capital. This enables us to separate ethnic capital effects from extended-family influences in the intergenerational transmission of outcomes.[7]

This approach can be generalized to include any set of nested averages. We extend equation (2) to also incorporate the role of the extended family:

$$y_i^c = \gamma_0 + \gamma_1 \bar{y}^p_{f(i)} + \gamma_2 \bar{y}^p_{d(i)} + \gamma_3 \bar{y}^p_{e(i)} + \varepsilon_i, \tag{3}$$

where $\bar{y}^p_{f(i)}$ is the average education of the parents, $\bar{y}^p_{d(i)}$ is the average education of the members of the extended family (denoted as d for "dynasty") in the parents' generation, and $\bar{y}^p_{e(i)}$ is the average education of the members of the country-of-origin group in the parental generation.

Assuming that $i \subset f \subset d \subset e$, we can again average equation (3) by country-of-origin group to obtain

$$\bar{y}^c_e = \gamma_0 + (\gamma_1 + \gamma_2 + \gamma_3)\bar{y}^p_e + \bar{\varepsilon}_e. \tag{4}$$

This approach thus allows us to decompose group-level convergence into three channels: the *nuclear family*, the *extended family*, and the *country-of-origin group*.

Appendix A formally shows that omitting the variable measuring extended-family human capital leads to an upward bias in the estimated coefficient on average human capital of the country-of-origin group, under the weak assumptions that, conditional on parental human capital, (i) country-of-origin group human capital is correlated with extended-family human capital,and (ii) extended-family human capital is positively associated with the child's educational outcome. Appendix A also establishes the analogous result for the coefficient on parental education.

If observed education in the parental generation fails to fully capture the human capital that is transmitted to children, the models discussed so far can yield misleading estimates. Classical measurement error of parents' education will bias an estimate of

[6]Different arguments for that ignoring other family members is likely to affect comparisons of intergenerational mobility between natives and immigrants have appeared in the previous literature—see e.g. Card et al. (2000) and Dustmann and Glitz (2011).

[7]A key difference relative to Adermon et al. (2021) is that their objective is to estimate the intergenerational transmission of latent human capital from parents to children, whereas the present paper aims to decompose the components of a group-level convergence parameter. Accordingly, we adopt a different standardization: whereas they normalize each group average to unit variance, we allow group averages to retain their original dispersion.

γ_1 downward and γ_3 upward.[8] Although we have excellent administrative data on the education of children, parents and the extended family, we also examine this by specifying a latent variables model (see Appendix B), and use observable proxies to get alternative estimates of the parameters in this model.[9]

Due to intermarriage, members of the same extended family may belong to different country-of-origin groups, making the assumption of perfectly nested groups ($f \subset d \subset e$) unlikely to hold in the data. In this case, equation (4), where we have $\alpha_1 = \gamma_1 + \gamma_2 + \gamma_3$, does not hold exactly. Since a large share of parent couples are formed outside the country-of-origin group (mostly natives, sometimes immigrants from other countries of origin), our main analyses include measures of human capital for the parents and the extended family that includes these individuals.

To take into account that groups can be non-nested, we also decompose the group level convergence parameter α_1 using the approach outlined in Gelbach (2016). Using the coefficients from the individual-level version of the baseline model from equation (1), and the full model from equation (3), the underlying sources of the intergenerational group persistence parameter can be decomposed as:

$$\underbrace{\alpha_1}_{\substack{\text{ethnic group}\\\text{convergence}}} = \underbrace{\gamma_1\delta_f}_{\substack{\text{parents'}\\\text{human capital}}} + \underbrace{\gamma_2\delta_d}_{\substack{\text{dynastic}\\\text{human capital}}} + \underbrace{\gamma_3}_{\substack{\text{ethnic}\\\text{human capital}}}, \quad (5)$$

where δ_f is the coefficient from regressing $\bar{y}^p_{f(i)}$ on $\bar{y}^p_{e(i)}$, and δ_d is the coefficient from regressing $\bar{y}^p_{d(i)}$ on $\bar{y}^p_{e(i)}$. Note that if the observations are nested, then $\delta_f = \delta_d = 1$, and we are back to the simple model we discussed above, where $\alpha_1 = \gamma_1 + \gamma_2 + \gamma_3$.

2.2 Other Specification and Inference Issues

In the models above, we include additional controls for observable characteristics beyond human capital to make immigrant families comparable along other dimensions. To account for variation in arrival timing, we include individual birth-year fixed effects and anchor immigrants' human capital to that of natives by subtracting the mean human

[8]This holds if measurement error is averaged out of the country-of-origin mean and parents' and country-of-origin mean schooling are positively correlated.

[9]This type of latent variable model has been discussed by, e.g., Adermon et al. (2021), Braun and Stuhler (2018), Clark (2014), Collado et al. (2023), and Stuhler (2014). The additional proxies we use are measures of mid-life labor income and an occupation-based index of social status (see Section 3.3). Accordingly, our estimates of the relationship between children's GPA and parental, extended-family, and country-of-origin group "human capital" (or social status) should be interpreted as capturing broader mechanisms not reflected in years of schooling alone. These include employers' valuation of human capital (including potential discrimination), occupational choice, and access to job networks.

capital of natives born in the same year.[10]

As pointed out in several previous studies (see, e.g., Borjas, 1995; Bratu and Bolotnyy, 2023; Ward, 2020), endogenous geographical sorting might play an important role in explaining patterns of intergenerational mobility among immigrant groups. If immigrants are not conditionally randomly assigned to neighborhoods and segregation is persistent across generations, group convergence might be driven by initial neighborhood location. Using detailed data on the *first* neighborhood at arrival for immigrant ancestors, we control for endogenous sorting of immigrant groups to regions with more or less potential for integration at the time of arrival. We do this by including fixed effects for first location.

To increase the precision of our estimates, we employ a feasible generalized least squares (FGLS) estimator with random effects at the country-group level, with standard errors clustered at the same level. This specification follows the recommendations of Wooldridge (2003, 2010) and Cameron and Miller (2015), who note that combining an explicit error-components model with cluster-robust inference can deliver efficiency gains while remaining robust to misspecification of the covariance structure.[11] Importantly, the FGLS and OLS point estimates are very similar and yield identical qualitative conclusions, while FGLS yields more precise estimates of the ethnic mean variable, with standard errors roughly 20 percent smaller in our baseline specifications.

3 Data and Brief Historical Background

Our dataset is constructed by linking individual-level data from several Swedish administrative registers. We use three generations of linked individuals, which we refer to as the *child*, *parent*, and *grandparent* generations. We observe educational outcomes for the child and parent generations, while the grandparent generation is used to assign country of origin, arrival neighborhood, and to identify extended family relationships.[12]

Our main sample consists of children born in Sweden between 1972 and 2006, the

[10] This adjustment is important because average years of schooling among natives trend upward over time. It also allows our estimates to be interpreted directly as convergence relative to natives.

[11] Cluster-robust standard errors allow for very general forms of heteroskedasticity and within-group dependence but may reduce statistical power, whereas GLS-type estimators can improve efficiency when the error structure is reasonably well approximated. Cameron and Miller (2015) note: "It is remarkable that current econometric practice with clustered errors ignores the potential efficiency gains of FGLS" (p. 326). The approach has been shown to lead to potentially large gains in power in the case of heteroskedasticity (Romano and Wolf, 2017) and difference-in-differences (Brewer et al., 2018), without compromising robustness.

[12] Siblings are identified via shared parentage. For immigrants, sibling links are typically observed only if individuals migrated to Sweden together with a parent before age 18.

cohorts for which grade point average (GPA) at the end of compulsory school is observed. Using the Swedish Multi-generation Register, we identify each child's grandparents, which allows us to reconstruct extended family networks—including aunts, uncles, and their spouses. In Figure C.2, we show the average number of individuals in each family category for the 18 most common countries of origin. Two patterns stand out. First, for almost all countries, we observe many relatives over whom we average education. Second, most individuals are born in Sweden, regardless of dynasty category. Taken together, although individuals have foreign ancestry, we measure parental-generation human capital using many relatives who largely received their education in Sweden.

We include children who have at least one parent who: (i) was born in Sweden; (ii) is of foreign background, defined as having both parents born outside the Nordic countries; and (iii) has at least one sibling in the data, which implies that the child has at least one aunt or uncle of foreign origin. About 9% of these children have missing GPA in our data. Missingness is weakly negatively related to parental education, and appears largely driven by migration: 31.5% of children with missing GPA have a recorded migration event (in- or out-migration), compared to about 6% among those with observed GPA.

The country composition of our sample reflects historical immigration flows. Most arrivals for the grandparent (first) generation occurred between 1945 and 1980.

3.1 Migration to Sweden 1945–1980

Sweden provides a valuable setting for studying long-term migration effects, having emerged as a stable and attractive destination country for migrants after World War II. The period from 1945 to 1980 encompasses three distinct phases of migration, shaped by geopolitical developments and domestic labor market needs (Lundh, 2005; OECD, various; Swedish Migration Agency, 2025).

1945–1950s: Post-war humanitarian and Nordic migration. Sweden initially received refugees and displaced persons from the Baltic states (Estonia, Latvia, Lithuania), Finland, Germany, and Poland. Following the creation of the Nordic Passport Union in 1954, migration from neighboring Nordic countries increased substantially (Nordic Council, 1954). The Cold War and increased political repression in Eastern Europe also triggered new refugee inflows from these countries (Byström, 2006).

1950s–1960s: Labor migration and industrial expansion. Sweden's economic boom created a demand for labor, prompting government-led recruitment—especially from Italy, Yugoslavia, Greece, and Turkey. Migrants often worked in construction and manufacturing, under regulated bilateral labor agreements (Lundh, 2005).

Late 1960s–1970s: Policy restriction and humanitarian turn. With rising unemployment in the late 1960s, Sweden imposed restrictions on labor immigration. Work permit requirements were introduced in 1967, and labor recruitment ended in 1972. The focus shifted toward family reunification and refugee admissions. Notable groups included refugees from Chile (after the 1973 coup), Uganda (expelled Asians), and Southeast Asia (post-Indochina wars). In 1975, Sweden adopted an official integration policy promoting multiculturalism and equal rights (Borevi, 2002).

Figure C.1 plots the year of immigration for ancestors from different country groups in our sample. The different immigration waves described above are all represented.

3.2 Ancestry Composition and Country of Origin Assignment

Our analysis excludes individuals of Nordic ancestry (i.e., those with grandparents from Denmark, Finland, Iceland, or Norway), as these populations are relatively similar to Sweden in terms of culture, language, and appearance.[13] Table C.1 presents the distribution of countries of origin for the child generation in our dataset. The largest groups originate from other European countries, with Germany and former Yugoslavia being particularly prominent. Turkey also constitutes a major group, while the US and Chile represent the largest non-European origins.

We assign country of origin to children based on their grandparents' countries of birth. If all observed grandparents are born in the same country, the child is assigned that country. If grandparents are born in different non-Nordic countries, the child is assigned proportional weights for each country. For example, a child with one set of grandparents from Germany and another from Poland is included in the analysis twice—once under each origin—with a weight of 0.5 in each case. This weighting scheme enables us to include children from mixed-background families without losing statistical representativeness or comparability across groups.

Our focus on third-generation immigrants allows us to study long-term outcomes without conditioning on parental marital patterns. In many cases, children have one native-born parent and one foreign-born parent. Figure C.2 displays the distribution of foreign and native ancestry across children in the sample.

For nearly all children in the dataset, we observe both parents. Approximately 86%

[13] In Section 4.3, we show that our results are robust to including them.

of children are assigned a single country of origin, while 13% are assigned two countries, and fewer than 1% are assigned three or more.

3.3 Variable Definitions

For the child generation, human capital is measured using Grade Point Average (*GPA*), defined as the average across all compulsory subjects in ninth grade—the final year of compulsory schooling (typically at age 16). We convert this measure into within-cohort percentile ranks.

For the parental generation, we use *years of schooling* to measure human capital. This measure is based on educational attainment recorded in the National Education Register (1985–2020) and census data from 1960, 1970, and 1990. Individuals educated outside Sweden report their education to the Migration Agency upon arrival.[14]

We calculate three sets of group averages of the human capital variables in the parental generation. First, *parental* outcomes are simply averaged over both parents. Second, we define the *extended family* as the child's aunts and uncles, their spouses, and those spouses' siblings. Siblings are defined as individuals sharing both biological parents. Third, the *country-of-origin mean* captures the average outcome for individuals in the parental generation with a given ancestry.

For each parent, the country mean is defined as the average human capital among individuals who (i) have the same ancestry as the parent,[15] (ii) are born within five years of the parent, (iii) are not members of the child's extended family, and (iv) are born in Sweden or moved to Sweden before age 9. The country mean assigned to the child is then the mean of the two parents' country means.[16] By focusing on individuals who are born in Sweden or came before age nine, we mitigate issues related to differences in age of arrival to Sweden between different country groups.[17]

We also define a child's *neighborhood* as the first parish that a foreign-born ancestor lived in when they first came to Sweden.

[14] In a robustness analysis, we also use *lifetime income* and an occupation-based *social stratification index* for the parental generation—see Section 4.3 and Appendix B

[15] Ancestry is defined as grandparents' country of birth or own country of birth if the individual is born outside Sweden.

[16] If a parent has two countries of origin (e.g., if the maternal grandparents were born in two different countries), we assign the parent one of these origins at random, and then proceed to calculate their country mean.

[17] Böhlmark (2008) shows that arriving after age nine has a strong negative impact on school performance.

4 Results

This section presents our results. The results from the core specification (see Section 2), which measures the excess intergenerational persistence from ethnic and dynastic human capital, are presented in Section 4.1. Section 4.3 assesses the robustness of our key findings, while Section 4.4 presents separate estimates by country-of-origin groups.

4.1 Parents, Extended Family and Ethnic Capital

Figure 2 highlights the paper's central motivation: slow convergence in human capital across country-of-origin groups. It plots average years of schooling in the second generation against that of the third generation by country group, defined by the first-generation immigrants (see equation (1)). The estimated slope implies that over 60 percent of the educational advantage observed in the second generation persists into the third—substantially higher than persistence typically found in individual-level Swedish data.

[Figure 2 about here.]

Table 1 presents our main results: a decomposition of group-level persistence into parental, extended-family, and country-of-origin components. Each column reports estimates from a separate specification in which children's educational outcomes are regressed on different combinations of human capital measures defined at these levels. All specifications include fixed effects for birth year, grandparents' initial neighborhood, and grandparents' year of migration to Sweden.

For interpretability, the coefficient on parental schooling is expressed in standard-deviation units. In column 2, for instance, a one-SD increase in parental schooling is associated with a 0.4-SD increase in child GPA. The extended-family and country-mean schooling variables are scaled using the same standard deviation, allowing direct comparison across coefficients.

[Table 1 about here.]

Column 1 mirrors the approach in Figure 2, but applies it at the individual level rather than the group level. As expected, the results confirm strong intergenerational persistence, with the GPA of the child closely related to the average years of schooling of their country group in the parental generation. Column 2 estimates a Markovian AR(1) model, widely used in research on intergenerational mobility since Becker and

Tomes (1986). The point estimate of 0.44 is in line with previous Swedish studies (e.g., Adermon et al., 2021).

Column 3 presents results from the canonical model of Borjas (1992) (see equation (2)), which includes the (leave-out) average educational attainment among individuals in the parental generation from the same country of origin. The coefficient on parental education remains similar to that in column 2, while the estimate for ethnic capital is moderate, substantially below the levels reported in some earlier studies using data on three generations (e.g., Borjas, 1992; Ward, 2020) that found ethnic capital to be comparable in magnitude to the coefficient for parental human capital. Although the coefficient is imprecisely estimated, the data rule out similarly large ethnic capital effects in this context.

Column 4 extends the baseline model by incorporating the mean schooling of extended family members, following the dynastic human capital framework of Adermon et al. (2021). This allows us to capture broader familial influences beyond the nuclear household. The results confirm that extended family education is a strong predictor of child outcomes, even after accounting for parental education, underscoring the importance of extended family networks in shaping educational attainment.

Finally, column 5 includes all three components: parental, extended family, and country-of-origin group, as in equation (3). In this full specification, the coefficient for ethnic capital becomes very close to zero, suggesting that once family-level human capital is accounted for, country-of-origin effects are no longer important for explaining differences in third-generation outcomes.[18]

As discussed in Section 2.1, we decompose the overall group convergence parameter into contributions from parents, the extended family, and ethnic capital. We implement this in two complementary ways: the first simply sums the relevant regression coefficients to recover the share of convergence explained by each channel (sums are reported at the bottom of Table 1), while the second applies a Gelbach decomposition, which accounts for the overlapping nature of family and group characteristics by reweighting the coefficients accordingly (see equation (5)).[19]

Using the estimates in column 5, the direct approach attributes 61 percent of group convergence to parents, 34 percent to the extended family, and 5 percent to ethnic

[18]As discussed in Section 2.2, Table 1 is estimated using a combination of FGLS and cluster-robust standard errors. In Table C.3, we present corresponding estimates using OLS with cluster-robust standard errors. With the exception of a somewhat larger country-of-origin coefficient in Column 1, estimates are very similar, although with substantially larger standard errors on the country-of-origin mean.

[19]The Gelbach weights are estimated by regressing parental and extended family education on the country mean, controlling for cohort and neighborhood fixed effects.

capital. The Gelbach decomposition, based on the overall country mean in column 1, yields similar results: parents account for 70 percent, the extended family for 25 percent, and country-of-origin for 5 percent.

Taken together, the results shown in Table 1 imply that the observed group-level differences in human capital across country-of-origin groups in the third generation are not driven by ethnic capital effects, such as culture, norms, or ethnic peer effects. Rather, the persistence seen in Figure 1 appears to reflect initial disparities in human capital and the strength of intergenerational transmission within families broadly defined, which have been ignored in previous studies.

4.2 The Extended Family as a Proxy for Ethnic Capital

An alternative interpretation of our results is that ethnic capital is absorbed by extended-family controls not because extended-family influences are omitted in the standard framework, but because the country-of-origin measure of ethnic capital is measured with error, and extended-family variables provide a better proxy for ethnic capital than the country mean.[20] Under this interpretation, our estimates would understate the importance of ethnic capital.

We assess this possibility in Table 2 by estimating specifications that distinguish between extended-family members of foreign and native origin. If extended-family variables primarily capture ethnic capital, we would expect (i) larger coefficients for relatives of foreign origin and (ii) a stronger attenuation of the country-mean coefficient when controlling for foreign-origin relatives than when controlling for native-origin relatives. We find neither pattern. Instead, the coefficient on native-origin relatives is at least as large as that on foreign-origin relatives, both when included separately (columns 2–3) and together (column 4). Furthermore, the country-mean coefficient is identical whether we include only native-origin or only foreign-origin extended family members (columns 2–3). We therefore conclude that extended-family variables capture family-specific factors distinct from ethnic capital.[21]

[Table 2 about here.]

[20]One reason is that a single country of origin may comprise multiple ethnic groups, which may be more accurately captured by extended-family links.

[21]The sample used differs slightly from that in Table 1. In the main analysis, we require at least one observation for each *type* of extended family member (i.e., aunts and uncles, their spouses, and the spouses' siblings). Applying this restriction to the separate extended family variables in this analysis would be too restrictive, so we instead require at least one observation for *any* extended family member of native and foreign origin, respectively.

4.3 Robustness of Key Results

We assess the robustness of our main estimates using five sensitivity checks:

1. *Neighborhood effects.* To examine whether our results are driven by neighborhood composition, we include fixed effects for children's neighborhoods.

2. *Inclusion of Nordic immigrants.* The main analysis excludes immigrants from other Nordic countries (Denmark, Finland, Iceland, and Norway), who could be ethnically relatively similar to Swedes.[22] As a robustness check, we expand the sample to include these groups.

3. *Reweighting of country groups.* To reduce the influence of larger origin groups, we reweight each observation by the inverse of the number of individuals in the group, so that each country of origin is given the same weight in the regressions.

4. *Multiple proxies for parental human capital.* Parental, extended family, and country-of-origin schooling measures are imperfect proxies for the respective groups' latent human capital. Since they are averages of the same underlying variable (years of schooling in the parental generation) over different sample sizes, they also differ in the extent of measurement error,[23] which may bias the estimated relative contributions.[24] We follow the proxy-variable framework in Lubotsky and Wittenberg (2006), combining years of schooling, labor income, and occupation-based social stratification to form a broader measure of parental human capital and to mitigate attenuation bias. See Appendix B for details.

5. *Excluding arrival neighborhood controls.* The main analysis includes fixed effects for the parish of arrival of the first ancestor to immigrate to Sweden. We check the sensitivity of our estimates by dropping these controls.

[Figure 3 about here.]

Figure 3 shows the results of each of these sensitivity checks.[25] The boxes correspond to the columns in Table 1, with rows showing the different human capital variables.

[22] However, Finnish immigrants and their descendants—who make up the vast majority of this group—had a distinct ethnic identity and were a socially disadvantaged group in Sweden well into the 1990s (see, e.g., Koivunen, 2017; Saarela and Finnäs, 2007; Saarela and Rooth, 2006; Weckström, 2011).

[23] If the error is classical, it will average out more in the country mean than in the extended family and parental means.

[24] Ward (2020) documents this in US data.

[25] These estimates are also presented in Table C.4.

Coefficients from the different sensitivity analyses outlined above are delineated using different shapes and colors.

Overall, our findings are robust across specifications. Ethnic capital plays only a modest role in explaining the slow convergence across groups, especially once dynastic human capital is incorporated into the Borjas framework.

In addition to these checks, Table C.2 reports results using alternative proxies for ethnic human capital. We consider education and cultural indices from the World Values Survey (WVS) as well as average years of schooling in the descendants' country of origin from Barro and Lee (2013) as alternative proxies for ethnic human capital. These measures yield a somewhat larger coefficient on the country mean. When we combine all proxies, including the country mean, the estimated contribution of ethnic human capital is very similar to that in the baseline model and the alternative proxies are jointly statistically insignificant.[26]

Finally, we estimate separate models by gender in the child generation. Results are qualitatively similar (see Table C.5). For men, the country-mean coefficient falls by half when controlling for the extended family, from 0.150 to 0.086. For women, the country-mean coefficient is negligible even before controlling for the country-of-origin mean, indicating that ethnic capital is more important for men than for women.

4.4 Between-Group Heterogeneity

Cross-country differences in intergenerational mobility are well documented (Corak, 2013; Manduca et al., 2024; Solon, 2004). Explanations typically emphasize inequality, welfare-state institutions, and cultural norms. Similar mechanisms may also operate across immigrant or ethnic groups within a given society. Immigrants often face language barriers and weaker networks, which may limit access to public institutions that facilitate human capital formation, such as schools and libraries. As a result, the family may play a larger role in the transmission of human capital, potentially leading to stronger intergenerational persistence within ethnic groups. If so, slower convergence across groups with different national origins may reflect differences in family-based transmission rather than an independent effect of *ethnic capital*.

Table 3 reports dynastic human capital models estimated separately by country-of-origin groups. We distinguish between parental education, the extended family, and the full family (our measure of within-group persistence). Estimates are shown for natives,

[26]The p-value from an F-test for the hyphotesis that all the alternative proxies in column (5) is equal to zero is 0.629.

all immigrants, and immigrant subgroups by region of origin. This allows us to compare how the strength of family-based transmission varies across groups.

The results provide limited support for the hypothesis that higher overall intergenerational persistence is the primary mechanism behind slow convergence across origin groups, or that cultural distance to the destination country drives this pattern. First, the point estimates for immigrants (0.559) and natives (0.554) are virtually identical. Second, we find no systematic evidence of higher persistence among groups originating from culturally more distant countries. For instance, immigrants from Southern Europe, Latin America, and MENA countries all fall in the lower half of the persistence ranking shown in Table 3.

[Table 3 about here.]

5 Conclusions

This paper shows that estimates of ethnic capital are highly sensitive to model specification and can largely reflect omitted extended-family influences rather than independent group-level effects. Using data that allow us to observe complete multigenerational family networks, we demonstrate that once extended-family human capital is accounted for, the apparent role of ethnic capital is substantially reduced. This pattern arises mechanically whenever human capital is correlated within dynasties beyond the nuclear family. Our findings therefore suggest a reinterpretation of existing evidence: standard specifications may conflate group effects with dynastic transmission. Distinguishing between these mechanisms is essential for understanding the sources of persistence and for designing policies aimed at reducing inequality across groups.

References

Abramitzky, R., L. Boustan, E. Jacome, and S. Perez (2021). “Intergenerational Mobility of Immigrants in the United States over Two Centuries”. *American Economic Review* 111.2, pp. 580–608.

Adermon, A., M. Lindahl, and M. Palme (2021). “Dynastic Human Capital, Inequality, and Intergenerational Mobility”. *American Economic Review* 111.5, pp. 1523–1548.

Amemiya, T. (1971). “The Estimation of the Variances in a Variance-Components Model”. *International Economic Review* 12.1, pp. 1–13.

Aydemir, A., W.-H. Chen, and M. Corak (2009). "Intergenerational Earnings Mobility among the Children of Canadian Immigrants". *The Review of Economics and Statistics* 91.2, pp. 377–397.

Barro, R. J. and J. W. Lee (2013). "A New Data Set of Educational Attainment in the World, 1950–2010". *Journal of Development Economics* 104, pp. 184–198.

Bauer, P. and R. T. Riphahn (2007a). "Heterogeneity in the Intergenerational Transmission of Educational Attainment: Evidence from Switzerland on Natives and Second-Generation Immigrants". *Journal of Population Economics* 20.1, pp. 121–148.

— (2007b). "Heterogeneity in the Intergenerational Transmission of Educational Attainment: Evidence from Switzerland on Natives and Second-Generation Immigrants". *Journal of Population Economics* 20.1, pp. 121–148.

Becker, G. S. and N. Tomes (1986). "Human Capital and the Rise and Fall of Families". *Journal of Labor Economics* 4.3, S1–S39.

Böhlmark, A. (2008). "Age at Immigration and School Performance: A Siblings Analysis Using Swedish Register Data". *Labour Economics* 15.6, pp. 1366–1387.

Borevi, K. (2002). *Välfärdsstaten i det mångkulturella samhället: Socialdemokratisk integrationspolitik under efterkrigstiden.* Uppsala: Acta Universitatis Upsaliensis.

Borjas, G. J. (1992). "Ethnic Capital and Intergenerational Mobility". *The Quarterly Journal of Economics* 107.1, pp. 123–150.

— (1995). "Ethnicity, Neighborhoods, and Human-Capital Externalities". *The American Economic Review* 85.3, pp. 365–390.

Boustan, L., M. F. Jensen, R. Abramitzky, E. Jácome, A. Manning, S. Pérez, A. Watley, A. Adermon, J. Arellano-Bover, O. Åslund, M. Connolly, N. Deutscher, A. C. Gielen, Y. Giesing, Y. Govind, M. Halla, D. Hangartner, Y. Jiang, C. Karmel, F. Landaud, L. Macmillan, I. Z. Martínez, A. Polo, P. Poutvaara, H. Rapoport, S. Roman, K. G. Salvanes, S. San, M. Siegenthaler, L. Sirugue, J. S. Espín, J. Stuhler, G. L. Violante, D. Webbink, A. Weber, J. Zhang, A. Zheng, and T. Zohar (2025). *Intergenerational Mobility of Immigrants in 15 Destination Countries.* Working Paper 33558. National Bureau of Economic Research.

Bratu, C. and V. Bolotnyy (2023). "Immigrant Intergenerational Mobility: A Focus on Childhood Environment". *European Economic Review* 151, p. 104353.

Braun, S. T. and J. Stuhler (2018). "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany". *The Economic Journal* 128.609, pp. 576–611.

Brewer, M., T. F. Crossley, and R. Joyce (2018). "Inference with Difference-in-Differences Revisited". *Journal of Econometric Methods* 7.1.

Byström, M. (2006). *En flykting korsar sitt spår: Folkbokföring, flyktingar och den svenska välfärdsstaten 1944–1951*. Stockholm: Almqvist & Wiksell.

Cameron, A. C. and D. L. Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference". *Journal of Human Resources* 50.2, pp. 317–372.

Card, D., J. DiNardo, and E. Estes (2000). "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s". In: *Issues in the Economics of Immigration*. University of Chicago Press, pp. 227–270.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020). "Race and Economic Opportunity in the United States: An Intergenerational Perspective". *The Quarterly Journal of Economics* 135.2, pp. 711–783.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press.

Collado, M. D., I. Ortuño-Ortín, and J. Stuhler (2023). "Estimating Intergenerational and Assortative Processes in Extended Family Data". *The Review of Economic Studies* 90.3, pp. 1195–1227.

Corak, M. (2013). "Income inequality, equality of opportunity, and intergenerational mobility". *Journal of Economic Perspectives* 27.3, pp. 79–102.

Dustmann, C. and A. Glitz (2011). "Migration and Education". In: *Handbook of the Economics of Education*. Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 4. Handbook of The Economics of Education. Elsevier, pp. 327–439.

Gelbach, J. B. (2016). "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics* 34.2, pp. 509–543.

Gielen, A. C. and D. Webbink (2025). "Unexpected colonial returns". *Journal of Human Resources*.

Hammarstedt, M. and M. Palme (2012). "Human Capital Transmission and the Earnings of Second-Generation Immigrants in Sweden". *IZA Journal of Migration* 1.1, pp. 1–23.

Katz, L. F. (2024). "The Economics of Immigration: A Festschrift in Honor of George J. Borjas". *ILR Review*, p. 00197939241264713.

Koivunen, A. (2017). "Economies of Pride and Shame: Politics of Affect in New Narratives about Sweden Finns". In: *Citizenships under Construction: Affects, Politics and Practices*. Ed. by K. De Graeve, R. Rossi, and K. Mäkinen. COLLeGIUM: Studies across Disciplines in the Humanities and Social Sciences 23. Helsinki: Helsinki Collegium for Advanced Studies, pp. 50–66.

Lambert, P. S. and E. Bihagen (2012). *CAMSIS Sweden*. `https://www.camsis.stir.ac.uk/Data/Sweden90.html`.

Lubotsky, D. and M. Wittenberg (2006). “Interpretation of Regressions with Multiple Proxies”. *Review of Economics and Statistics* 88.3, pp. 549–562.

Lundh, C. (2005). *Invandringens arbetsmarknad: Ett historiskt perspektiv.* Stockholm: SNS Förlag.

Manduca, R., M. Hell, A. Adermon, J. Blanden, E. Bratberg, A. C. Gielen, H. van Kippersluis, K. Lee, S. Machin, M. D. Munk, M. Nybom, Y. Ostrovsky, S. Rahman, and O. Sirniö (2024). “Measuring Absolute Income Mobility: Lessons from North America and Europe”. *American Economic Journal: Applied Economics* 16.2, pp. 1–30.

Nerlove, M. (1971). “Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections”. *Econometrica* 39.2, pp. 359–382.

Nielsen, H. S., M. Rosholm, N. Smith, and L. Husted (2003). “The School-to-Work Transition of 2nd Generation Immigrants in Denmark”. *Journal of Population Economics* 16.4, pp. 755–786.

Nordic Council (1954). *Nordic Passport Union Agreement (1954).* `https://www.norden.org/en/info-norden/nordic-passport-union`. Accessed: 2025-05-13.

OECD (various). *International Migration Outlook: Sweden Country Notes and Labor Migration Reports.* `https://www.oecd.org/migration`. Accessed: 2025-05-13.

Prokic-Breuer, T., S. Vermeulen, and D. Webbink (forthcoming). “On the economic prospects of non-Western migrants in Europe”. *Journal of Population Economics.*

Romano, J. P. and M. Wolf (2017). “Resurrecting Weighted Least Squares”. *Journal of Econometrics* 197.1, pp. 1–19.

Saarela, J. and F. Finnäs (2007). “Adjustment failures in an immigrant population: Finns in Sweden”. *Social Indicators Research* 82.3, pp. 545–563.

Saarela, J. and D.-O. Rooth (2006). “How Integrated are Finns in the Swedish Labour Market? Outcomes of Free Labour Mobility”. *International Migration* 44.2, pp. 119–152.

Solon, G. (2004). “A Model of Intergenerational Mobility Variation Over Time and Place”. In: *Generational Income Mobility in North America and Europe.* Ed. by M. Corak. Cambridge: Cambridge University Press, pp. 38–47.

— (2018). “What do we know so far about multigenerational mobility?” *The Economic Journal* 128.612, F340–F352.

Stuhler, J. (2014). “Mobility Across Multiple Generations: The Iterated Regression Fallacy”. Mimeo.

Swedish Migration Agency (2025). *Official statistics and historical overview.* `https://www.migrationsverket.se`. Accessed: 2025-05-13.

Sweetman, A. and G. Dicks (1999). "Education and Ethnicity in Canada: An Intergenerational Perspective". *The Journal of Human Resources* 34.4, pp. 668–696.

Vosters, K. (2018). "Is the Simple Law of Mobility Really a Law? Testing Clark's Hypothesis". *The Economic Journal* 128.612, F404–F421.

Vosters, K. and M. Nybom (2017). "Intergenerational Persistence in Latent Socioeconomic Status: Evidence from Sweden and the United States". *Journal of Labor Economics* 35.3, pp. 869–901.

Ward, Z. (2020). "The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration". *American Economic Journal: Applied Economics* 12.4, pp. 73–102.

Weckström, L. (2011). *Representations of Finnishness in Sweden.* Studia Fennica Linguistica 16. Helsinki: Finnish Literature Society.

Wooldridge, J. M. (2003). "Cluster-Sample Methods in Applied Econometrics". *The American Economic Review* 93.2, pp. 133–138.

— (2010). *Econometric Analysis of Cross Section and Panel Data.* Second Edition. Cambridge: The MIT Press.

Zhao, L. and L. G. Drouhot (2024). "The Grandchildren of Immigrants in Western Europe: Patterns of Assimilation Among the Emerging Third Generation". *Demography* 61.2, pp. 463–491.

Zorlu, A. and W. van Gent (2024). "Economic Assimilation of the "Third Generation": An Intergenerational Mobility Perspective". *International Migration Review* 58.2, pp. 734–763.

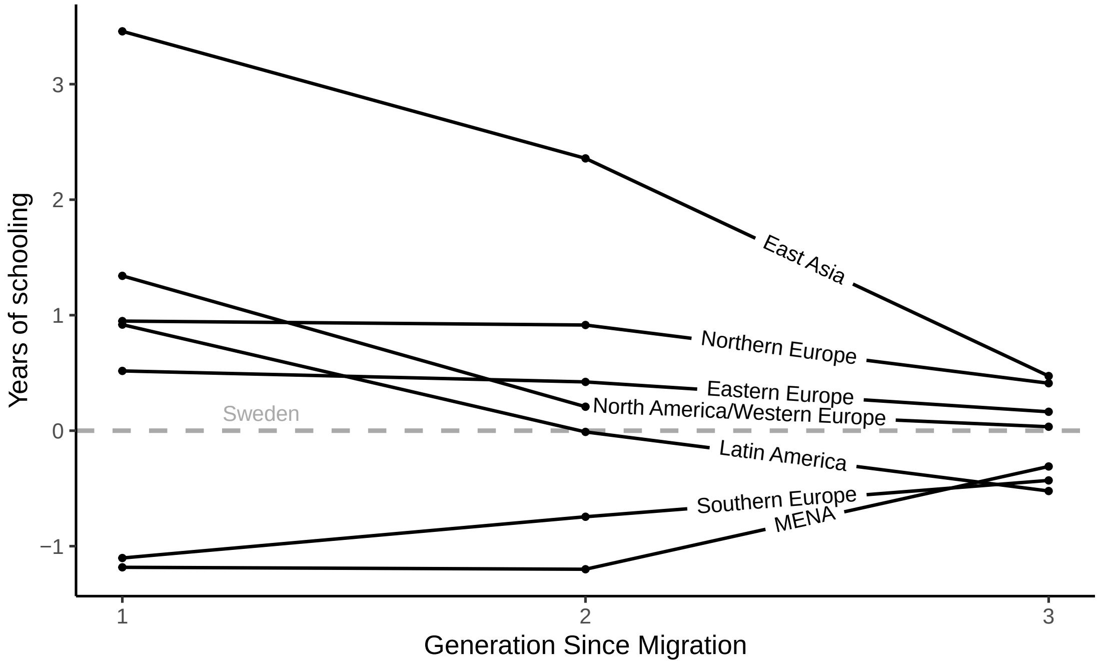

Figure 1: Convergence Between Ethnic Groups

Note: Average years of schooling relative to Swedes born the same year for different origin groups. 1 = Migrants, 2 = Children, 3 = Grandchildren. Second- and third-generation individuals are weighted by ancestral country group. Sample restricted to dynasties where second-generation immigrants were born in Sweden or arrived before age 9.

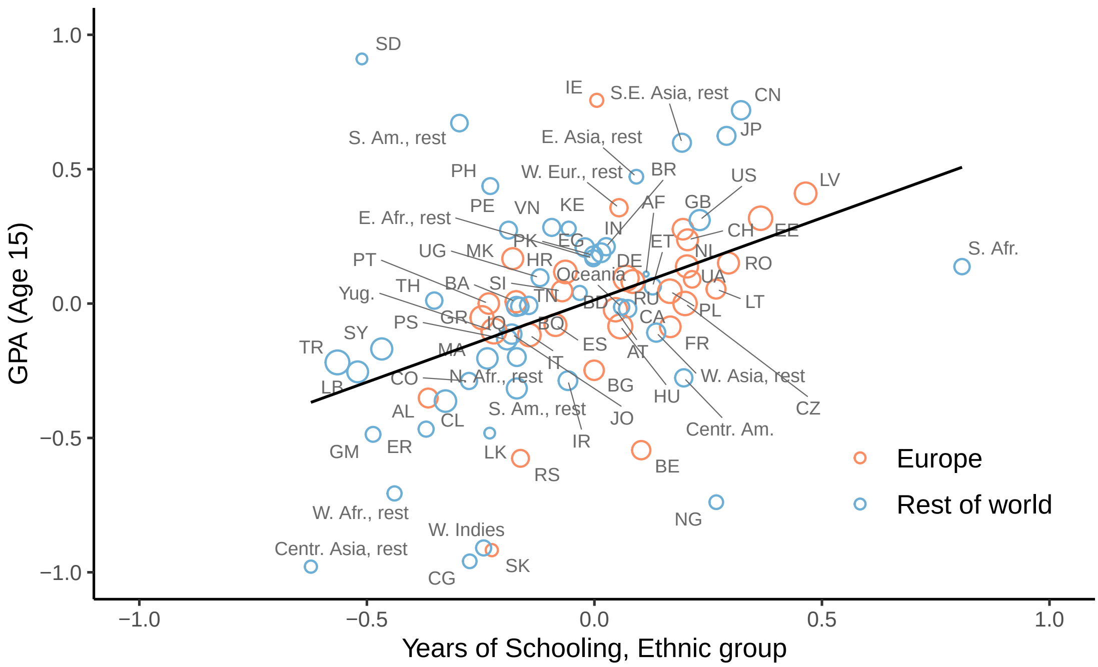

Figure 2: Country-Level Scatter Plot

Note: This figure plots average GPA against average country group means by country of origin. Circle sizes are proportional to the logarithm of the number of children in our sample from each country group. Nordic countries are excluded.

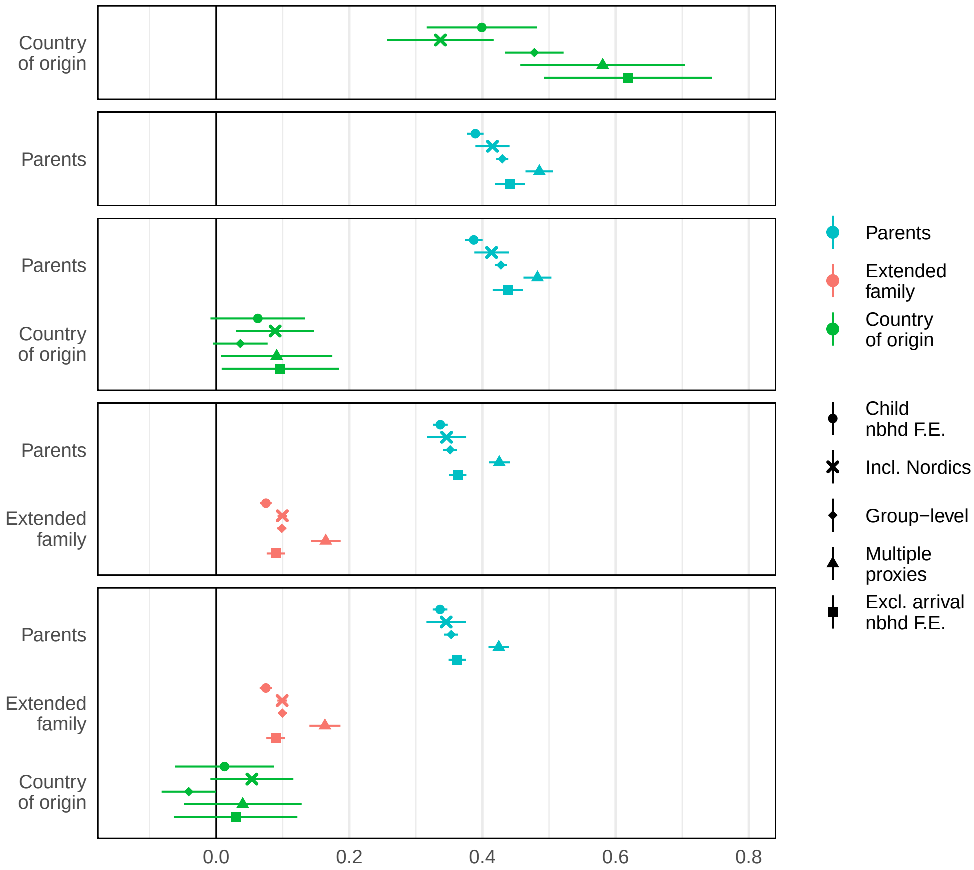

Figure 3: Robustness to alternative specifications

Note: This figure displays results for five alternative specifications. Each subfigure corresponds to one column in Table 1. **Child nbhd. F.E.** displays coefficients where we include neighborhood fixed effects for the children. **Incl Nordics** displays results where we include children of Nordic descent. **Group-level** displays coefficients where we weight each observation by the inverse of the number of individuals in each country group. **Multiple proxies** displays coefficients based on Lubotsky-Wittenberg index constructed based on years of schooling, income and an occupation stratification index. **Exl. arrival nbhd F.E.** displays coefficients where we exclude fixed effects for ancestors' first neighborhood of arrival. All models are estimated using FGLS with random effects and cluster-robust standard errors at the country-of-origin level.

Table 1: GPA schooling regressions

	(1)	(2)	(3)	(4)	(5)
Parents		0.433	0.430	0.362	0.361
		(0.012)	(0.012)	(0.007)	(0.007)
Extended family				0.200	0.199
				(0.016)	(0.016)
Country-of-origin mean	0.540		0.090		0.029
	(0.058)		(0.042)		(0.045)
Sum	0.540	0.433	0.520	0.561	0.589
	(0.058)	(0.012)	(0.041)	(0.018)	(0.043)
R^2	0.085	0.230	0.230	0.241	0.241
Num. Ind	36 305	36 305	36 305	36 305	36 305

Note: Each column shows results from separate specifications where standardised Grade Point Average (GPA) is regressed on years of schooling. The years of schooling variables are divided by the standard deviation of years of schooling for parents. All regressions include fixed effects for birth year, age at migrating to Sweden and first parish of arrival of the ancestors. Each parental generation outcome is the average across all members of the given category of relatives. Models are estimated using FGLS to allow for country group random effects estimated using the method of Amemiya (1971). Standard errors clustered by country group in parentheses.

Table 2: Separate extended family variables by origin

	(1)	(2)	(3)	(4)
Parents	0.360	0.387	0.387	0.363
	(0.008)	(0.010)	(0.008)	(0.008)
Extended family	0.189			
	(0.015)			
Extended family (Natives)		0.122		0.097
		(0.011)		(0.011)
Extended family (Foreign Origin)			0.103	0.081
			(0.012)	(0.012)
Country-of-origin mean	0.041	0.064	0.065	0.040
	(0.044)	(0.044)	(0.042)	(0.045)
R^2	0.235	0.231	0.232	0.235
Num. Ind	39 556	39 556	39 556	39 556

Note: Each column shows results from a separate regression of GPA on years of schooling averages. Column 1 corresponds to column 5 of Table 1, while columns 2–4 show results where we include separate the extended-family measures for relatives born in Sweden (native) versus born outside Sweden (foreign origin). Variable definitions, control variables, and estimation method otherwise corresponds to our main results. See the note to Table 1 for details.

Table 3: Heterogeneous transmission

	Parents	Extended Family	Parents + Extended Family	Obs.
Swedish Origin	0.351	0.203	0.554	53 193
	(0.005)	(0.007)	(0.006)	
Foreign Origin	0.363	0.195	0.559	36 305
	(0.006)	(0.008)	(0.006)	
Nordics	0.405	0.251	0.656	70 433
	(0.005)	(0.007)	(0.006)	
North America/ Western Europe	0.359	0.206	0.565	9 919
	(0.011)	(0.015)	(0.013)	
Southern Europe	0.381	0.182	0.563	9 811
	(0.012)	(0.016)	(0.014)	
Eastern Europe	0.369	0.245	0.614	8 004
	(0.012)	(0.016)	(0.012)	
Northern Europe	0.349	0.120	0.469	4 524
	(0.015)	(0.022)	(0.019)	
MENA	0.343	0.151	0.495	3 144
	(0.019)	(0.031)	(0.027)	
Latin America	0.412	0.127	0.540	507
	(0.047)	(0.074)	(0.070)	
East Asia	0.320	0.284	0.604	300
	(0.055)	(0.075)	(0.063)	

Note: Each row shows results from a separate OLS regression. The dependent variable is standardised GPA. All regressions include controls for birth year fixed effects. Robust standard errors in parentheses.

A Omitted variable bias

The true model is:

$$y_i^c = \gamma_0 + \gamma_1 \bar{y}^p_{f(i)} + \gamma_2 \bar{y}^p_{d(i)} + \gamma_3 \bar{y}^p_{e(i)} + \varepsilon_i, \tag{a.1}$$

Suppose $\bar{y}^p_{d(i)}$ is unobserved and we instead estimate the short regression

$$y_i^c = \gamma_0' + \gamma'_1 \bar{y}^p_{f(i)} + \gamma_3' \bar{y}^p_{e(i)} + u_i. \tag{a.2}$$

Let all variables be expressed in deviations from their sample means and define

$$\begin{aligned}
\sigma_f^2 &= \text{Var}\left(\bar{y}^p_{f(i)}\right), \\
\sigma_e^2 &= \text{Var}\left(\bar{y}^p_{e(i)}\right), \\
\rho_{ef} &= \text{Cov}\left(\bar{y}^p_{e(i)}, \bar{y}^p_{f(i)}\right), \\
\rho_{df} &= \text{Cov}\left(\bar{y}^p_{d(i)}, \bar{y}^p_{f(i)}\right), \\
\rho_{de} &= \text{Cov}\left(\bar{y}^p_{d(i)}, \bar{y}^p_{e(i)}\right),
\end{aligned}$$

and

$$D = \sigma_f^2 \sigma_e^2 - \rho_{ef}^2. \tag{a.3}$$

Then the omitted variable biases can be shown to be

$$\text{plim}\, \hat{\gamma}_1' - \gamma_1 = \gamma_2 \frac{\sigma_e^2 \rho_{df} - \rho_{ef}\rho_{de}}{D}, \tag{a.4}$$

$$\text{plim}\, \hat{\gamma}_3' - \gamma_3 = \gamma_2 \frac{\sigma_f^2 \rho_{de} - \rho_{ef}\rho_{df}}{D}. \tag{a.5}$$

By assumption, $\gamma_2 \geq 0$. The denominator D is always non-negative due to the Cauchy-Schwarz inequality. The signs of the biases then depend only on the numerators.

We start with equation (a.4). Divide the numerator by σ_e^2 to get

$$\rho_{df} - \frac{\rho_{ef}\rho_{de}}{\sigma_e^2} = \rho_{df} - \text{Cov}(\hat{\lambda}_{f|e} \bar{y}^p_{e(i)}, \hat{\lambda}_{d|e} \bar{y}^p_{e(i)}), \tag{a.6}$$

where

$$\hat{\lambda}_{f|e} = \frac{\rho_{ef}}{\sigma_e^2}$$
$$\hat{\lambda}_{d|e} = \frac{\rho_{ed}}{\sigma_e^2}$$

are coefficients from regressing $\bar{y}^p_{f(i)}$ and $\bar{y}^p_{d(i)}$, respectively, on $\bar{y}^p_{e(i)}$.

The second term in equation (a.6) is thus the covariance between the bivariate linear projections of the country-level mean $\bar{y}^p_{e(i)}$ on the parental and extended family means $\bar{y}^p_{f(i)}$ and $\bar{y}^p_{d(i)}$. If the parental and extended family means are *only* related through their respective covariance with country-level means, then this term would fully capture their covariance ρ_{fd}, and the two terms would cancel out. In this scenario, $\hat{\gamma}_1$ is unbiased.

If, as is more plausible, parental and extended family means are directly positively related, we have

$$\rho_{df} > \mathrm{Cov}(\hat{\lambda}_{f|e}\bar{y}^p_{e(i)}, \hat{\lambda}_{d|e}\bar{y}^p_{e(i)}),$$

so that $\hat{\gamma}_1$ is upwards biased.

The analogous argument shows that $\hat{\gamma}_2$ is upwards biased if the extended family and the country mean covary positively beyond the link through their respective association with the parental mean.

B Latent Variables Model

We here specify the latent variables model, which we estimate in section 4.2. Specifically, let the observed child outcome be determined as

$$y_i^c = \beta_0 + \beta_1 y_{f(i)}^{p*} + \beta_2 y_{d(i)}^{p*} + \beta_3 y_{e(i)}^{p*} + \varepsilon, \tag{b.1}$$

where $y_{f(i)}^{p*}$ is the parents' latent human capital, $y_{d(i)}^{p*}$ latent human capital of the extended family, and $y_{e(i)}^{p*}$ latent human of the ethnic group. While these latent variables are fundamentally unobservable, we observe a set of proxy variables $y_j^p = y_j^{p*} + v_j$, for $j \in \{f, d, e\}$, where v^j is an i.i.d. measurement error term. We can view equation (3) above as an empirical version of this model, where we use average level of education among parents, the extended family, and the ethnic group as proxies for the corresponding latent variables.

If we have access to multiple proxies, y_{jk}^p, for y_j^{p*}, we can improve our estimates by using the approach proposed by Lubotsky and Wittenberg (2006).[27] This approach proceeds in three steps: first, we estimate an extended model by regressing child outcome on the full set of proxies,

$$y_i^c = \sum_j \sum_k b_{jk} y_{ijk}^p + \varepsilon_i. \tag{b.2}$$

We then calculate weighted averages of the coefficients from equation (b.2) as

$$\hat{\beta}_j = \sum_k \rho_{jk} \hat{b}_{jk}, \tag{b.3}$$

where the weights are functions of covariances $\rho_{jk} = \frac{\mathrm{Cov}(y^c, y_{jk}^p)}{\mathrm{Cov}(y^c, y_{j1}^p)}$. Notice that the weights are normalized to the scale of one of the proxies (y_{j1}^p)—we normalize to years of schooling, so that the Lutbotsky-Wittenberg (LW henceforth) coefficients can be interpreted on the same scale as our main years of schooling regressions.[28]

The additional proxies we use are *lifetime income*, including unemployment insurance and sickness benefits[29]; and a *social stratification index* based on the Swedish version

[27]For each latent variable j, the measurement errors for each proxy k are assumed to be uncorrelated with the latent variable, but are allowed to be correlated with each other.

[28]The Lubotsky and Wittenberg estimator has previously been used to improve estimates of intergenerational mobility by, e.g., Adermon et al. (2021), Vosters (2018), and Vosters and Nybom (2017).

[29]We first regress annual log income on a full set of gender, birth-year, and calendar-year controls using the full income panel observed between ages 35 and 55. Lifetime income is then defined as the individual-level average of the residuals from this regression.

of the occupation-based CAMSIS index (Lambert and Bihagen, 2012).[30] We assign the index using each individual's occupation at, or as close as possible to, age 50.

This means that our resulting estimates of the relationship between child's GPA and parents', extended families', and ethnic groups' "human capital" (or social status) should be interpreted as capturing broader channels not captured by years of schooling, including employers' valuation of human capital (including discrimination), individuals' occupational choices, and job networks.

[30] CAMSIS measures social distance based on the occupations of married couples.

C Additional Tables and Figures

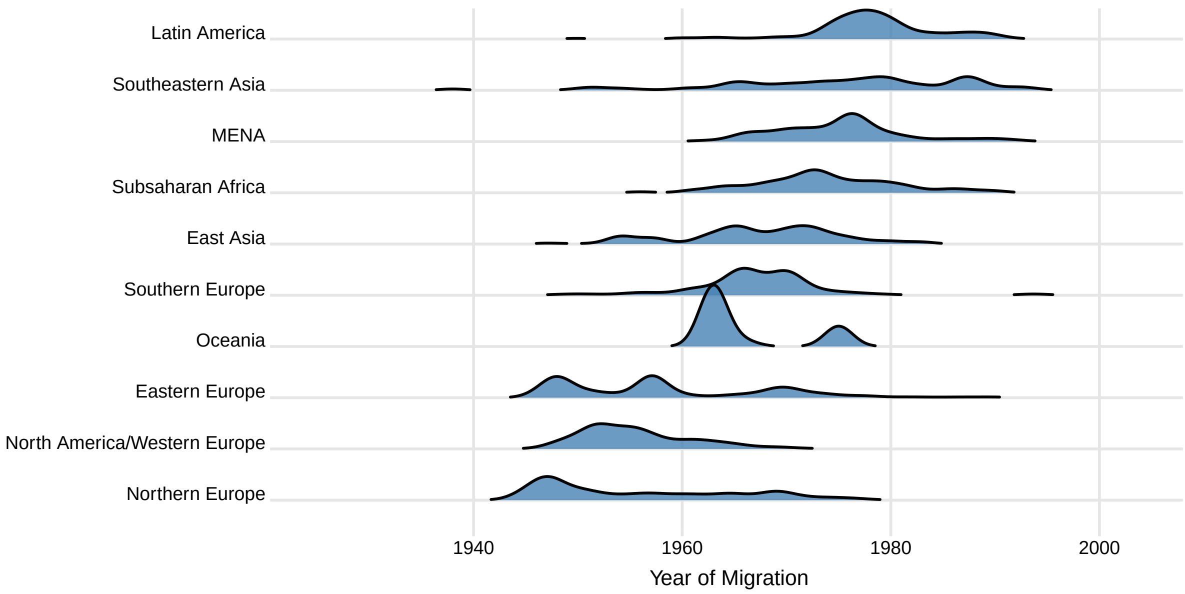

Figure C.1: Year of immigration to Sweden for the grandparent generation.

Note: This figure shows the distribution of first arrival years to Sweden for the grandparents of individuals in our sample split up by broad contry groups.

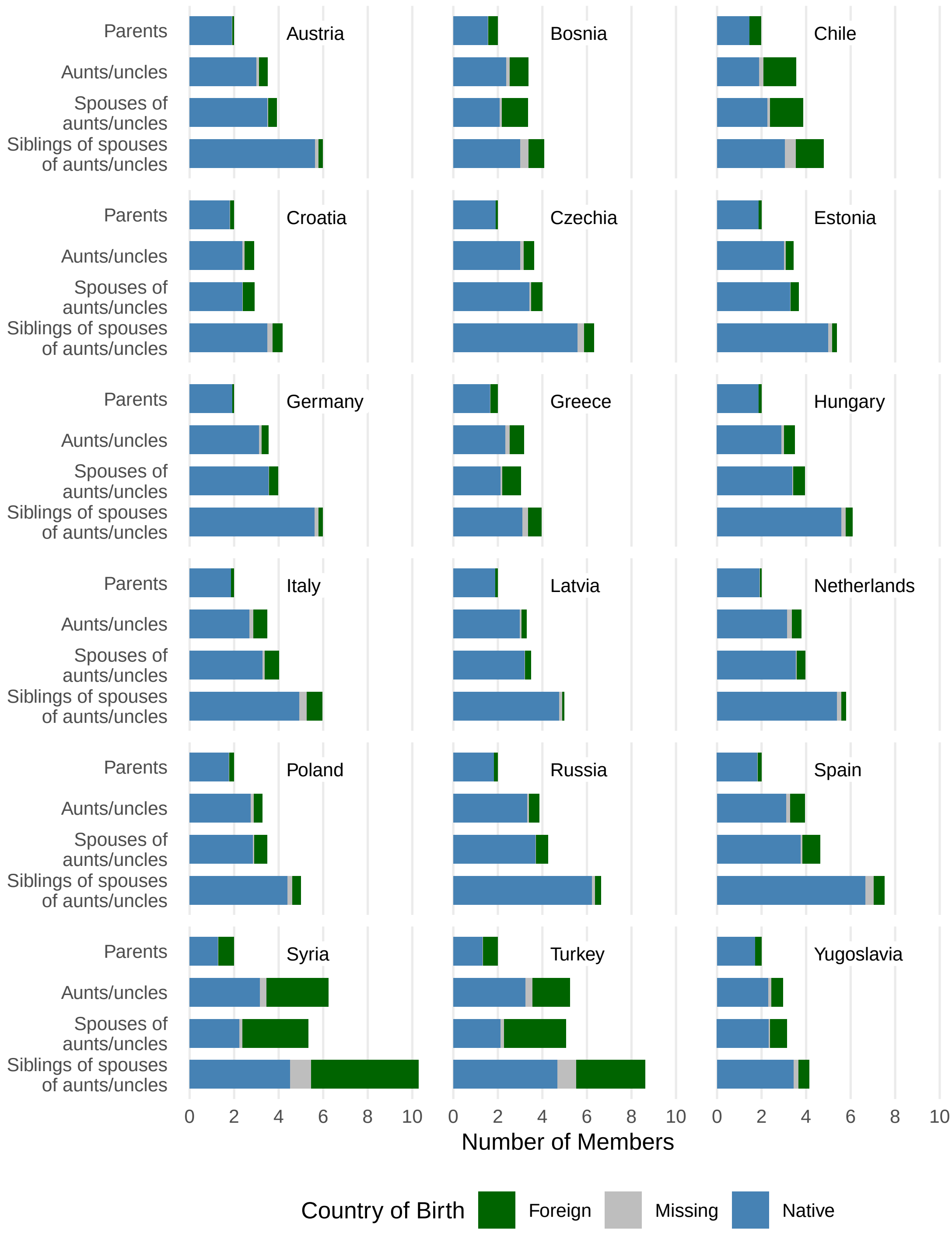

Figure C.2: Composition of native and foreign-born individuals per child.

Note: Green bars indicate the average number of foreign-born individuals (with observed schooling) per child; blue bars indicate native-born individuals; gray bars represent foreign-born individuals for whom schooling data is missing.

Table C.1: Number of observations by country

Countrygroup	N. Ind.	N. Origin mean	Countrygroup	N. Ind.	N. Origin mean
Germany	8 900	3 912	Chile	392	2 756
Yugoslavia	6 601	6 172	Romania	379	866
Estonia	3 979	1 920	Switzerland	310	893
Hungary	3 257	1 670	Bosnia	304	1 667
Poland	2 488	1 757	Syria	302	1 310
Turkey	2 480	4 004	United Kingdom	289	782
Czechia	1 973	1 296	North Macedonia	274	2 060
Austria	1 883	1 486	Lebanon	254	1 350
Russia	1 824	1 177	France	243	975
Greece	1 413	2 468	Portugal	213	481
Italy	1 149	1 101	Slovenia	203	1 785
Croatia	899	2 342	Morocco	178	596
Latvia	832	670	South America, rest	168	1 271
Netherlands	763	904	Palestine	129	387
Spain	606	713	Bulgaria	101	1 669
USA	401	1 021	Others (50 regions)	1 455	925

Note: N. ind reports the number of children in the main sample by countrygroup. N. origin mean reports the average number of individuals used to compute the country-of-origin mean for each parent of children in the sample. The table reports the number of child-country observations by country which is slightly higher than the number of individuals in the sample. Category "Others" include the following countries: Lithuania, Ireland, Belgium, Slovakia, Western Europe (rest), Ukraine, Eastern Europe (rest), Serbia, Albania, Canada, Central Asia (rest), Jordan, Iraq, Western Asia (rest), Iran, Afghanistan, Pakistan, India, Bangladesh, Sri Lanka, Tunisia, Egypt, Sudan, North Africa (rest), Nigeria, Gambia, West Africa (rest), Congo, Eritrea, Ethiopia, Somalia, Uganda, Kenya, East Africa (rest), Southern Africa, Central America, Colombia, Brazil, Peru, Bolivia, China, South Korea, Japan, East Asia (rest), Vietnam, Thailand, Philippines, South east Asia (rest), Oceania, West Indies.

Table C.2: Alternative Ethnic Capital Proxies

	(1)	(2)	(3)	(4)	(5)
Parents	0.367	0.367	0.367	0.367	0.367
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Extended family	0.197	0.197	0.197	0.197	0.197
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Country-of-origin mean	0.024	0.027	0.026	0.015	0.018
	(0.078)	(0.077)	(0.077)	(0.079)	(0.078)
WVS education index		0.006			0.030
		(0.021)			(0.042)
WVS culture index			0.008		0.010
			(0.016)		(0.021)
Barro Lee Years of schooling				0.015	0.036
				(0.019)	(0.028)
R^2	0.253	0.253	0.253	0.253	0.254
Num. Obs	19 698	19 698	19 698	19 698	19 698

Note: Each column shows results from separate regressions. The dependent variable is standardised GPA in both panels. WVS education index, WVS culture index and Barro and Lee (2013) years of schooling are standardised. Models are estimated using FGLS to allow for country group random effects estimated using the method of Nerlove (1971). Standard errors clustered by country group in parentheses.

Table C.3: GPA schooling regressions, OLS

	(1)	(2)	(3)	(4)	(5)
Parents		0.436	0.432	0.365	0.365
		(0.009)	(0.011)	(0.007)	(0.007)
Extended family				0.196	0.196
				(0.013)	(0.014)
Country-of-origin mean	0.631		0.089		0.014
	(0.063)		(0.055)		(0.056)
Sum	0.631	0.436	0.521	0.561	0.574
	(0.063)	(0.009)	(0.049)	(0.015)	(0.048)
R^2	0.096	0.236	0.237	0.247	0.247
Num. Ind.	36 305	36 305	36 305	36 305	36 305

Note: Each column shows results from separate specifications where standardised Grade Point Average (GPA) is regressed on years of schooling. The years of schooling variables are divided by the standard deviation of years of schooling for parents. All regressions include fixed effects for birth year, age at migrating to Sweden and first parish of arrival of the ancestors. Each parental generation outcome is the average across all members of the given category of relatives. Standard errors are clustered by country group in parentheses.

Table C.4: Alternative specifications

	(1)	(2)	(3)	(4)	(5)
	Child nbhd F.E.				
Parents		0.389	0.387	0.337	0.336
		(0.006)	(0.007)	(0.006)	(0.006)
Extended family				0.075	0.075
				(0.004)	(0.005)
Country-of-origin mean	0.399		0.063		0.013
	(0.042)		(0.036)		(0.038)
	Excl. arrival nbhd F.E.				
Parents		0.441	0.438	0.363	0.362
		(0.012)	(0.012)	(0.007)	(0.007)
Extended family				0.090	0.089
				(0.007)	(0.007)
Country-of-origin mean	0.618		0.096		0.029
	(0.064)		(0.045)		(0.047)
	Group-level				
Parents		0.430	0.428	0.352	0.353
		(0.005)	(0.005)	(0.005)	(0.005)
Extended family				0.099	0.099
				(0.003)	(0.003)
Country-of-origin mean	0.478		0.036		-0.041
	(0.022)		(0.021)		(0.021)
	Incl. Nordics				
Parents		0.415	0.414	0.346	0.346
		(0.013)	(0.013)	(0.015)	(0.015)
Extended family				0.099	0.099
				(0.004)	(0.004)
Country-of-origin mean	0.337		0.089		0.054
	(0.041)		(0.030)		(0.032)
	Multiple proxies				
Parents		0.485	0.483	0.425	0.425
		(0.011)	(0.011)	(0.008)	(0.008)
Extended family				0.165	0.163
				(0.011)	(0.012)
Country-of-origin mean	0.581		0.091		0.040
	(0.063)		(0.043)		(0.045)

Note: This table reports the coefficient point estimates and corresponding standard errors underlying the point estimates and confidence intervals in Figure 3. See the note to that figure for a description of the robustness checks. All models are estimated using FGLS with random effects and cluster-robust standard errors at the country-of-origin level.

Table C.5: Regressions by gender

	(1)	(2)	(3)	(4)	(5)
Panel A: Men					
Parents		0.448	0.443	0.376	0.375
		(0.012)	(0.013)	(0.010)	(0.010)
Extended family				0.198	0.196
				(0.018)	(0.018)
Country-of-origin mean	0.610		0.150		0.086
	(0.071)		(0.063)		(0.064)
Sum	0.610	0.448	0.593	0.574	0.656
	(0.071)	(0.012)	(0.060)	(0.018)	(0.058)
Num. Ind	18 722	18 722	18 722	18 722	18 722
R^2	0.108	0.260	0.260	0.271	0.271
Panel B: Women					
Parents		0.433	0.432	0.358	0.358
		(0.015)	(0.015)	(0.011)	(0.011)
Extended family				0.207	0.208
				(0.018)	(0.018)
Country-of-origin mean	0.516		0.037		-0.026
	(0.055)		(0.042)		(0.044)
Sum	0.516	0.433	0.469	0.565	0.540
	(0.055)	(0.015)	(0.041)	(0.022)	(0.045)
Num. Ind	17 583	17 583	17 583	17 583	17 583
R^2	0.115	0.256	0.256	0.269	0.269

Notes: Each column shows results from a separate regression of GPA on years of schooling averages. Panel A shows results for men in the child generation, while Panel B shows results for women. Variable definitions, control variables, and estimation method otherwise corresponds to our main results. See the note to Table 1 for details.