IARIW 2025

IARIW 2025

Thursday, October 2 & Friday, October 3

Sibling Correlations in Schooling around the World: A New Database

Ercio Munoz

(Inter-American Development Bank)

Gabriela Jaque

(University of Chile)

erciom@iadb.org

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on "Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies" October 2-3, 2025

Session 4A: Income Measurement in Developing Countries
Time Slot: Friday, October 3, 8:00-9:30 AM

Sibling Correlations in Schooling Around the World: A New Database*

Ercio Muñoz† Gabriela Jaque‡

August 6, 2025

Abstract

We estimate sibling correlations in schooling (i.e., the fraction of inequality in educational outcomes attributable to factors shared by siblings) for 128 countries accounting for 94% of the world's population. With this new database, we document several findings. On average, at least 56% of the inequality in schooling can be attributed to factors shared by siblings. There are significant regional differences, with Europe and Central Asia showing the lowest and South Asia the highest levels of sibling correlation. There is also substantial heterogeneity within some regions. The average sibling correlation has been decreasing across cohorts; in some regions, however, it has stagnated. At the global level, educational mobility appears strongly correlated with several social and economic variables.

JEL Codes: D63, I24, J62.
Keywords: Sibling correlations, Intergenerational mobility, Education, Family background.

*We thank the participants at the ECINEQ LAC conference in Mexico and the Eleventh ECINEQ in Washington DC for providing comments and suggestions. Any views expressed in this paper are our own and should not be attributed to the Inter-American Development Bank.

†Corresponding author. Social Sector, Inter-American Development Bank. Email: erciom@iadb.org.
‡Universidad de Chile. Email: gjaqueh@fen.uchile.cl.

I Introduction

How much of the variation in schooling around the globe is due to family background? Social scientists have long been interested in understanding how family background shapes individual socioeconomic outcomes, including education. Economists have made important progress in particular in documenting the intergenerational relationship between parent and offspring outcomes (e.g., Van der Weide et al. 2024, Munoz & Van der Weide 2025). However, evidence on broader measures of family impact remains scarce.[1] This paper aims to fill this gap by estimating the level of sibling similarity in educational outcomes (i.e., the fraction of total inequality in educational attainment attributable to factors shared by siblings, such as family background) across a large group of developed and developing countries.

We use a large number of census and survey samples spanning 128 countries to construct a new database of cross-generational educational mobility with coverage of 94% of the global population. We use household rosters to identify cohabiting siblings aged 21 to 30 years and estimate sibling correlations in educational outcomes. Using this database, we document patterns across countries and over time and explore the association between this measure of intergenerational mobility and a rich set of social and economic variables.

We document several empirical findings. First, we find that, on average, at least 56% of the inequality in schooling can be attributed to shared sibling background. Second, we report mobility by region, showing that, on average, countries in Europe and Central Asia have the lowest sibling correlations, while the highest average is seen in countries in South Asia. Third, there is also important heterogeneity within some regions of the developing world. For example, some regions include both countries among the world's twenty *most* mobile and countries among the world's twenty *least* mobile. Fourth, we show that mobility in education (as measured by sibling correlations) has been increasing globally but has stagnated in some regions. Fifth, we show that the measure of educational mobility employed in this paper is

[1]See Table A1 for an overview of the published estimates of sibling correlations in schooling, which span only a handful of high-income countries.

strongly correlated with several variables related to the economy, education, health, labor market, demography, infrastructure, and governance.

This paper makes several contributions to the literature. First, it expands the cross-national empirical evidence on educational mobility across generations (e.g., Van der Weide et al. 2024) by estimating an indicator that goes beyond the parent–offspring association or the role of a given set of circumstances. This can help connect the evidence gathered through different approaches in the study of how individual outcomes are connected to family background (for a review of related approaches, see Björklund & Jäntti 2020). Second, it significantly improves on the coverage of previous studies on sibling correlations by accounting for 94% of the global population, encompassing both developed and developing countries (see Table A1 for an overview of existing estimates).[2] Third, the paper presents a novel set of stylized facts on global intergenerational mobility by cohort and region. Last, it examines how sibling similarity is associated with a diverse range of social and economic variables at the country level.

The remainder of the paper is organized as follows. Section II outlines the data and methods employed in this paper. Section III describes the main patterns we see in the sibling correlation estimates, and Section IV explores the statistical association between educational mobility and a set of variables previously used in the literature. Finally, Section V offers some concluding remarks.

II Data and Methods

Data. To construct a comprehensive database for the estimation of global sibling correlations, we combine and harmonize data from multiple sources: IPUMS International, the Luxembourg Income Study (LIS), the World Bank's Poverty and Inequality Platform (PIP), and several individual household surveys and censuses obtained directly from national statistics

[2] Concurrent work in Ahsan et al. (2023) offers estimates for 53 developing countries using Demographic and Health Surveys.

offices. These sources provide nationally representative microdata, organized by household, that contain information on the educational attainment and demographic characteristics of the population. We use data from a total of 128 countries, which ensures comprehensive geographic representation.

For ninety-four countries, we use population and housing census data obtained from IPUMS International (Ruggles et al. 2024) and hosted at the University of Minnesota Population Center, which reports harmonized representative samples (usually 10%) of full-count census microdata for a large number of economies.[3] These censuses are conducted with the goal of obtaining updated social and demographic information about the entire population and their housing conditions.

For nineteen countries, we use household survey data from the LIS, which compiles and harmonizes socioeconomic microdata from mostly middle- and high-income nations and provides remote online access. Similarly, data for three additional countries are sourced from the World Bank's PIP, which also provides online access to anonymized microdata.[4]

Last, for three additional countries, we use full-count microdata from their population and housing censuses, which we obtain directly from their respective National Statistics Offices (NSOs). For the remaining nine countries, we rely on household and labor force surveys obtained from the countries' NSO websites.

For each country, we estimate sibling correlations for up to seven cohorts based on the available data, with each cohort representing a specific period. For the vast majority of countries, we use a single sample to represent a single cohort; however, in selected cases for which the samples are small (e.g., with fewer than 500 observations), we pool adjacent waves to increase the sample size. The birth cohorts span the 1960s to the 2020s and approximately correspond to the decennial census rounds, depending on the availability of data, which allows us to analyze sibling correlations across different periods. Table A2 outlines the source of

[3]Census data have been used to study intergenerational educational mobility in, for example, Alesina et al. (2021, 2023), Card et al. (2022), Derenoncourt (2022), and Munoz (2024).

[4]The platform is available at https://pip.worldbank.org/home.

information for each country and the specific years used for each cohort studied. This provides a clear overview of the data sources and periods covered in our analysis.

Educational attainment. There are two items about educational attainment in the harmonized dataset obtained from IPUMS, the main data source. The first reports the total *years* of schooling completed by each individual (formal schooling regardless of the track or kind of study), and the second is recoded by IPUMS to capture educational attainment in terms of the *level* of schooling completed,[5] with four categories: (1) less than primary completed, (2) primary completed, (3) secondary completed, and (4) university completed. We use years of schooling for all countries for which this indicator is available. For countries with only schooling levels available, we impute years based on the average years of schooling computed for each level of education, using samples for which both variables are available. These averages are then scaled by the average ratio between estimates with both alternatives, also computed from the samples for which both variables are available.[6]

For the remaining data sources, we use years of schooling when this indicator is available; otherwise, we impute the average years from the corresponding levels on a case-by-case basis. In particular, the PIP data provide harmonized educational categories that are compatible with IPUMS, allowing us to impute average years of schooling as we do with the IPUMS categorical data. Similarly, the full-count census data for Belize provide education categories compatible with IPUMS, and we assign average years accordingly. Finally, the full-count census data for Barbados report years of schooling in intervals, and we use the value of the midpoint in each interval.[7]

Country coverage. The dataset includes a total of 128 countries, representing 94% of the global population. For most regions of the world, the countries represent more than 68%

[5]Note that this variable does not necessarily reflect any particular country's definition of the various levels of schooling in terms of terminology or number of years of schooling. The variable applies, to the extent possible, the United Nations standard of six years of primary schooling, three years of lower secondary schooling, and three years of upper secondary schooling.

[6]Figure A1 shows that for countries with both variables, the estimates with years of schooling imputed from categories align closely with those obtained with years of schooling as originally reported.

[7]See Table A2 for details on which countries have data on years of schooling.

of their population (see Table 1). North America and Latin America and the Caribbean (LAC) are the regions with the highest population coverage. For developing countries, coverage reaches 94%, and for high-income countries, 95% (see Table A3). To the best of our knowledge, this database provides the largest coverage in the literature on sibling correlations in terms of number of countries and population accounted for.

Table 1: Country coverage

Region	Number of countries	Population share, %
East Asia & Pacific	17	98.01
Europe & Central Asia	35	85.26
Latin America & Caribbean	28	99.77
Middle East & North Africa	9	67.99
North America	2	99.98
South Asia	6	98.53
Sub-Saharan Africa	31	89.16
World	128	94.05

Notes: The table shows the number of countries covered by the database and the population share that they account for. Table A3 shows the country coverage by income level.

Identification of siblings. Data collection is organized at the household level in all samples used, so it is possible to link individuals who live in the same household at the time of the interview through the variable that details the relationship between each individual and the household head (or person of reference). Using this variable, we identify individuals likely to be siblings. For example, individuals classified as children of the household head are listed as siblings (the appendix provides identification details). We restrict the sample of siblings to individuals aged between 21 and 30 years, with our aim being to choose an age range that maximizes the likelihood that respondents have completed their education and during which their probability of cohabitation remains relatively high. A potential concern that may emerge from this sample restriction is that our estimates could be biased by coresidence bias; however, recent work provides evidence that this type of bias is small for measures such as sibling correlations (Ahsan et al. 2025).[8]

Measurement. Following previous literature (see, for example, Grätz et al. 2021), we estimate sibling correlations in educational attainment with the expression:

$$y_{ij} = \beta' X_{ij} + \epsilon_{ij} \tag{1}$$

where y_{ij} denotes the measure of educational attainment (e.g., years of schooling) for the j^{th} sibling in family i, X_{ij} is a vector of exogenous variables that account for the life cycle (e.g., a cubic in age) with β as the associated vector of coefficients,[9] and ϵ_{ij} is a disturbance term that represents educational attainment net of life-cycle effects.

The disturbance term has two components:

$$\epsilon_{ij} = \alpha_i + \mu_{ij} \tag{2}$$

where α_i is a permanent component common to all siblings in family i and μ_{ij} is an idiosyncratic individual component. We assume that these two components are orthogonal, which implies that the variance of the disturbance term can be expressed as:

$$\sigma_\epsilon^2 = \sigma_\alpha^2 + \sigma_\mu^2 \tag{3}$$

Hence, the correlation of educational attainment among siblings is

$$\rho = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\mu^2) \tag{4}$$

which is the proportion of the population variance in educational attainment that is due to

[8] Munoz & Siravegna (2023) show that the bias due to the cohabitation restriction is also small in the context of upward mobility and that rankings derived from other mobility indicators such as the intergenerational regression coefficient on the basis only of coresident samples are reliable.

[9] Our baseline estimates do not include age, and they barely change when age is included (see Figure A2).

factors shared by siblings.[10]

III Sibling Correlations Around the World

In this section, we document patterns observed in the estimated sibling correlations. First, we analyze patterns across geographical regions of the world. Second, we explore the variation in educational mobility between countries. Finally, we analyze patterns for different birth cohorts based on the year in which the data were collected.

Evidence across regions. As described in our data section, the coverage by region varies from nearly 68% to almost 100% (see Table 1). The lowest coverage is for Europe and Central Asia and the Middle East and North Africa (MENA) (85% and 68% of their population, respectively). Nonetheless, we aggregate the available information and compare the evidence on the importance of shared background across these regions.

Table 2 summarizes our main results. Given the heterogeneity in the number of samples available for each country and the year in which data collection occurred, we aggregate the information in three different ways: keeping all cohorts, keeping only the first cohort of each country, and keeping only the last cohort of each country. We compute simple averages of country cohorts (i.e., data samples) within each region to reflect the level of mobility in the average country in a given region. We find that South Asia shows the highest sibling correlations (lowest educational mobility). Conversely, we see the lowest sibling correlations (highest educational mobility) for Europe and Central Asia. North America and MENA also show higher mobility levels than other regions, while East Asia and the Pacific show lower mobility. Interestingly, sub-Saharan Africa and LAC fall in the middle of the seven regions when we consider all cohorts or only the last cohort but show very low mobility when we consider only the first cohort, which suggests that mobility levels have been changing across

[10] We estimate this correlation in Stata using the command `mixed` for estimation of linear mixed-effects models by means of restricted maximum likelihood. Estimates that account for the censored nature of years of schooling would yield larger values with similar rankings (see Figure A3).

cohorts in this region.[11] When we group the countries by income level (see Table A5), the sibling correlations are, on average, lower for high-income than for developing countries (0.46 vs. 0.61 when the average cohort is considered).

Table 2: Sibling correlation by region

Region	Average	First cohort	Last cohort	Countries
Europe & Central Asia	0.45	0.44	0.43	35
North America	0.49	0.52	0.39	2
Middle East & North Africa	0.52	0.52	0.52	9
East Asia & Pacific	0.55	0.54	0.53	17
Sub-Saharan Africa	0.62	0.62	0.62	31
Latin America & Caribbean	0.62	0.64	0.57	28
South Asia	0.64	0.60	0.60	6
World	0.56	0.55	0.53	128

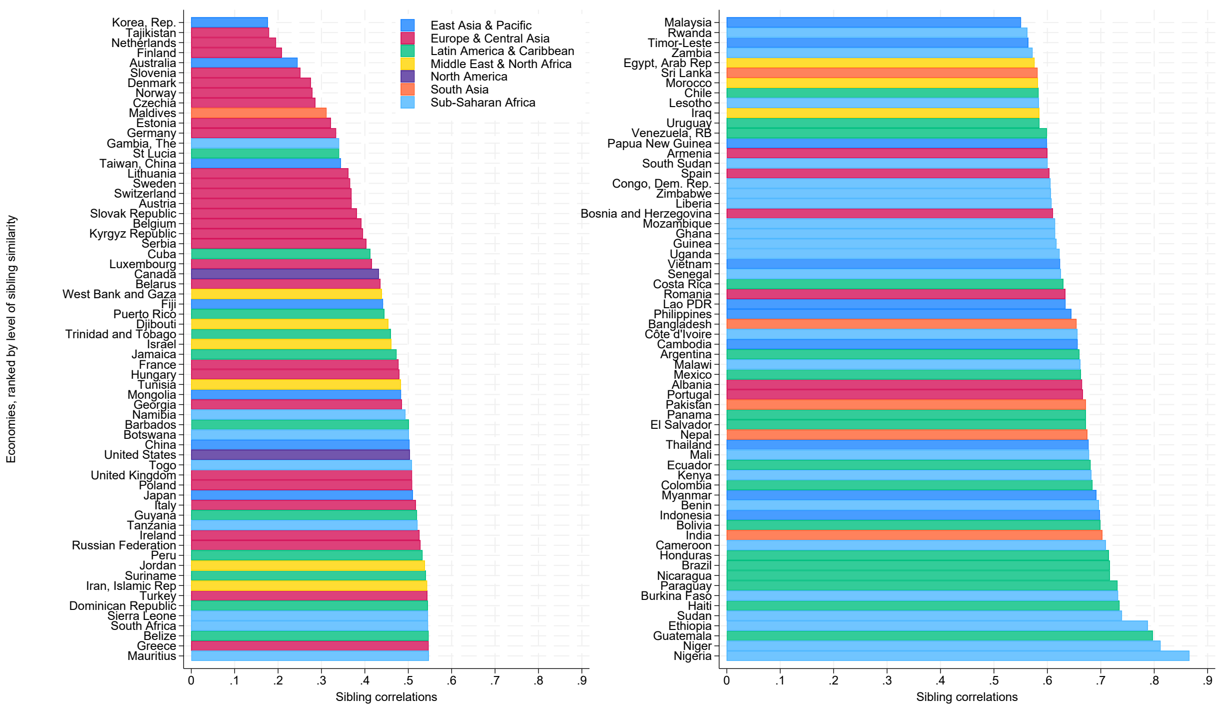
Notes: The table reports simple averages of our estimates for each region. The columns "Average," "First cohort," and "Last cohort," respectively, use all the birth cohorts available for each country, only the first cohort available for each country, and only the latest birth cohort available for each country. The column "Countries" reports the number of countries with data for each region. Rows are sorted according to the "Average" column. Table A5 provides these statistics by income level.

Evidence across countries. Figure 1 presents an overview of the level of sibling correlations in schooling across the world using the most recent cohort for each country. The figure ranks countries in ascending order by sibling correlation, dividing them into two groups (bottom half and upper half). Consistent with the findings in the previous section (see Table 2), North American countries fall predominantly in the lower half of the distribution, indicating high mobility, while most South Asian countries fall in the upper half, reflecting lower mobility. An exception within South Asia is the Maldives, which ranks among the ten most mobile countries. Other regions display a wider range of outcomes. For instance, although most countries in Europe and Central Asia are highly mobile, a few, such as Albania and Portugal, fall into the less mobile group. In LAC and sub-Saharan Africa, most countries fall among the least mobile half, yet notable exceptions such as Gambia and Saint

[11] The results are qualitatively similar if we restrict the samples to only those with years of schooling (see Table A4).

Lucia appear in the most mobile half, underscoring the considerable variation within these regions. MENA countries tend to cluster around the middle of the distribution, leaning slightly toward higher mobility. Last, East Asia and the Pacific is the most diverse region in terms of mobility: While countries such as Myanmar and Indonesia are among the least mobile, others such as Australia and South Korea (which ranks first) are among the most mobile in the world.

Figure 1: Sibling correlations for countries around the world (from most mobile to least mobile)



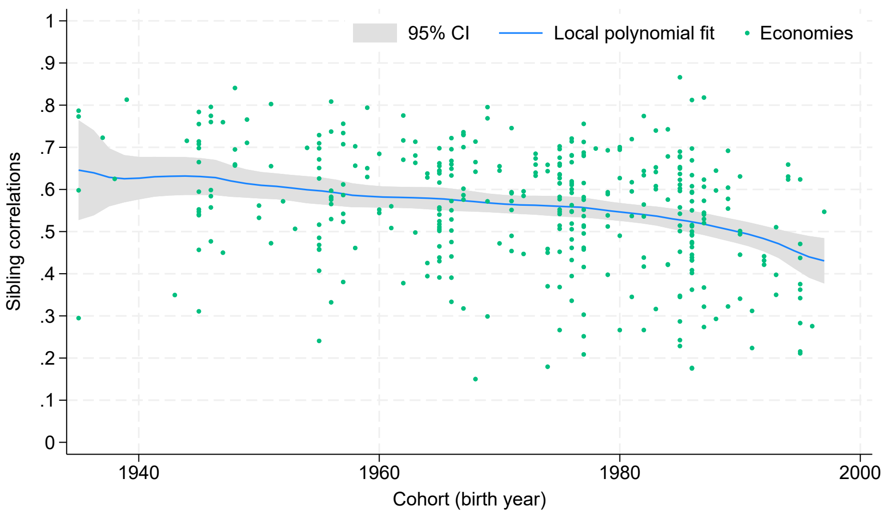
Notes: The figure presents simple averages of our estimates for each country, using all available birth cohorts.

Sibling correlations over time. Given the time coverage of our dataset, we document how the level of mobility has changed across birth cohorts at the global level. Figure 2 shows that the average level of mobility has been increasing across cohorts (i.e., the average level of sibling correlations has decreased across cohorts).[12] This pattern characterizes both

[12]Figures A4 and A5 show that this pattern holds when we restrict the sample to economies with at least

high-income and developing economies (see Appendix Figure A7). However, it does not hold for all regions, as shown in Figure A8: While LAC and North America exhibit a clear downward trend in their average level of sibling correlations, other regions, including MENA and sub-Saharan Africa, do not show a clear trend across cohorts.

Figure 2: Sibling correlation in schooling over time



Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection.

IV Correlates of Sibling Similarity

We use our newly created database to investigate the relationship between our estimates of sibling similarity in schooling and a set of correlates examined in the literature on intergenerational mobility. In particular, we replicate the choice of correlates in Van der Weide et al. (2024), which examines the association between these and several indicators of intergenerational mobility.[13] We contrast our findings with those reported for the intergenerational

four estimates or to those with data on years of schooling. Moreover, the trend is not driven by an increase in levels (i.e., a larger share of people with the maximum level of schooling in our sample), as shown in Figure A6, which plots the trend over time obtained with a tobit model.

[13]This selection is guided by data availability and previous literature (see Alesina et al. 2021, Asher et al. 2024, Chetty et al. 2014).

regression coefficient, keeping the discussion short given that Appendix D in Van der Weide et al. (2024) details how these correlates may be associated with intergenerational mobility in education.[14] Figure 3 reports the coefficients obtained from regressing our estimates of sibling similarity on a wide set of correlates related to the economy, education, health, labor market, demography, infrastructure, and governance.[15]

We find that higher GDP per capita, government revenue, expenditure, education expenditure, and income equality are all negatively associated with sibling similarity. All the education and health indicators considered show a negative association with the sibling correlations. All of these associations are statistically significant, except for the proportion of the population without HIV. Among labor market variables, employment shares across different sectors are associated with the level of sibling similarity, whereas labor force participation is not. In terms of demographic factors, we find that population size and its growth are positively associated with sibling similarity whereas net migration and the international migrant stock are negatively associated. Indicators of gender equality, such as early marriage and early pregnancy, are negatively correlated with sibling similarity. Both infrastructure indicators are negatively correlated with sibling similarity. Finally, we find significant associations between sibling similarity and governance indicators.[16]

Most of the correlations are consistent with the findings documented in Van der Weide et al. (2024) for intergenerational mobility in education estimated by means of a regression between parents' and children's schooling. However, there are a few differences worth highlighting. We find significant associations with the proportion of children not wasted, the number of years of compulsory schooling, and the log of population. In contrast, we do not find statistically significant associations with labor force participation. Last, we see

[14]Figure A9 compares our estimates with those in Van der Weide et al. (2024). Although the estimates are positively associated, several countries appear to be more or less mobile when we consider their sibling correlations.

[15]We omit the variable "less than 25 battle deaths" because only 32 countries in our database have values for this indicator and most of them are zeroes.

[16]The results are qualitatively similar when we restrict to the samples with years of schooling (see Figure A10).

a statistically significant association with the opposite sign only for religious homogeneity, although the relationship is quite weak (Figure A11).[17]

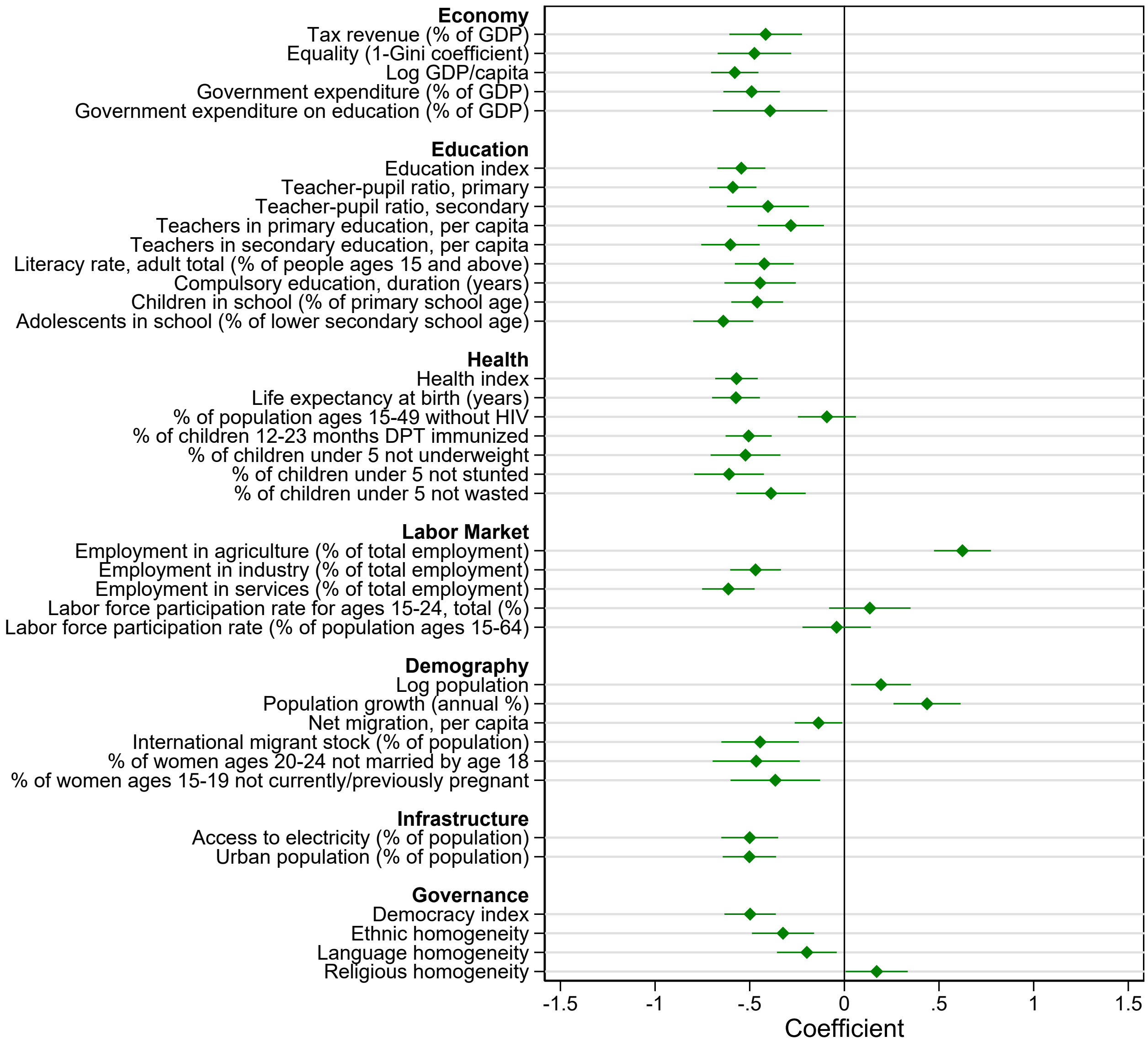
V Concluding Remarks

In this paper, we create and examine a new database of sibling correlations in educational outcomes that spans 128 developed and developing countries. This database significantly expands the cross-national empirical evidence on sibling similarity and complements recent estimates of intergenerational mobility at the global level. Our estimates indicate that at least 56% of the inequality in schooling can be attributed to factors shared by siblings. Countries in Europe and Central Asia have the lowest sibling correlations, while countries in South Asia have the highest. However, there is substantial heterogeneity within some regions of the developing world. We also find that average sibling correlations in schooling have been decreasing globally, but not for all regions. Finally, we show that educational mobility is strongly correlated with several variables related to the economy, education, health, labor market, demography, infrastructure, and governance, in line with recent evidence on other indicators of intergenerational mobility.

Future research could expand this work in several directions. First, the data used in this paper could be used to explore how the patterns of sibling similarity in schooling at the global level vary by sibling composition (e.g., all-male sibships, all-female sibships and mixed-gender sibships). Second, census data often include information about religion and ethnicity. Hence, researchers could explore how sibling correlation patterns vary with these characteristics (see, for example, Alesina et al. 2023).

[17] A comparison of the correlation coefficients would be qualitatively similar but would yield more discrepancies in terms of statistical significance, as Van der Weide et al. (2024) find nonsignificant relationships for many more indicators.

Figure 3: Correlates of sibling similarity



Notes: The figure plots the coefficients from a regression of our measure of sibling correlations (standardized to have mean 0 and variance 1) on each covariate (standardized to have mean 0 and variance 1), including 95% confidence intervals with standard errors clustered by country. We fill data gaps for the covariates using interpolation. The confidence intervals do not account for the uncertainty in the estimate of sibling correlation. Each sample is matched with a correlate at its value 10 years earlier, such that a given cohort aged 21–30 years is associated with the value of the correlates when the cohort members were around primary or secondary school age. Sudan is excluded from the tax revenue correlation, as the corresponding value appears implausibly high.

References

Ahsan, M. N., Emran, M. S., Jiang, H., Han, Q. & Shilpi, F. (2023), 'Growing up together: Sibling correlation, parental influence, and intergenerational educational mobility in developing countries', *Policy Research working paper 10285* .

Ahsan, M. N., Emran, M. S., Jiang, H. & Shilpi, F. (2025), 'Making the most of coresident data: Credible evidence on intergenerational mobility with sibling correlation', *Journal of Development Economics* **176**.

Alesina, A., Hohmann, S., Michalopoulos, S. & Papaioannou, E. (2021), 'Intergenerational Mobility in Africa', *Econometrica* **89**(1), 1–35.

Alesina, A., Hohmann, S., Michalopoulos, S. & Papaioannou, E. (2023), 'Religion and educational mobility in africa', *Nature* **618**(7963), 134–143.

Asher, S., Novosad, P. & Rafkin, C. (2024), 'Intergenerational Mobility in India: New Measures and Estimates Across Time and Social Groups', *American Economic Journal: Applied Economics* **16**, 66–98.

Björklund, A. & Jäntti, M. (2012), 'How Important is Family Background for Labor-economic Outcomes?', *Labour Economics* **19**(4), 465–474.

Björklund, A. & Jäntti, M. (2020), 'Intergenerational Mobility, Intergenerational Effects, Sibling Correlations, and Equality of Opportunity: A Comparison of Four Approaches', *Research in Social Stratification and Mobility* **70**.

Björklund, A., Jäntti, M. & Lindquist, M. J. (2009), 'Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932-1968', *Journal of Public Economics* **93**(5-6), 671–680.

Björklund, A., Lindahl, L. & Lindquist, M. J. (2010), 'What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents', *B.E. Journal of Economic Analysis and Policy* **10**(1).

Björklund, A. & Salvanes, K. G. (2011), Education and Family Background: Mechanisms and Policies, *in* E. A. Hanushek, S. J. Machin & L. Woessmann, eds, 'Handbook of the Economics of Education', Vol. 3, North Holland, Amsterdam, chapter 3, pp. 201–247.

Bredtmann, J. & Smith, N. (2018), 'Inequalities in Educational Outcomes: How Important Is the Family?', *Oxford Bulletin of Economics and Statistics* **80**(6), 1117–1144.

Card, D., Domnisoru, C. & Taylor, L. (2022), 'The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility', *Journal of Labor Economics* **40**(S1), S39–S95.

Chetty, R., Hendren, N., Kline, P. & Saez, E. (2014), 'Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States', *The Quarterly Journal of Economics* **129**(4), 1553–1623.

Conley, D. & Glauber, R. (2008), 'All in the family? Family Composition, Resources, and Sibling Similarity in Socioeconomic Status', *Research in Social Stratification and Mobility* **26**(4), 297–306.

Derenoncourt, E. (2022), 'Can You Move to Opportunity? Evidence from the Great Migration', *American Economic Review* **112**(2), 369–408.

Grätz, M., Barclay, K. J., Wiborg, Ø. N., Lyngstad, T. H., Karhula, A., Erola, J., Präg, P., Laidley, T. & Conley, D. (2021), 'Sibling similarity in education across and within societies', *Demography* **58**(3), 1011–1037.

Hällsten, M. & Thaning, M. (2018), 'Wealth vs . education , occupation and income: Unique and overlapping influences of SES in intergenerational transmissions', *The Department of Sociology Working Paper Series, Stockholm University* (35).

Marks, G. N. & Mooi-Reci, I. (2016), 'The declining influence of family background on educational attainment in Australia: The role of measured and unmeasured influences', *Social Science Research* **55**, 171–185.

Mazumder, B. (2008), 'Sibling Similarities, Differences and Economic Inequality', *Journal of Population Economics* **21**(3), 685–701.

Munoz, E. (2024), 'The Geography of Intergenerational Mobility in Latin America and the Caribbean', *Economía LACEA Journal* **23**, 333–354.

Munoz, E. & Siravegna, M. (2023), 'When Measure Matters: Coresidence Bias and Intergenerational Mobility Revisited', *IDB Working Paper* **01469**.

Munoz, E. & Van der Weide, R. (2025), 'Intergenerational income mobility around the world: A new database', *World Bank Policy Research Working Paper* **11166**.

Raaum, O., Salvanes, K. G. & Sørensen, E. (2006), 'The neighbourhood is not what it used to be', *Economic Journal* **116**(508), 200–222.

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q. & Lee, J. (2024), 'Integrated Public Use Microdata Series, International: Version 7.5 [dataset]'.

Schnitzlein, D. (2014), 'How Important Is the Family? Evidence from Sibling Correlations in Permanent Earnings in the USA, Germany, and Denmark', *Journal of Population Economics* **27**(1), 69–89.

Sieben, I., Huinink, J. & De Graaf, P. M. (2001), 'Family background and sibling resemblance in educational attainment trends in the former FRG, the former GDR, and the Netherlands', *European Sociological Review* **17**(4), 401–430.

Sobek, M. & Kennedy, S. (2009), 'The development of family interrelationship variables for international census data', *Minneapolis, "University of Minnesota Population Center Working Paper Series* (2009-02).

Van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A. & Gupta, R. (2024), 'Intergenerational mobility around the world: A new database', *Journal of Development Economics* **166**, 103167.

Appendices

Identification of siblings

For IPUMS, the microdata from each census are organized at the household level, which allows us to link individuals living in the same household at the time of the interview via a variable that uses 62 distinct values to specify each member's relationship to the household head. This information enables us to identify sibling groups in two ways.

First, we use the variables *pernum_pop* and *pernum_mom*, which are constructed by IPUMS (Sobek & Kennedy 2009) and indicate the likely father and mother within the household for each individual. For example, for individuals classified as children of the household head, these variables typically identify the head as either the probable father or mother, depending on their gender. All individuals sharing the same probable father are considered siblings. Likewise, when information about the probable father is missing, individuals with the same probable mother are considered siblings.

Second, for individuals with missing values on both *pernum_pop* and *pernum_mom*, we identify sibling groups using the *Related* variable, which describes the individual's relationship to the head of household. We consider as siblings the household head and all individuals classified as his or her siblings or stepsiblings. Similarly, the spouse or partner of the household head and all individuals classified as his or her siblings or stepsiblings are also treated as sibling groups.

The Barbados census follows the same structure and approach as IPUMS. For the remaining data sources, which lack variables identifying the parental position within the household, we rely solely on the relationship variable. In these cases, we define siblings as individuals identified as children of the household head.[18]

In the case of the LIS data, the availability of variables varies across countries: Some

[18] This applies to the censuses of Belize and Guyana, as well as the labor force and household surveys and the PIP data.

provide parental position variables, others provide only a relationship variable, and some provide both. When parental position variables are available, we apply the same method as for IPUMS and Barbados. When only the relationship variable is available, we follow the same approach described above. In cases where both types of variables are available, we use the one that yields the largest number of observations.

Tables and Figures

Table A1: Overview of previous estimates available in the literature

Country	Study	Sibship type	Estimate
Australia	Marks & Mooi-Reci (2016)[2]	Mixed	0.34–0.58
Denmark	Bredtmann & Smith (2018)[1]	Mixed	0.33
Denmark	Bredtmann & Smith (2018)[1]	Brothers	0.31
Denmark	Bredtmann & Smith (2018)[1]	Sisters	0.39
Finland	Grätz et al. (2021)	Mixed	0.36
Germany	Schnitzlein (2014)[1]	Brothers	0.66
Germany	Schnitzlein (2014)[1,2]	Sisters	0.55
Netherlands	Sieben et al. (2001)[1,2]	Mixed	0.45
Norway	Raaum et al. (2006)[2]	Brothers	0.42
Norway	Raaum et al. (2006)[2]	Sisters	0.46–0.47
Norway	Björklund et al. (2010)[1]	Mixed	0.41
Norway	Björklund & Salvanes (2011)[2]	Mixed	0.40–0.42
Norway	Grätz et al. (2021)	Mixed	0.41
Sweden	Björklund et al. (2009)[2]	Brothers	0.46–0.48
Sweden	Björklund & Jäntti (2012)[2]	Sisters	0.39
Sweden	Björklund & Jäntti (2012)[3]	Brothers	0.46
Sweden	Björklund & Jäntti (2012)[1]	Sisters	0.40
Sweden	Hällsten & Thaning (2018)[2]	Mixed	0.38
Sweden	Grätz et al. (2021)	Mixed	0.44
UK	Grätz et al. (2021)	Mixed	0.42
USA	Conley & Glauber (2008)[2]	Mixed	0.63
USA	Mazumder (2008)[1,2]	Mixed	0.60
USA	Grätz et al. (2021)	Mixed	0.51

Notes: [1] Estimates compiled in Björklund & Jäntti (2020). [2] Estimates compiled in Grätz et al. (2021). Different estimates refer to estimates reported for different cohorts. [3] Björklund & Jäntti (2020) reports 0.43. but the paper has 0.46.

Table A2: Country-wave coverage and estimate overview

Country	Waves	Source	Estimate	Obs.	Educ
Albania	2002–2005	LSMS	0.51	1,256	Y
Albania	2012	LSMS	0.82	5,106	Y
Argentina	1970	IPUMS	0.71	9,446	Y
Argentina	1980	IPUMS	0.71	55,633	Y
Argentina	1991	IPUMS	0.67	89,881	Y
Argentina	2001	IPUMS	0.62	131,850	Y
Argentina	2010	IPUMS	0.60	105,572	Y
Armenia	2001	IPUMS	0.59	14,466	N
Armenia	2011	IPUMS	0.61	19,365	N
Australia	2004, 2008, 2010	LIS	0.27	792	Y
Australia	2014, 2016, 2018	LIS	0.22	1,025	Y
Austria	1971	IPUMS	0.48	6,896	N
Austria	1981	IPUMS	0.33	13,442	N
Austria	1991	IPUMS	0.33	20,657	N
Austria	2001	IPUMS	0.34	14,332	N
Bangladesh	1991	IPUMS	0.65	197,496	Y
Bangladesh	2001	IPUMS	0.68	228,654	Y
Bangladesh	2011	IPUMS	0.63	162,709	Y
Barbados	2010	Census	0.50	7,004	Y
Belarus	1999	IPUMS	0.45	12,131	N
Belarus	2009	IPUMS	0.42	21,728	N
Belgium	1985, 1988, 1992	LIS	0.39	1,005	Y
Belgium	2008–2011	LIS	0.34	1,117	Y
Belgium	2019–2021	LIS	0.44	1,213	Y
Belize	2022	Census	0.55	9,719	N
Benin	1979	IPUMS	0.70	9,256	Y
Benin	1992	IPUMS	0.73	14,183	Y
Benin	2002	IPUMS	0.71	16,100	Y
Benin	2013	IPUMS	0.64	40,970	Y
Bolivia	1976	IPUMS	0.80	5,293	Y
Bolivia	1992	IPUMS	0.74	8,506	Y
Bolivia	2001	IPUMS	0.71	18,511	Y
Bolivia	2012	IPUMS	0.54	25,761	Y
Bosnia and Herzegovina	2001	LSMS	0.61	637	Y
Botswana	1981	IPUMS	0.58	2,080	Y
Botswana	1991	IPUMS	0.50	4,005	Y
Botswana	2001	IPUMS	0.48	6,839	Y
Botswana	2011	IPUMS	0.45	7,606	Y
Brazil	1960	IPUMS	0.79	378,257	Y
Brazil	1970	IPUMS	0.78	592,754	Y
Brazil	1980	IPUMS	0.73	789,289	Y
Brazil	1991	IPUMS	0.70	637,297	Y
Brazil	2000	IPUMS	0.66	701,150	Y
Brazil	2010	IPUMS	0.65	696,070	N
Burkina Faso	1996	IPUMS	0.75	28,122	N
Burkina Faso	2006	IPUMS	0.72	40,588	N
Cambodia	1998	IPUMS	0.66	16,292	Y
Cambodia	2008	IPUMS	0.65	49,409	Y
Cambodia	2019	IPUMS	0.66	62,446	Y
Cameroon	1976	IPUMS	0.66	13,842	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Cameroon	1987	IPUMS	0.78	16,098	Y
Cameroon	2005	IPUMS	0.70	65,280	Y
Canada	2011	IPUMS	0.43	17,602	N
Chile	1970	IPUMS	0.71	21,473	Y
Chile	1982	IPUMS	0.61	42,139	Y
Chile	1992	IPUMS	0.58	49,177	Y
Chile	2002	IPUMS	0.58	45,447	Y
Chile	2017	IPUMS	0.44	45,126	Y
China	1982	IPUMS	0.38	218,661	N
China	1990	IPUMS	0.50	285,353	N
China	2000	IPUMS	0.63	248,010	N
Colombia	1973	IPUMS	0.69	48,707	Y
Colombia	1985	IPUMS	0.68	123,404	Y
Colombia	1993	IPUMS	0.66	129,857	Y
Colombia	2005	IPUMS	0.69	109,373	Y
Congo, Dem. Rep.	2004	E123	0.59	2,135	Y
Congo, Dem. Rep.	2012	E123	0.62	2,760	Y
Costa Rica	1973	IPUMS	0.66	5,370	Y
Costa Rica	1984	IPUMS	0.63	10,590	Y
Costa Rica	2000	IPUMS	0.66	10,902	Y
Costa Rica	2011	IPUMS	0.57	19,162	Y
Cuba	2002	IPUMS	0.46	19,839	N
Cuba	2012	IPUMS	0.37	17,964	Y
Czechia	1992, 1996	LIS	0.30	1,757	Y
Czechia	2002, 2004, 2007	LIS	0.27	1,196	Y
Czechia	2010, 2013, 2016	LIS	0.29	1,231	Y
Côte d'Ivoire	1988	IPUMS	0.68	19,815	N
Côte d'Ivoire	1998	IPUMS	0.63	54,981	Y
Denmark	2021	LIS	0.28	1,776	Y
Djibouti	1996	EDAM	0.45	1,033	Y
Dominican Republic	1981	IPUMS	0.58	16,509	Y
Dominican Republic	2002	IPUMS	0.50	26,731	Y
Dominican Republic	2010	IPUMS	0.56	30,602	Y
Ecuador	1974	IPUMS	0.77	11,830	Y
Ecuador	1982	IPUMS	0.71	20,740	Y
Ecuador	1990	IPUMS	0.64	30,106	Y
Ecuador	2001	IPUMS	0.66	29,419	Y
Ecuador	2010	IPUMS	0.62	37,235	Y
Egypt, Arab Rep	1986	IPUMS	0.56	308,738	N
Egypt, Arab Rep	1996	IPUMS	0.55	244,149	N
Egypt, Arab Rep	2006	IPUMS	0.62	390,287	N
El Salvador	1992	IPUMS	0.70	11,097	Y
El Salvador	2007	IPUMS	0.64	20,327	Y
Estonia	2007, 2010, 2013, 2016	LIS	0.32	1,228	Y
Ethiopia	1984	IPUMS	0.79	21,937	Y
Ethiopia	1994	IPUMS	0.80	62,661	Y
Ethiopia	2007	IPUMS	0.77	18,178	Y
Fiji	1976	IPUMS	0.47	1,648	Y
Fiji	1986	IPUMS	0.51	2,561	Y
Fiji	1996	IPUMS	0.49	2,134	Y
Fiji	2007	IPUMS	0.42	3,005	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Fiji	2014	IPUMS	0.32	3,820	Y
Finland	1987, 1991, 1995, 2000, 2004 2007, 2010, 2013, 2016	LIS	0.21	725	Y
France	1968	IPUMS	0.35	18,979	N
France	1975	IPUMS	0.53	25,162	N
France	1982	IPUMS	0.54	26,682	N
France	1990	IPUMS	0.51	31,927	N
France	1999	IPUMS	0.46	34,890	N
France	2011	IPUMS	0.47	187,345	N
Gambia, The	2015	IHS	0.34	1,905	Y
Georgia	2010–2012	LIS	0.50	1,217	Y
Georgia	2017–2022	LIS	0.47	1,080	Y
Germany	1984–1989	LIS	0.38	1,221	Y
Germany	1990–1994	LIS	0.32	1,127	Y
Germany	2000–2003	LIS	0.30	1,015	Y
Germany	2010–2014	LIS	0.27	1,198	Y
Germany	2016–2020	LIS	0.40	1,222	Y
Ghana	1984	IPUMS	0.65	51,358	Y
Ghana	2000	IPUMS	0.58	42,888	Y
Ghana	2010	IPUMS	0.61	99,256	Y
Greece	1971	IPUMS	0.56	14,702	N
Greece	1981	IPUMS	0.57	13,187	N
Greece	1991	IPUMS	0.57	23,975	N
Greece	2001	IPUMS	0.55	42,315	N
Greece	2011	IPUMS	0.49	31,289	N
Guatemala	1964	IPUMS	0.81	3,249	Y
Guatemala	1973	IPUMS	0.84	4,393	Y
Guatemala	1981	IPUMS	0.81	5,618	Y
Guatemala	1994	IPUMS	0.77	18,214	Y
Guatemala	2002	IPUMS	0.76	34,012	Y
Guinea	1983	IPUMS	0.66	13,619	Y
Guinea	1996	IPUMS	0.59	25,158	Y
Guinea	2014	IPUMS	0.60	45,860	Y
Guyana	2012	Census	0.52	9,358	Y
Haiti	1971	IPUMS	0.77	9,402	Y
Haiti	1982	IPUMS	0.73	3,295	Y
Haiti	2003	IPUMS	0.70	25,140	Y
Honduras	1974	IPUMS	0.71	5,040	Y
Honduras	1988	IPUMS	0.71	10,490	Y
Honduras	2001	IPUMS	0.72	16,180	Y
Hungary	1970	IPUMS	0.59	2,868	N
Hungary	1980	IPUMS	0.47	5,277	Y
Hungary	1990	IPUMS	0.46	3,731	Y
Hungary	2001	IPUMS	0.46	10,673	N
Hungary	2011	IPUMS	0.41	10,232	N
India	1983	IPUMS	0.70	14,828	Y
India	1993	IPUMS	0.71	16,679	N
India	2004	IPUMS	0.69	22,029	N
Indonesia	1971	IPUMS	0.76	7,455	N
Indonesia	1980	IPUMS	0.70	103,739	Y
Indonesia	1990	IPUMS	0.66	20,817	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Indonesia	2000	IPUMS	0.70	572,661	N
Indonesia	2010	IPUMS	0.68	682,226	N
Iran, Islamic Rep	2006	IPUMS	0.54	79,965	N
Iran, Islamic Rep	2011	IPUMS	0.55	93,701	N
Iraq	1997	IPUMS	0.58	165,155	N
Ireland	1971	IPUMS	0.58	5,568	N
Ireland	1981	IPUMS	0.58	10,647	N
Ireland	1991	IPUMS	0.53	12,928	N
Ireland	2002	IPUMS	0.46	15,433	N
Ireland	2011	IPUMS	0.48	10,881	N
Israel	1972	IPUMS	0.45	8,725	N
Israel	1983	IPUMS	0.46	18,630	N
Israel	1995	IPUMS	0.47	18,649	N
Italy	2001	IPUMS	0.52	129,452	N
Italy	2011	IPUMS	0.52	85,746	N
Jamaica	1982	IPUMS	0.52	4,568	Y
Jamaica	1991	IPUMS	0.39	7,514	Y
Jamaica	2001	IPUMS	0.51	6,622	Y
Japan	2016–2020	LIS	0.51	761	Y
Jordan	2004	IPUMS	0.54	39,933	N
Kenya	1969	IPUMS	0.72	6,798	Y
Kenya	1989	IPUMS	0.63	16,346	Y
Kenya	1999	IPUMS	0.64	33,909	Y
Kenya	2009	IPUMS	0.74	103,474	Y
Korea, Rep.	2010, 2012	LIS	0.18	1,460	Y
Kyrgyz Republic	1999	IPUMS	0.37	16,763	N
Kyrgyz Republic	2009	IPUMS	0.42	31,956	N
Lao PDR	1995	IPUMS	0.64	8,215	Y
Lao PDR	2005	IPUMS	0.63	15,760	Y
Lao PDR	2015	IPUMS	0.63	30,239	Y
Lesotho	1996	IPUMS	0.57	7,607	Y
Lesotho	2006	IPUMS	0.60	7,619	Y
Liberia	2008	IPUMS	0.61	9,873	Y
Lithuania	2009–2013	LIS	0.36	920	Y
Luxembourg	1987–1993	LIS	0.51	1,032	Y
Luxembourg	1998–2003	LIS	0.39	1,014	Y
Luxembourg	2010–2012	LIS	0.40	1,018	Y
Luxembourg	2018–2021	LIS	0.36	1,369	Y
Malawi	1987	IPUMS	0.67	6,017	Y
Malawi	1998	IPUMS	0.67	9,811	Y
Malawi	2008	IPUMS	0.64	15,238	Y
Malaysia	1970	IPUMS	0.55	2,564	N
Malaysia	1980	IPUMS	0.49	5,856	Y
Malaysia	1991	IPUMS	0.55	11,867	Y
Malaysia	2000	IPUMS	0.61	14,480	N
Maldives	2016	PIP	0.31	1,023	N
Mali	1987	IPUMS	0.72	20,735	N
Mali	1998	IPUMS	0.64	25,983	Y
Mali	2009	IPUMS	0.68	42,057	Y
Mauritius	1990	IPUMS	0.55	7,369	Y
Mauritius	2000	IPUMS	0.58	5,538	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Mauritius	2011	IPUMS	0.51	3,624	Y
Mexico	1970	IPUMS	0.70	9,226	Y
Mexico	1990	IPUMS	0.64	290,562	Y
Mexico	2000	IPUMS	0.70	354,834	Y
Mexico	2010	IPUMS	0.65	384,977	Y
Mexico	2020	IPUMS	0.62	484,430	Y
Mongolia	1989	IPUMS	0.43	5,183	N
Mongolia	2000	IPUMS	0.54	10,357	N
Morocco	1982	IPUMS	0.59	38,985	Y
Morocco	1994	IPUMS	0.57	95,110	Y
Morocco	2004	IPUMS	0.59	118,096	Y
Morocco	2014	IPUMS	0.58	192,938	Y
Mozambique	1997	IPUMS	0.60	15,384	N
Mozambique	2007	IPUMS	0.63	27,309	N
Myanmar	2014	IPUMS	0.69	181,107	Y
Namibia	2015	PIP	0.49	867	N
Nepal	2001	IPUMS	0.68	51,979	Y
Nepal	2011	IPUMS	0.67	84,856	Y
Netherlands	2008-2014	LIS	0.18	1,176	Y
Netherlands	2019–2021	LIS	0.22	1,058	Y
Nicaragua	1971	IPUMS	0.80	3,645	Y
Nicaragua	1995	IPUMS	0.66	13,653	Y
Nicaragua	2005	IPUMS	0.70	22,158	Y
Niger	2011	ECVMA	0.81	1,806	Y
Nigeria	2010	IPUMS	0.87	17,321	Y
Norway	1991, 1995	LIS	0.15	1,059	Y
Norway	2000, 2004, 2007	LIS	0.40	2,125	Y
Norway	2010	LIS	0.29	1,923	Y
Norway	2020	LIS	0.28	3,351	Y
Pakistan	1973	IPUMS	0.66	38,561	N
Pakistan	1998	IPUMS	0.68	497,720	N
Panama	1960	IPUMS	0.77	829	Y
Panama	1970	IPUMS	0.76	2,967	Y
Panama	1980	IPUMS	0.65	5,240	Y
Panama	1990	IPUMS	0.62	8,876	Y
Panama	2000	IPUMS	0.63	8,644	Y
Panama	2010	IPUMS	0.60	9,706	Y
Papua New Guinea	1980	IPUMS	0.63	4,437	Y
Papua New Guinea	1990	IPUMS	0.56	6,062	Y
Papua New Guinea	2000	IPUMS	0.61	13,303	Y
Paraguay	1962	IPUMS	0.72	1,855	Y
Paraguay	1972	IPUMS	0.76	5,534	Y
Paraguay	1982	IPUMS	0.76	9,495	Y
Paraguay	1992	IPUMS	0.73	10,217	Y
Paraguay	2002	IPUMS	0.69	15,009	Y
Peru	1993	IPUMS	0.64	75,952	Y
Peru	2007	IPUMS	0.54	95,857	Y
Peru	2017	IPUMS	0.42	98,065	Y
Philippines	1990	IPUMS	0.66	277,650	Y
Philippines	2000	IPUMS	0.64	339,080	Y
Philippines	2010	IPUMS	0.64	454,831	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Poland	1978	IPUMS	0.51	68,071	N
Poland	2002	IPUMS	0.51	105,025	N
Portugal	1981	IPUMS	0.74	8,791	N
Portugal	1991	IPUMS	0.72	14,201	N
Portugal	2001	IPUMS	0.61	14,986	N
Portugal	2011	IPUMS	0.59	10,463	N
Puerto Rico	1970	IPUMS	0.54	600	Y
Puerto Rico	1980	IPUMS	0.46	3,841	Y
Puerto Rico	1990	IPUMS	0.43	5,094	N
Puerto Rico	2000	IPUMS	0.45	4,205	N
Puerto Rico	2010	IPUMS	0.35	665	Y
Romania	1977	IPUMS	0.57	20,511	N
Romania	1992	IPUMS	0.59	38,678	N
Romania	2002	IPUMS	0.68	64,873	N
Romania	2011	IPUMS	0.70	47,330	N
Russian Federation	2002	IPUMS	0.51	127,695	N
Russian Federation	2010	IPUMS	0.54	201,157	N
Rwanda	2002	IPUMS	0.55	16,880	Y
Rwanda	2012	IPUMS	0.57	34,050	Y
Senegal	1988	IPUMS	0.66	31,402	Y
Senegal	2002	IPUMS	0.62	54,443	Y
Senegal	2013	IPUMS	0.60	71,408	Y
Serbia	2010–2012	LIS	0.43	1,368	Y
Serbia	2019, 2021	LIS	0.38	1,019	Y
Sierra Leone	2004	IPUMS	0.59	14,100	Y
Sierra Leone	2015	IPUMS	0.50	16,158	Y
Slovak Republic	1992	LIS	0.60	637	Y
Slovak Republic	2004, 2007	LIS	0.35	2,139	Y
Slovak Republic	2010	LIS	0.23	1,261	Y
Slovak Republic	2017–2018	LIS	0.35	1,237	Y
Slovenia	2002	IPUMS	0.25	7,837	N
South Africa	2001	IPUMS	0.60	169,990	Y
South Africa	2011	IPUMS	0.49	187,176	Y
South Sudan	2008	IPUMS	0.60	12,691	N
Spain	1991	IPUMS	0.66	95,859	N
Spain	2001	IPUMS	0.55	122,542	N
Sri Lanka	2002	HIES	0.58	3,459	Y
St Lucia	1980	IPUMS	0.24	260	Y
St Lucia	1991	IPUMS	0.44	546	N
Sudan	2008	IPUMS	0.74	187,896	N
Suriname	2012	IPUMS	0.54	1,744	N
Sweden	2000–2004	LIS	0.42	1,056	Y
Sweden	2005–2010	LIS	0.32	1,582	Y
Switzerland	1970	IPUMS	0.31	2,706	N
Switzerland	1980	IPUMS	0.41	2,480	N
Switzerland	1990	IPUMS	0.39	4,650	N
Switzerland	2000	IPUMS	0.37	3,327	N
Taiwan, China	1981	LIS	0.53	1,721	Y
Taiwan, China	1991	LIS	0.48	1,930	Y
Taiwan, China	2000	LIS	0.27	1,932	Y
Taiwan, China	2010	LIS	0.24	1,998	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
Taiwan, China	2020	LIS	0.21	1,617	Y
Tajikistan	1999	LSS	0.18	1,467	Y
Tanzania	1988	IPUMS	0.50	32,171	Y
Tanzania	2002	IPUMS	0.54	45,880	Y
Tanzania	2012	IPUMS	0.53	68,545	Y
Thailand	1970	IPUMS	0.66	17,890	Y
Thailand	1980	IPUMS	0.67	11,447	Y
Thailand	1990	IPUMS	0.70	20,948	Y
Thailand	2000	IPUMS	0.68	18,077	Y
Timor-Leste	2007	SLS	0.56	855	Y
Togo	1960	IPUMS	0.29	485	N
Togo	1970	IPUMS	0.54	850	N
Togo	2010	IPUMS	0.69	13,507	Y
Trinidad and Tobago	1970	IPUMS	0.46	1,980	N
Trinidad and Tobago	1980	IPUMS	0.46	5,070	Y
Trinidad and Tobago	1990	IPUMS	0.44	6,017	Y
Trinidad and Tobago	2000	IPUMS	0.51	4,741	Y
Trinidad and Tobago	2011	IPUMS	0.44	6,501	Y
Tunisia	2005	PIP	0.49	4,573	N
Tunisia	2010	PIP	0.51	4,832	N
Tunisia	2015	PIP	0.45	8,091	N
Turkey	1985	IPUMS	0.54	46,028	N
Turkey	1990	IPUMS	0.54	59,556	N
Turkey	2000	IPUMS	0.55	111,958	N
Uganda	1991	IPUMS	0.63	28,403	Y
Uganda	2002	IPUMS	0.68	30,239	Y
Uganda	2014	IPUMS	0.55	49,947	Y
United Kingdom	2000, 2002	LIS	0.43	1,109	Y
United Kingdom	2010–2012	LIS	0.46	1,027	Y
United Kingdom	2017–2021	LIS	0.63	1,388	Y
United States	1960	IPUMS	0.60	10,160	Y
United States	1970	IPUMS	0.55	12,676	Y
United States	1980	IPUMS	0.52	145,049	Y
United States	1990	IPUMS	0.52	193,472	N
United States	2000	IPUMS	0.55	146,544	N
United States	2010	IPUMS	0.45	31,994	Y
United States	2020	IPUMS	0.34	178,596	Y
Uruguay	1963	IPUMS	0.63	4,987	Y
Uruguay	1975	IPUMS	0.56	4,001	Y
Uruguay	1985	IPUMS	0.55	5,292	Y
Uruguay	1996	IPUMS	0.59	5,578	Y
Uruguay	2011	IPUMS	0.59	7,196	N
Venezuela, RB	1971	IPUMS	0.60	25,402	Y
Venezuela, RB	1981	IPUMS	0.60	58,386	Y
Venezuela, RB	1990	IPUMS	0.64	63,319	Y
Venezuela, RB	2001	IPUMS	0.56	86,338	Y
Vietnam	1989	IPUMS	0.64	91,152	Y
Vietnam	1999	IPUMS	0.66	79,831	Y
Vietnam	2009	IPUMS	0.58	434,594	Y
Vietnam	2019	IPUMS	0.62	223,177	Y
West Bank and Gaza	1997	IPUMS	0.45	15,614	Y

Continued on next page

Table A2 – continued from previous page

Country	Waves	Source	Estimate	Obs.	Educ
West Bank and Gaza	2007	IPUMS	0.44	15,684	N
West Bank and Gaza	2017	IPUMS	0.43	34,519	Y
Zambia	1990	IPUMS	0.52	23,552	Y
Zambia	2000	IPUMS	0.51	24,619	Y
Zambia	2010	IPUMS	0.68	32,422	Y
Zimbabwe	2012	IPUMS	0.61	11,936	Y

Notes: Column *Educ* refers to the educational variable used in the estimation. *Y* stands for "years of schooling," and *N* stands for "educational levels or categories."

Table A3: Country coverage by income level

Income group/Region	Number of countries	Population share, %
High income	40	95.27
Developing countries	88	93.74
East Asia & Pacific	13	98.64
Europe & Central Asia	9	49.52
Latin America & Caribbean	21	99.94
Middle East & North Africa	8	73.35
South Asia	6	98.53
Sub-Saharan Africa	31	89.17
World	128	94.05

Notes: The table shows the number of countries covered by the dataset and the population share that they account for. The country classification follows the World Bank's income grouping and geographic regions for fiscal year 2025, available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

Table A4: Sibling correlation by region, estimated with years of schooling

Region	Average	First cohort	Last cohort	Countries
Europe & Central Asia	0.38	0.38	0.39	19
North America	0.49	0.60	0.34	1
Middle East & North Africa	0.52	0.50	0.49	3
East Asia & Pacific	0.54	0.55	0.52	15
Sub-Saharan Africa	0.62	0.63	0.62	26
Latin America & Caribbean	0.63	0.64	0.56	26
South Asia	0.66	0.65	0.65	4
World	0.57	0.57	0.53	94

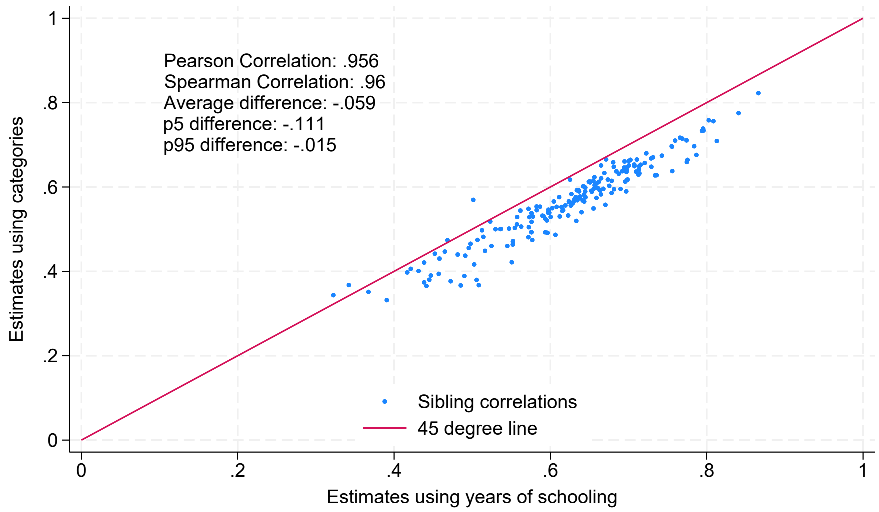
Notes: The table reports simple averages of our estimates for each region. The columns "Average," "First cohort," and "Last cohort," respectively, use all the birth cohorts available for each country, only the first cohort available for each country, and only the latest birth cohort available for each country. The column "Countries" reports the number of countries with data for each region. Rows are sorted according to the "Average" column.

Table A5: Sibling correlations by income level

Region	Average	First cohort	Last cohort	Countries
High income	0.46	0.46	0.41	40
Developing countries	0.61	0.60	0.59	88
Europe & Central Asia	0.49	0.47	0.49	9
Middle East & North Africa	0.53	0.52	0.52	8
East Asia & Pacific	0.60	0.60	0.61	13
Sub-Saharan Africa	0.62	0.62	0.62	31
South Asia	0.64	0.60	0.60	6
Latin America & Caribbean	0.65	0.66	0.59	21

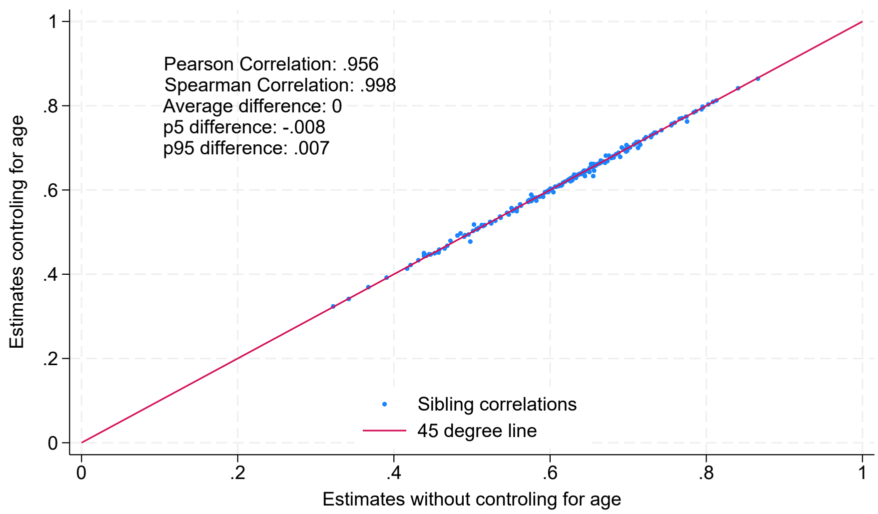
Notes: The table reports simple averages of our estimates for each region by income level. The columns "Average," "First cohort," and "Last cohort," respectively, use all the birth cohorts available for each country, only the first cohort (or census sample) available for each country, and only the latest birth cohort (or census sample) available for each country. The column "Countries" reports the number of countries with data for each region. Rows are sorted according to the "Average" column.

Figure A1: Estimates obtained with levels of education after imputation of years of schooling



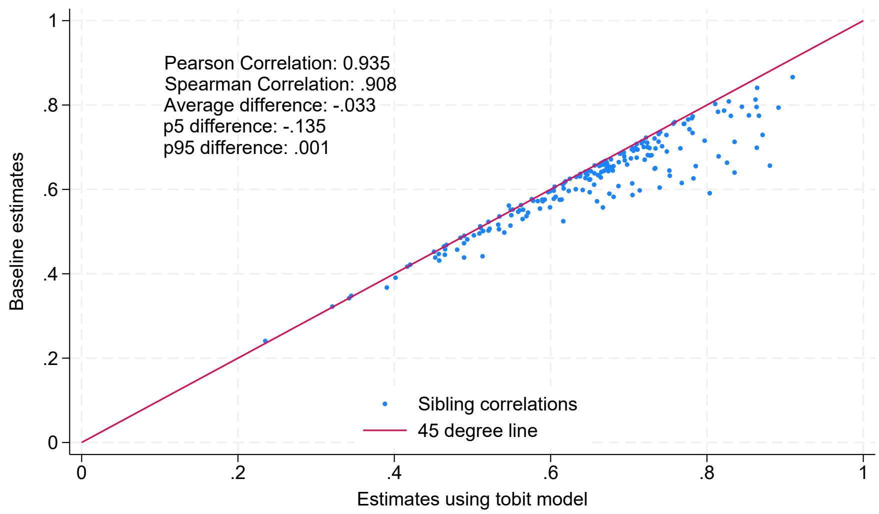
Notes: The figure compares estimates computed on the basis of years of schooling with those obtained by imputation of years from the variable based on levels of education. We use samples from IPUMS International, for which years and levels are available.

Figure A2: Sensitivity of sibling correlations to controls for age



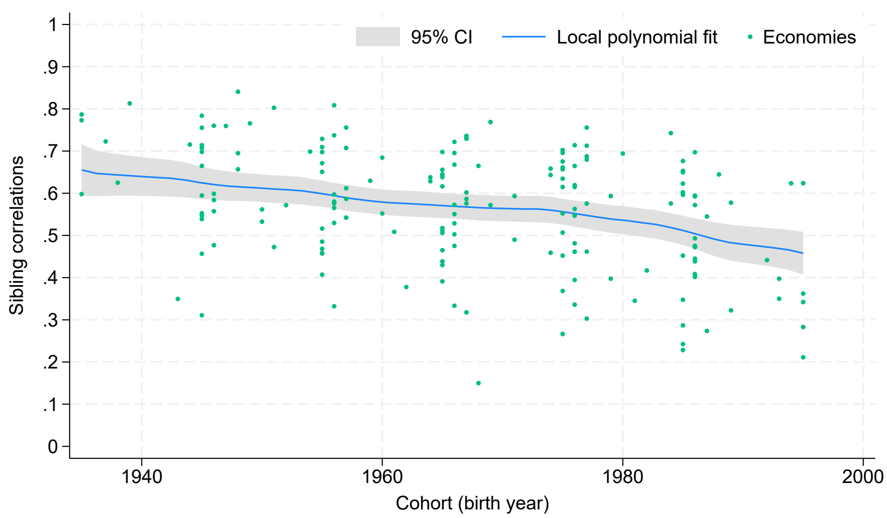
Notes: The figure compares estimates from a regression including a quadratic polynomial with age with estimates obtained without this age control. We use the samples obtained from IPUMS International.

Figure A3: Sensitivity of sibling correlations to censoring of schooling



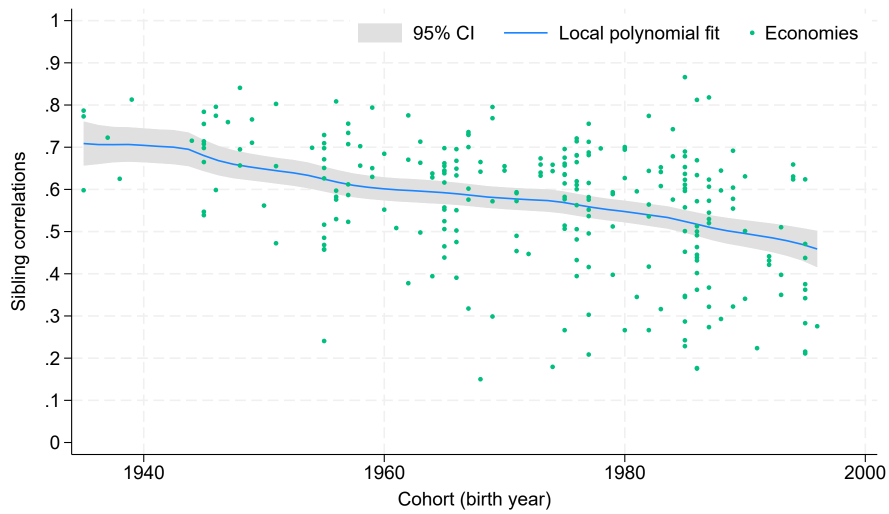
Notes: The figure compares estimates we compute using years of schooling as in our baseline analysis with those we obtain by estimating a tobit model that accounts for the censoring of years of schooling. We use samples from IPUMS International for which years of schooling are available.

Figure A4: Sibling correlations over time (economies with at least four estimates)



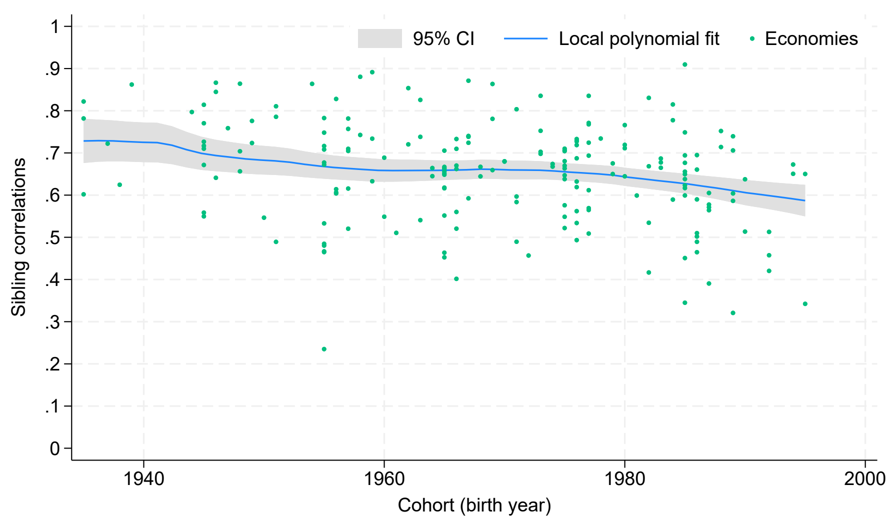
Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection.

Figure A5: Sibling correlation in years of schooling over time



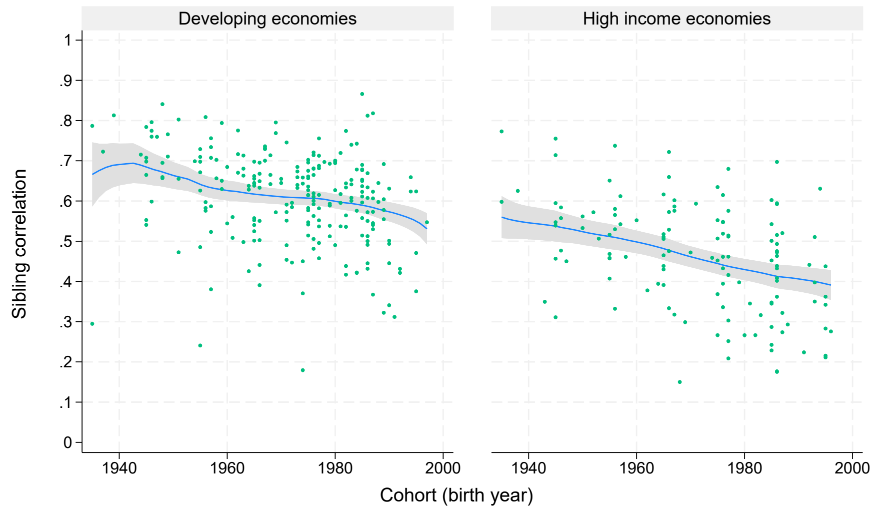
Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection.

Figure A6: Sibling correlation over time (accounting for censorship in education)



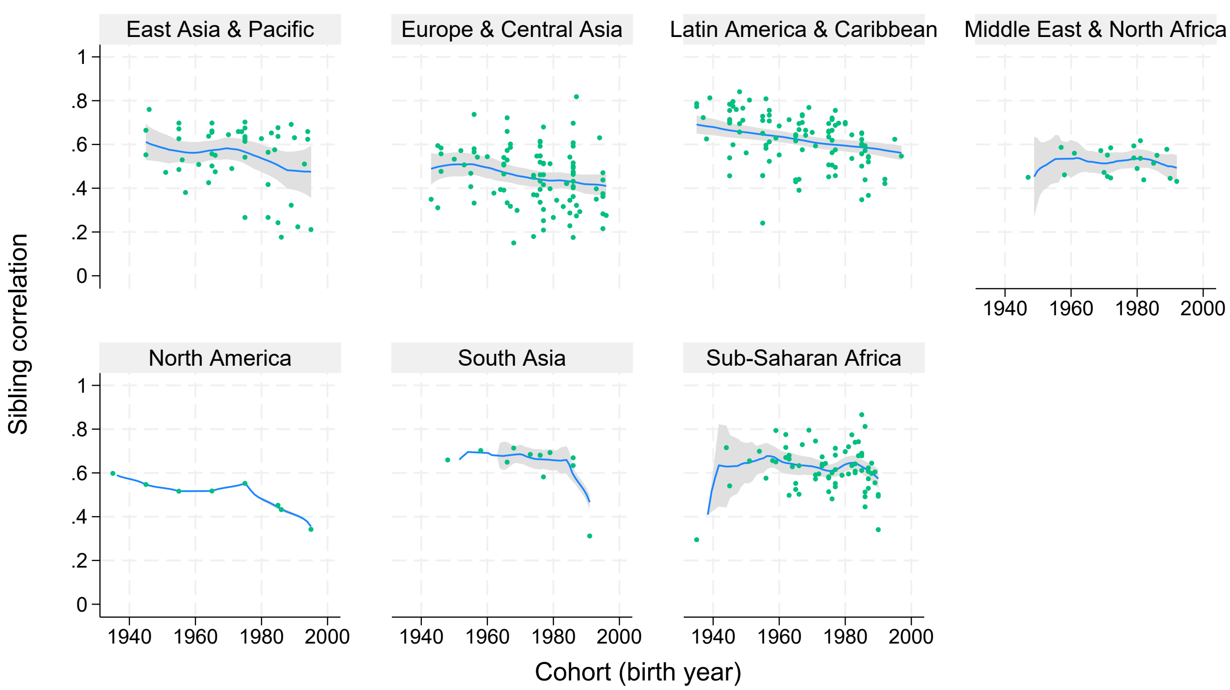
Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection. We use samples from IPUMS International for which years of schooling are available.

Figure A7: Sibling correlations over time by country income level



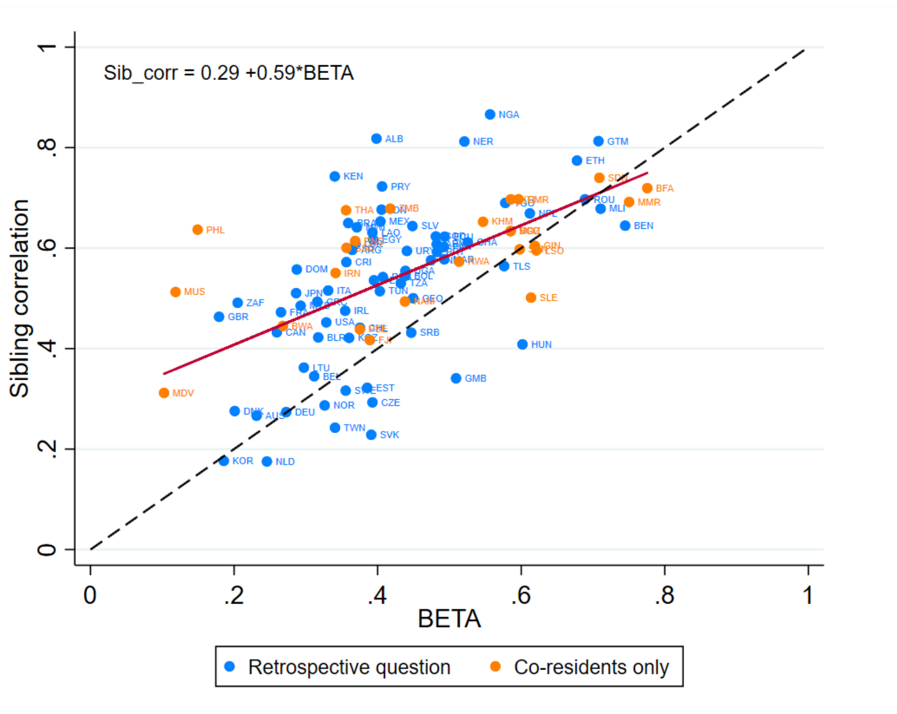
Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection. We categorize economies by income level using the World Bank's income classification.

Figure A8: Sibling correlations over time by region



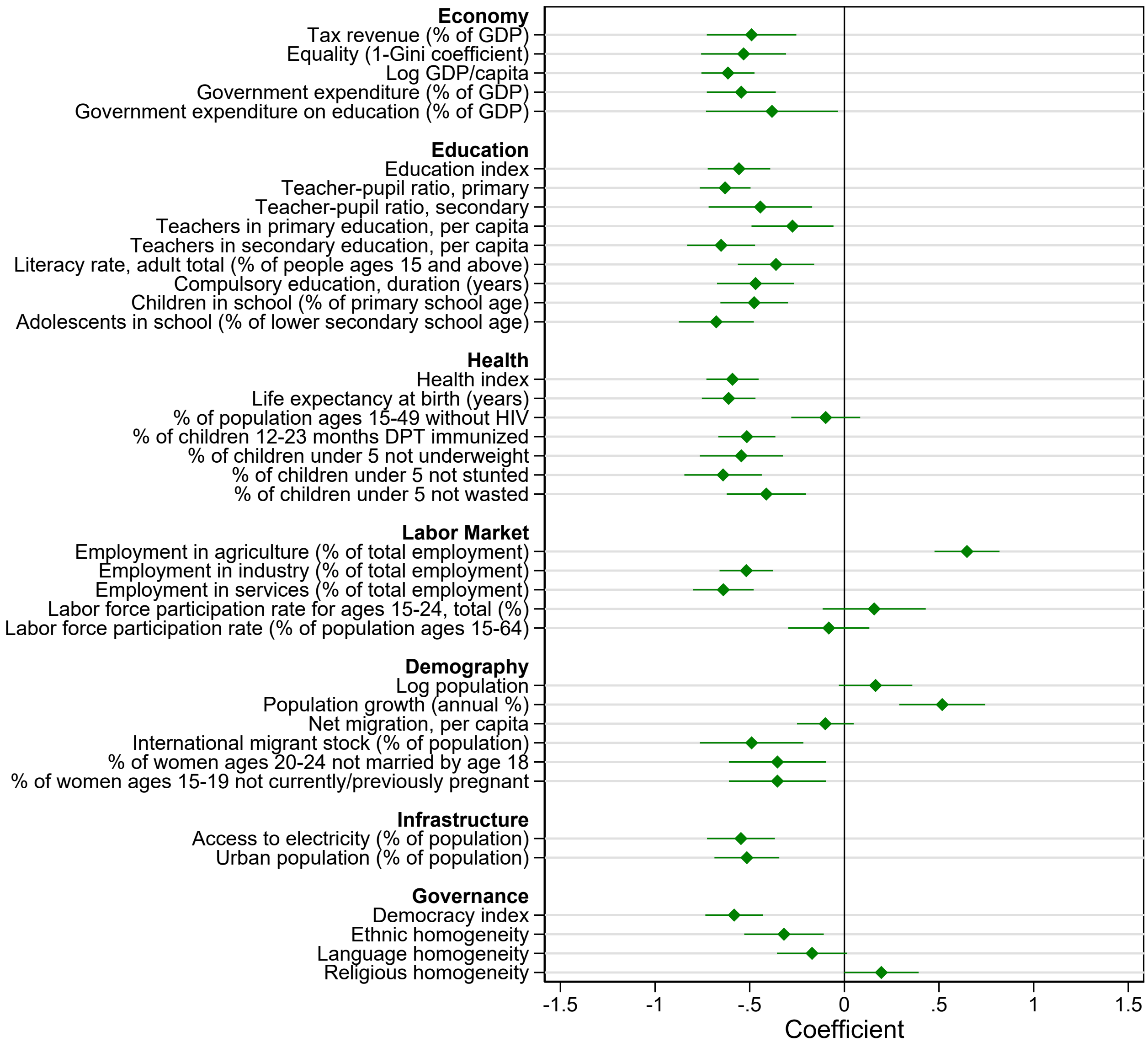
Notes: The figure presents estimates of sibling correlations by birth cohort for individuals aged 21–30 years. We assign the birth year by subtracting 25 from the year of data collection. We categorize economies by region using the World Bank's regional classification.

Figure A9: Sibling similarity and the intergenerational regression coefficient



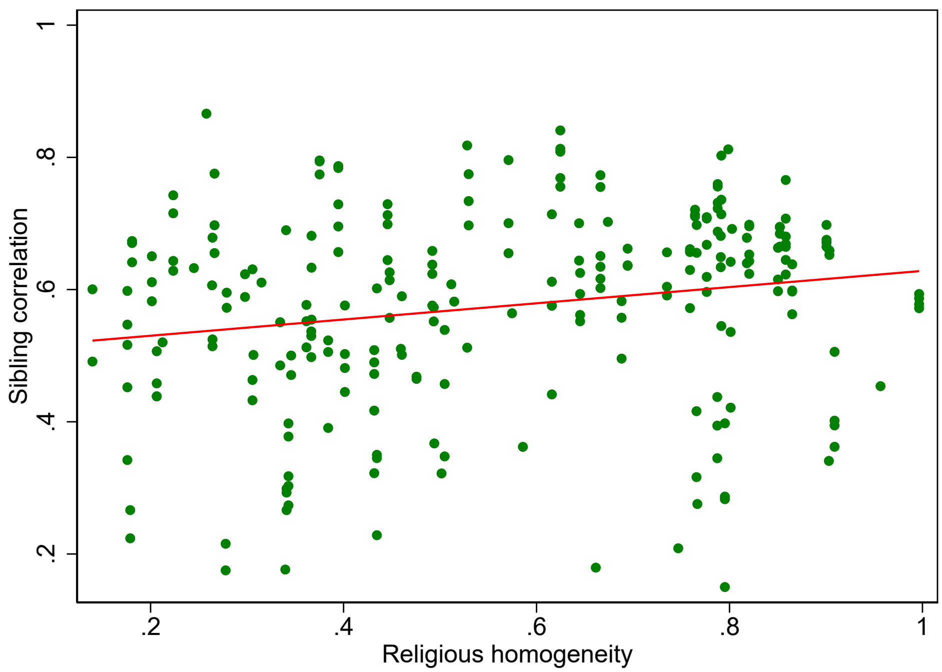
Notes: The indicator of intergenerational mobility on the x axis corresponds to the one computed by means of a regression of children's schooling on their parents' average schooling in Van der Weide et al. (2024) for the cohort born in the 1980s. These estimates are obtained from surveys with retrospective information and coresident samples, as highlighted. The dashed line is a 45-degree line. The indicator of sibling similarity corresponds to the most recent cohort available for each country in our database.

Figure A10: Correlates of sibling similarity estimated with years of schooling



Notes: The figure plots the coefficients from a regression of our measure of sibling correlations (standardized to have mean 0 and variance 1) on each covariate (standardized to have mean 0 and variance 1), including 95% confidence intervals with standard errors clustered by country. We fill data gaps for the covariates using interpolation. The confidence intervals do not account for the uncertainty in the estimate of sibling correlation. Each sample is matched with the value of a correlate at 10 years earlier, such that a given cohort aged 21–30 years is associated with the value of the correlates when the cohort members were around primary or secondary school age.

Figure A11: Sibling similarity and religious homogeneity



Notes: Red line is the linear fit.