Intergenerational Mobility in Africa*

Alberto Alesina
Harvard University, CEPR and NBER

Sebastian Hohmann
London Business School

Stelios Michalopoulos
Brown University, CEPR and NBER

Elias Papaioannou
London Business School and CEPR

This draft: July 2018

Abstract

We examine intergenerational mobility in educational attainment in Africa, using census data covering close to 20 million people in 23 countries. We define the former as the likelihood that children of uneducated parents complete (at least) primary schooling. In the first part of the paper, we characterize the geography of intergenerational mobility across and within countries. The observed cross-sectional heterogeneity is tightly and positively linked to the level of literacy of the "old generation" in the region, suggesting strong persistence of initial conditions. Inertia is stronger for rural, as compared to urban, households and present both for boys and girls. In the second part, we identify the causal effects of regions on educational mobility exploiting within-family variation from migrant households. The analysis reveals that while sorting is sizeable, there are large regional exposure effects. In the third part, we explore the geographic, historical, and contemporary correlates of intergenerational mobility across regions. Colonial investments, in terms of the transportation infrastructure, and some geographic features correlate strongly with educational mobility beyond their relationship with the level of schooling among the old.

Keywords: Africa, Development, Education, Inequality, Intergenerational Mobility.

JEL Numbers. N00, N9, O10, O43, O55

*For their comments we are grateful to Oriana Bandiera, Larry Katz, Nathan Nunn and Sara Lowes, conference participants at the University of Zurich, UPF, Monash University, and seminar participants at Brown, UPF, Harvard, Queen Mary, and Williams College. We also thank Remi Jedwab, Adam Storeygard Julia Cagé, Valeria Rueda and Nathan Nunn for generously sharing their data.

1 Introduction

There is rising optimism about Africa's future, a continent with 1.2 billion opportunities, as the Economist (2017) recently touted. The formerly "hopeless continent" is gradually becoming the "hopeful one" (The Economist, 2000 and 2011). Educational attainment is rising, health is improving, and incomes are growing, so that many now talk of an African growth miracle (Young, 2012). At the same time, anecdotal evidence – data are scant – suggests large inequalities, implying that the recent gains in aggregate may not be broadly shared. Similarly, narratives regarding inequities in social mobility and opportunity abound, but a comprehensive assessment is lacking.

In this paper, we take the first step in mapping and exploring mobility in educational attainment and its correlates across African countries, provinces, and regions, using matched parents-children census-level data covering 23 countries. Taking a holistic approach, we shed light on some intriguing questions on the distribution of opportunity across and within African countries since independence. Where is the land of educational opportunity? How wide are differences in the intergenerational mobility of education across African countries and regions? How large are gender disparities and how big is the rural-urban gap? What is the association between intergenerational mobility and literacy across countries and regions? How much of the observed social mobility reflects sorting/migration or region-specific factors? Which elements of a region's history, geography and at-independence features correlate with educational mobility?

Results Preview We organize the analysis into four parts.

In the first part, we construct new cross-country and within-country across-region measures of educational opportunity in Africa. Following recent works on intergenerational mobility in income (Chetty *et al.*, 2014; Chetty and Hendren, 2018a,b) and education (Card, Domnisoru, and Taylor, 2018) we compile measures of social mobility reflecting the likelihood that children whose parents have not completed primary schooling will manage to complete at least primary schooling ("absolute mobility"). We use data from close to 20 million individuals, who are residing with at least one parent in the time of the census. To account for "selection on cohabitation", we follow the simple and transparent approach of Card, Domnisoru, and Taylor (2018), who propose looking at children in the $14-18$ age range (around 8.5 million individuals); by that age primary schooling attendance is mostly over while children in this age bracket still reside with their parents. Using the uncensored statistics (computed from all individuals above the age of 14) and the measures looking solely at children aged $14-18$, we provide mappings of intergenerational educational mobility (social mobility) across 23 African countries and $2,440$ regions (admin 2/3 geographical units).

The analysis uncovers considerable heterogeneity both across and within countries. Social mobility is considerably higher (lower) in regions and countries with relatively higher (lower) literacy levels, suggesting strong inertia. Variation in literacy rates of the old generation across regions explains more than half of the observed variability in intergenerational mobility. This result implies sizeable persistence of initial educational levels across space. Inertia is stronger for rural, as compared to urban, households, which is consistent with the rising African urbanization during the past decades (Jedwab and Storeygard, 2017). We also document a gender gap in educational mobility that varies across the continent.

In the second part, we move beyond a descriptive analysis and try to identify the causal effect of

regions on social mobility. We do so by focusing on multi-children households that have moved over time across different regions. We start our analysis estimating household fixed-effects specifications looking at the effect of children's' place-of-birth-literacy among the old on educational mobility. These specifications that look at households with children born in different locations mitigate the issue of sorting. The latter rises when families that value education more are already found in regions where educational opportunities abound and vice versa, making it hard to disentangle the role of family characteristics from the location-specific attributes. The richness of the census data allows us to estimate these specifications across $265,000$ individuals in $96,000$ households. The results reveal that sorting is non-negligible,complementing the findings from the recent work of Young (2014). Nevertheless, the association between social mobility and place of birth literacy remains significant, even in these restrictive specifications.

To further isolate the role of regional features on social mobility, we build on Hendren and Chetty (2018a) and exploit differences in the age at which children of migrant households move. We uncover significant "*regional exposure effects*". Social mobility of children of uneducated parents increases when kids move to regions with higher educational attainment (and higher mobility); and this relationship is approximately linear for kids when they move in the $5-15$ age window that is relevant for primary schooling. These results uncover the causal effects of regions on educational mobility across Africa.

In the third part, we attempt to open the black box of region-specific correlates of social mobility. Building on a vibrant body of research on the deep origins of contemporary African development (summarized below) we investigate the relationship between geographic and historical variables and educational mobility, conditional on country fixed effects. While these correlations do not identify causal effects, we conduct the correlation analysis using both the simple statistics of educational mobility (part I) and the within-household statistics that net sorting (part II). Among various geographic variables, terrain ruggedness is significantly positively correlated to mobility, consistent with ruggedness shielding the local populations from the adverse effects of the slave trades (Nunn and Puga (2012) and Nunn (2008)). Distance to the coast and to a lesser extent distance to the capital (negatively) correlate with IM.

Colonial investments in railroads and schools are positively correlated with both measures of social mobility; and the correlations retain significance even when we account for the literacy of the old, which is the most significant correlate of educational mobility (parts I and II). Social mobility is unrelated to ethnic fractionalization and polarization. And while educational mobility, as well as literacy rates, are related to regional industrial specialization, the significantly positive (negative) correlation between intergenerational mobility and employment share in manufacturing-services (agriculture) turns insignificant once we condition on the literacy of the "old" generation.

Related Literature Our work contributes to two strands of the literature that have, however, moved in parallel.

The first is the growing research that studies intergenerational mobility in well-being across countries and regions. Corak (2013) and Solon (1999) and Black and Devereux (2011) review works on intergenerational mobility in income/wealth and education, respectively.[1] A key challenge has been the matching

[1]Olivetti and Paserman (2015) and Oliveti, Paserman, and Salisbury (2017) study IM in income in the United States

of children to parental outcomes; as such most earlier works rely on relatively small samples.[2] Of most relevance to our paper are the recent studies of Card, Domnisoru, and Taylor (2018) and Chetty, *et al.* (2014) and Chetty and Hendren (2018a,b). Card, Domnisoru, and Taylor use census data from the entire US population in 1940 to map educational mobility by looking at children residing with at least one parent (as we do). The large sample they work with allows them to focus on individuals aged $14-18$, when the overwhelming majority of children still reside with their parents and are at an age that allows them to meaningfully assess educational attainment. They show rising mobility during the first-half of the 20th century, which however differs across race and states. Chetty *et al.* (2014) provide a mapping of IM in income across US counties and explore its correlates running simple univariate specifications to understand the vast regional differences. Chetty and Hendren (2018a,b) use matched parents-children administrative tax records of moving families to isolate the effect of neighbourhood exposure on mobility from sorting. Our main contribution to this body of research is looking within Africa. We take a panoramic approach that not only constructs measures of educational mobility across 23 African countries, but also tries to isolate the role of regions (building on Chetty and Hendren, 2018a,b)) and explore the correlates of mobility (as Chetty *et al.* 2014)).

Our paper also relates to the vibrant research agenda on African development. There has been a growing interest in Africa both in properly measuring well-being, poverty, and output (Young (2012), Pinkovskiy and Sala-i-Martin (2014)) and on the correlates of African development (e.g., Henderson *et al.* (2018)). The literature has moved from mostly cross-country approaches focusing on national features (e.g., Collier and Gunning (1999), Bates (2006), Easterly and Levine (1997)), to within-country analyses that connect Africa's contemporary development to its colonial and pre-colonial past. This body of research has uncovered strong evidence of historical continuity as well as instances of rupture in the evolution of the economy and polity (Michalopoulos and Papaioannou (2018) provide a comprehensive review).[3] Our contribution to this research agenda is multifold. First, on measurement, we use census data and trace

from 1850 till 1940. Charles and Hurst (2003) use PSID data to estimate intergenerational persistence in wealth across US households. Alesina, Stancheva, and Teso (2018) compare actual social mobility and perceptions in several industrial countries. Early studies on IM in education include Bowles (1972), Blake (1985), and Spady (1996). More recently, Hertz et. al. (2007) estimate country-level IM coefficients for various cohorts across 42 countries. Hilger (2015) calculates educational IM in the United States since WWII. Azam and Bhatt (2015) and Golley and Kong (2013) estimate IM in education in India and China, respectively.

[2]In parallel and independent work, Narayan et al. (2018) measure intergenerational mobility in education and income around the world using survey data. Our work differs from their because (a) we use millions of observations from census data rather than thousands from survey data; (b) we deal with co-habitation selection directly by focusing on 14-18 year old individuals; (c) we examine IM across regions inside countries; (d) distinguish IM by gender and urban/rural residence; (e) we employ migrants to identify the causes of IM; (f) we document geographic, historic, and contemporary correlates of IM.

[3]Nunn (2010, 2014), Cage and Rueda (2016, 2017), Wantchekon, Klasnja, and Novtna (2015), Okoye and Pongou (2014), Mantovanelli (2014), Jedwab, Meier, and Moradi, (2017) and Huillery (2010) examine the role of Christian missions and colonial investments in human capital played in Africa's development, among many others. Building on the country case evidence of Kerby, Jedwab and Moradi (2017) in Kenya, Jedwab and Moradi (2016) on Ghana, and Okoye and Pongou (2017) in Nigeria, Jedwab and Storeygard (2017) examine the role of colonial roads and railroads on spatial development across Africa. Acemoglu et al. (2016) focus on the indirect colonial rule in Sierra Leone, while Lowes and Montero (2017) study King Leopold's concessionary agreements in Congo. Michalopoulos and Papaioannou (2016) connect ethnic partitioning during Africa's Scramble to contemporary conflict. Starting with Nunn (2008) many works trace various aspects of Africa's underdevelopment to the slave trade epoch. Other studies uncover the legacy of pre-colonial institutional, cultural, and economic features for contemporary development (e.g., Gennaioli and Rainer (2007), Michalopoulos and Papaioannou (2014), Michalopoulos, Putterman and Weil (2017), Mosconna, Nunn, and Robinson (2017), Fenske (2015)).

the evolution of education and mobility across Africa since independence. We are thus able to provide mappings of the land of opportunity (and education across cohorts) covering 2,440 districts across 23 African countries. Second, the analysis that exploits within-household variation from multi-child families that have moved during a child's formative years provides evidence that locations have strong causal effects on social mobility. This finding adds to the recent research that stresses the role of historical factors, but without accounting for sorting. Our analysis shows that sorting and migration are important features that future research should seriously consider. Third, the uncovered strong inertia between literacy and mobility provide large-scale evidence consistent with this research's main result, that pre-independence features correlate strongly with contemporary development. Fourth, using a common simple estimating framework we find that colonial investments in education-health and railroads are strongly correlated with educational mobility. And while this association does not have a causal interpretation, together with the strong link between literacy and educational mobility and the equally significant association between colonial investments and literacy at independence, suggests that by shaping initial at-independence conditions, colonial investments have lasting legacies most likely because economic activity concentrates into these places (see also Jedwab and Storeygard (2017) and Kerby, Jedwab, and Moradi (2016) for a similar argument).

Structure The paper is organized as follows. In Section 2 we present the census data on educational attainment and detail the construction of the intergenerational educational mobility measures. Section 3 starts with a portrait of the distribution of IM across countries and gives the main cross-country patterns. Then it provides a detailed mapping of the land of opportunity across African regions and established the tight link between upward mobility and the stock of literacy of the previous generation. Section 4 digs deeper into the literacy - social mobility nexus; exploiting within-household variation on children's place of birth, the analysis shows significant and strong regional exposure effects. Section 5 explores the geographic, historical and contemporary correlates of educational mobility. In Section 7 we summarize and discuss directions for future research.

2 Data

In this Section, we first motivate our focus on education. Second, we discuss the Census data. Third, we present the methodology in compiling statistics of absolute intergenerational mobility in education (social mobility). Fourth, we discuss how we account for cohabitation selection.

2.1 Why Education?

We focus on IM in education for several reasons. First, income and wealth data are scant, available only for a tiny share of the African population and only for a handful of countries. For instance, Alvaredo *et al.* (2017) report that for countries like Ghana, Kenya, Tanzania, Nigeria, and Uganda, the income data encompass less than 1% of the adult population. Likewise, consumption data for African countries are noisy, cover small samples, and are not spatially disaggregated. Complicating things further, the share

of the underground economy in Africa is large (La Porta and Shleifer (2008, 2016)) and tax evasion is rampant. In contrast, education data are available through various African censuses since the late 1960s. Not only measurement error in educational attainment is a lesser concern compared to that of reported income or wealth, but the education data are available at a fine temporal and geographic resolution.

Moreover, education is useful in mapping and studying intergenerational mobility, as people tend to complete (primary) schooling by (15) $18-20$ and so, unlike lifetime earnings, the analysis can start when adults are relatively early in the life-cycle. Second, education is strongly correlated with income/wealth both across countries (e.g., Barro and Lee (2014)) and regions (Gennaioli *et al.* (2013, 2014)); a large body of research in labour economics shows that education causally affects lifetime income (Card (1999), Krueger and Lindahl (2001)). Individual (Mincerian) returns to schooling are sizeable and possibly larger in low-income countries. [4] Third, besides wages, education is related to the quality of life and people's aspirations. In Appendix A using data from the Demographic and Health Surveys (DHS) and the Afrobarometer Surveys, we show that education positively correlates with various proxies of well-being across Africa. These include living conditions, child mortality and fertility, hopes and aspirations, attitudes toward domestic violence, and proxies of political and civic engagement.

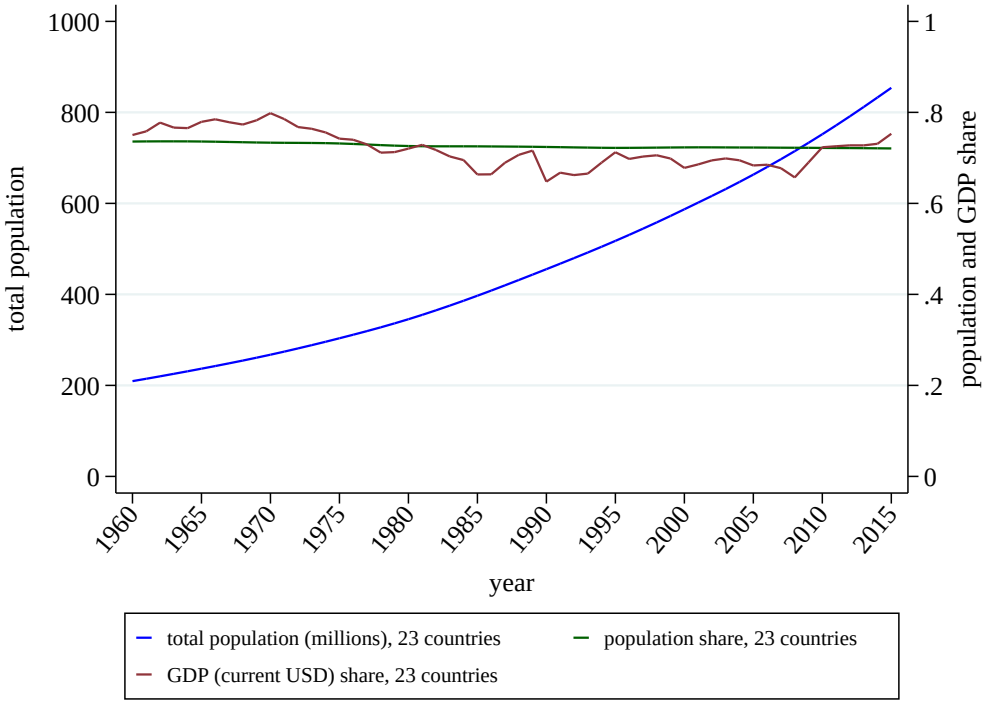
2.2 Sample Characteristics

Our analysis is based upon individual records, retrieved from 63 national censuses from 23 African countries: Botswana, Burkina Faso, Cameroon, Egypt, Ethiopia, Ghana, Guinea, Kenya, Liberia, Malawi, Mali, Morocco, Mozambique, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, South Sudan, Tanzania, Uganda, and Zambia. These data are available from IPUMS (Integrated Public Use Microdata Series) International. This database, hosted at the University of Minnesota Population Centre, takes representative samples from national censuses (typically 10%), harmonizes the data series and makes them available in the public domain.[5] Appendix Table B.1 gives information on our sample (countries, census years, coverage rates, and the number of individuals, provinces and districts). Figure 2.1 shows the evolution of the coverage. As of 2015, the countries included in the sample were home to slightly more than 850 million people, representing 72 percent of Africa's population and 75 percent of its GDP.

[4] See Psacharopoulos (1994), Caselli (2014), Patrinos (2014). Young (2013) reports Mincerian returns in the range of 11.3% (OLS) to 13.9% (2SLS) in a sample of 14 Sub-Saharan African countries with data on labour income from the Demographic and Health Surveys. These estimates are higher than in 11 non-SSA low income countries [range between 8.7% (OLS) and 10.4% (2SLS)] and the "consensus" estimate of 6.5%-8.5% in high income countries. Caselli (2016) reports lower Mincerian returns in Sub-Saharan Africa (around 8.5%), though in line with the earlier work of Psacharopoulos he also estimates a negative relationship between Mincerian returns and years of schooling (which is steeper in 1995 as compared to 2005). Moreover, Patrinos (2014) estimates higher Mincerian returns in SSA (12.5%) compared to the rest of the world (9.7%).

[5] The data from Nigeria come from household surveys conducted in consecutive years over $2006-2010$. We aggregate these yearly observations and count them as one census-year.

Figure 2.1: Total population, (African) GDP share covered by countries in the sample



Appendix Table B.4 gives further sample details. To estimate IM, we match individuals to their parents, using Census information for individuals of various generations who cohabitate.[6] Since not all individuals live with their parents, the total sample shrinks.[7] The cleaned dataset includes information on $19,039,076$ " young" individuals who cohabitate with at least one member of an "older" generation and are older than 14 years old. Estimating IM based on a sample of individuals, who reside with their parents, raises "cohabitation selection" concerns that we discuss below. IPUMS reports information on respondents' current residence allowing us to assign individuals to "coarse" and "fine" current administrative units. Districts are typically admin-2 divisions, though in some countries these are admin-3 areas (e.g., Sudan or Mali). Provinces are larger, almost always admin-1 areas (e.g., provinces in South Africa or states in Nigeria).[8] We have information from 346 provinces and $2,444$ districts across the 23 countries.

[6]Observations from three early censuses, Burkina Faso 1985, Kenya 1979, and Liberia 1974 drop out because there are no household identifiers that would allow us to assign individuals to different generations. For households with three or more generations, an individual's education can appear both as the education of an "old" generation (say vis a vis one's children) and as the education of a "young" generation (say vis a vis one's parents).

[7]We have dropped individuals without schooling information; while this is not a serious issue for most countries, it is for Ethiopia in 2007, and Burkina Faso in 1985, where respectively around 85% and 45% of the observations have missing schooling data. We are also unable to use the Burkina Faso 1985 census, since it lacks the variable "relationship to household head", which allows us to assign individuals to generations. We are also forced to drop Kenya's 1979 census and Liberia's 1974 census, since they lack household- in addition to individual identifiers.

[8]For Botswana and Nigeria, IPUMS reports just one administrative unit, "Districts" in Botswana and "States" in Nigeria; we thus use this aggregation both for districts and for provinces. In a few instances (in Ghana after 1984, in Burkina Faso in 1985, in Ethiopia in 1984, in Malawi in 1987, and in South Africa after 1996) the number of districts and regions changes between censuses in given country, as administrative boundaries are sometimes redrawn. For our analysis, we have harmonized administrative boundaries.

For more than 15 million individuals from 21 countries, we also have information on place of birth as well as their residence at the time of the census, which allow us to assess their migrant status. For a further subset of 10.8 million individuals, we additionally have information about the timing of their move, if any (see appendix table B.2 for details).

3 Methodology

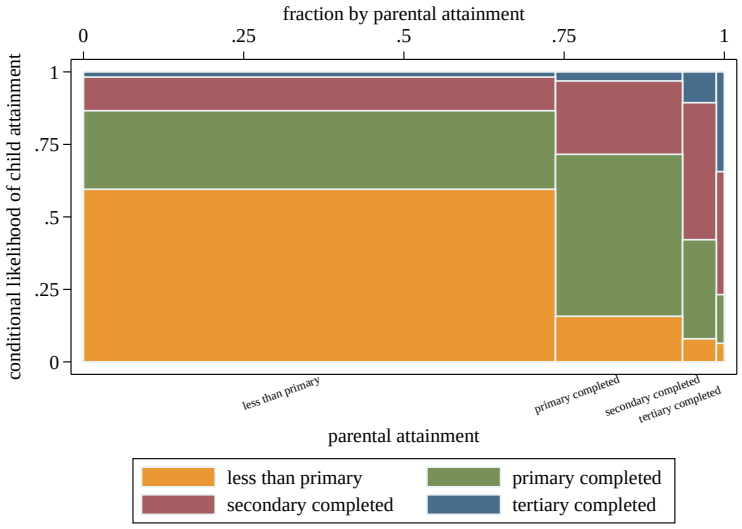
We measure IM as the transmission of education from " old" to "young" within a household. Following Chetty *et al.* (2014), Chetty and Hendren (2018a,b) and Card *et al.* (2018) we construct measures of *absolute* intergenerational mobility that reflect the likelihood that young individuals (children) acquire higher educational attainment than the older generation (parents).[9] The main attainment categories are: no schooling, less than primary, some or completed primary, some or completed secondary, and some or completed tertiary. Individuals with incomplete primary are assigned to less-than-primary, individuals with incomplete secondary to completed primary and individuals with incomplete tertiary to completed secondary. The education of the old assigned to individual i is the average attainment of individuals one generation older than individual i in the same household. We round the previous generation's average attainment to the nearest integer.

We obtain 4 by 4 matrices of "absolute" IM in education. Figure 3.1 (a) shows the Africa-wide transition matrix using all censuses, while Figures 3.1 (b) and (c) reproduce the transition matrix for Tanzania and Mozambique, respectively. To estimate transition likelihood for the completion of schooling beyond secondary we use a different sample restriction than the 14-year old age cutoff we employ in the rest of the paper: First, we require individuals to be at least 18, and second, we require them to be at least 9 years older than their years of schooling. Since schooling usually begins at age 6, this gives a reasonable 3-year buffer to make sure that individuals' educational attainment is not misclassified. We use these sample restrictions only for this initial inspection of the data and do not rely on it anywhere else.

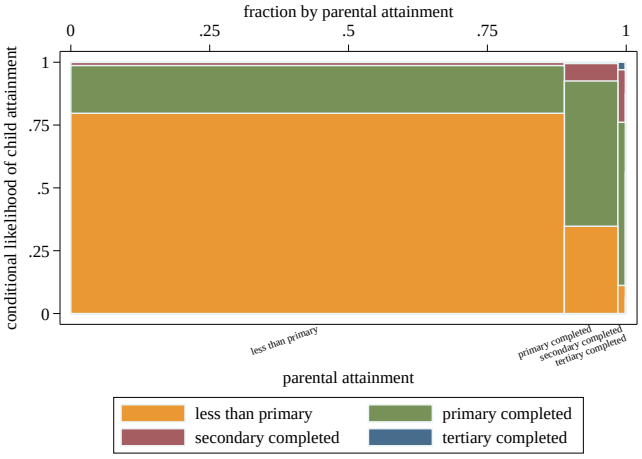
[9]The earlier literature has often relied on measures of relative intergenerational mobility (Black and Devereux, 2011). These are based on regressing the schooling of the young to the schooling of their parents, controlling also for demographic features and adding cohort and census fixed effects. In the earlier draft of the paper, we also used relative IM measures. As the results are similar to the ones with absolute mobility, and as it is apparent from Figure 3.1 (a) below educational mobility in Africa is primarily about the transition from zero schooling to primary education, we focus on the latter.

Figure 3.1: Visualization of transition likelihoods

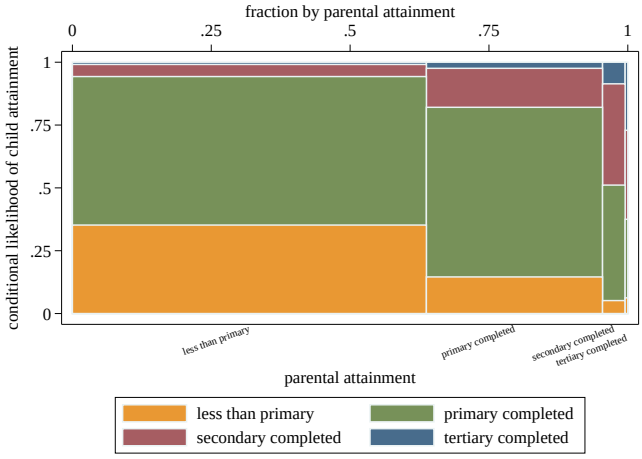
(a) Africa



(b) Mozambique



(c) Tanzania



The height of each cell in the plot indicates the probability that the child has the educational attainment of the respective color conditional on his/her parents having the educational attainment shown at the bottom of each bar (a group of cells). The width of each bar indicates how many parents have the educational attainment in question. Across Africa roughly 75% of the "old" generation has not completed primary schooling; with only 1.2% of the "old" having tertiary education. However, 26% of Africans whose parents have not completed primary schooling, manage to complete primary education, 12% finish high-school and 2% even manage to get a college degree. [10]

[10] Please note that since we drop younger individuals in censuses in the 1990s and 2000s so as to allow Africans to complete schooling, the statistics regarding the likelihood of a child completing tertiary education does not capture the sizeable expansion of tertiary education during the last 20 years.

Since three-fourths of "old" Africans did not have any schooling, in the rest of the paper we focus on the likelihood that kids from parents without any schooling or less than completed primary (that we label as "illiterate") manage to complete primary education (we label them as " literate"). We construct absolute IM measures at the country level and for each country at the admin-1 province and at the admin-2 district level.

To maximize coverage we pool information from the different censuses and birth-cohorts and then define the following indicator variables:

- $\mathbb{I}^{lit}_{0,ibct} = 1$ if the parent of individual i born in birth-decade b in country c and observed in census-year t is literate and zero otherwise.
- $\mathbb{I}^{lit,illit}_{1,ibct} = 1$ if a child i born to illiterate parents in birth-decade b in country c and observed in census-year t is literate and zero otherwise. Again, we define literacy as having completed at least primary education.

Pooling observations across all countries and censuses, we run the following regressions:

$$\mathbb{I}^{lit}_{0,ibct} = \alpha^o_c + [\gamma^o_b + \delta^y_b + \theta_t] + \epsilon_{ict} \tag{3.1}$$

$$\mathbb{I}^{lit,illit}_{1,ibct} = \alpha^y_c + [\gamma^o_b + \delta^y_b + \theta_t] + \epsilon_{ict}, \tag{3.2}$$

where $\mathbb{I}^{lit}_{0,ict}$ and $\mathbb{I}^{lit,illit}_{1,ict}$ are the indicators for parental and child education, respectively. We estimate specification 3.1 in the full sample of individuals for which we observe previous-generation education; and we estimate specification 3.2 for all children of illiterate parents. This ensures that the country fixed effects (α^y_c) can be interpreted as conditional proportions – that is, we want to know what proportion of children of uneducated parents become educated. To account for unobserved factors we also estimate specifications conditioning on birth-decade fixed effects (separately for the "young" (δ^y_b) and the "old" (γ^o_b)) and census-year fixed effects (θ_t).

Then, we run similar specifications for each country and estimate corresponding measures of average parental education (α^o_r and α^o_e) and absolute IM in education (α^y_r and α^y_e) at the region level. Specifically, we estimate country-by-country:

$$\mathbb{I}^{lit}_{0,itbcr} = [\alpha^o_r + \alpha^o_e] + [\gamma^o_b + \delta^o_b + \theta_t] + \epsilon_{ict} \tag{3.3}$$

$$\mathbb{I}^{lit,illit}_{1,itbcr} = [\alpha^y_r + \alpha^y_e] + [\gamma^o_b + \delta^o_b + \theta_t] + \epsilon_{ict}, \tag{3.4}$$

once unconditionally, and once conditioning on birth-cohort effects for old and young, as well as census-year fixed effects.

3.1 Cohabitation Selection

We can only estimate mobility of individuals who reside with their parents. This raises concerns of sample selection if the transmission of education between parents and kids who live apart systematically differs from that of co-resident parents and kids. By itself, this should push us to include only young children in

the sample. The problem is, of course, that young children may not have completed their schooling, so the younger we make the sample, the greater the risk of misclassifying individuals as "less-than-primary" when in fact they would complete primary education one or two years after we observe them in the census.

We deal with this tension between cohabitation selection and education misclassification in two ways. To address cohabitation selection, we follow Card, Domnisoru, and Taylor (2018) and begin by estimating IM for the sample of individuals aged $14-18$ for whom the in-sample co-residence rate is 86% over the entire sample.

Table 3.1: Co-residence rates ages 14-18 and 14-100 vs. age 8, education data observed for all

country	age 8	age 14-18	age 14-100	Δ^8_{14-18}	Δ^8_{14-100}
Burkina Faso	96.75	78.47	27.94	-18.90	-71.12
Botswana	89.52	79.54	35.29	-11.14	-60.57
Cameroon	96.33	79.12	29.55	-17.86	-69.32
Egypt	98.27	94.62	36.99	-3.72	-62.36
Ethiopia	98.02	83.24	27.79	-15.08	-71.65
Ghana	95.05	88.03	36.75	-7.38	-61.33
Guinea	94.38	77.71	29.11	-17.67	-69.16
Kenya	96.55	86.76	31.28	-10.13	-67.60
Liberia	95.99	90.63	35.52	-5.58	-63.00
Morocco	99.37	97.14	49.82	-2.24	-49.87
Mali	97.49	82.25	34.23	-15.63	-64.88
Mozambique	95.77	79.85	25.42	-16.63	-73.46
Malawi	99.92	85.35	22.24	-14.58	-77.75
Nigeria	93.97	86.68	30.58	-7.76	-67.45
Rwanda	94.53	86.57	32.30	-8.42	-65.83
Sudan	96.42	84.38	38.15	-12.49	-60.44
Senegal	97.98	92.26	47.21	-5.84	-51.81
Sierra Leone	91.58	82.28	38.89	-10.15	-57.53
South Sudan	95.58	88.23	37.80	-7.69	-60.45
Tanzania	96.15	87.05	28.86	-9.46	-69.99
Uganda	97.68	80.70	25.81	-17.39	-73.58
South Africa	88.47	83.03	35.02	-6.15	-60.41
Zambia	96.79	86.48	35.63	-10.64	-63.19
overall	94.86	86.07	33.37	-9.27	-64.82

This table shows the number of individuals of different age ranges for whom previous generation education as well as their own is observed as a percentage of all individuals with data on their own educaction as well as their relationship to the household head. The latter does not exclude single-person households, since these individuals will be labelled "head". The columns titled Δ^8_{m-n} show the *proportionate* reduction in the percentages relative to individuals aged 8.

We then follow the estimates with the restricted $14-18$ sample with the full sample of individuals aged 14 and above (co-residence rate is 33%) and compare estimates of IM from the two sample to gauge the extent to which our results may be driven by cohabitation selection.

To address the problem of education misclassification, we use census information on individual school enrolment to upward-correct educational attainment of individuals close to completing primary education. Specifically, if individuals have four or five years of schooling at the time we observe them and have

educational attainment of "less-than-primary", we record their attainment as "completed primary" in a second "student-corrected" educational attainment variable. We then present results with and without this student-correction.

4 Social Mobility across African Countries and Regions

In this section we first present the country-level estimates of social mobility and the main cross-country patterns. Second, we report the newly-constructed statistics of social mobility across African regions, illustrate the land of opportunity in Africa, and discuss the main regional patterns.

4.1 Country-level Statistics and Patterns

4.1.1 Educational IM across African Countries

Table 4.1 reports the country-level estimates of IM (in (1)-(4)), alongside some sample statistics (in columns (5)-(8)).

Table 4.1: Country-level estimates of IM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IM estimates				N / share children of illiterate parents			
age range	14-18	14-100	14-18	14-100	14-18	14-100	14-18	14-100
fixed effects	none	none	y, b	y, b	N with e_0 obs	N with e_0 obs	share with $e_0 = 0$	share with $e_0 = 0$
South Africa	0.779	0.739	0.591	0.314	986,330	2,788,228	0.31	0.39
Botswana	0.683	0.634	0.597	0.306	42,863	113,479	0.47	0.57
Egypt	0.630	0.600	0.501	0.196	2,073,589	4,737,281	0.71	0.74
Nigeria	0.615	0.664	0.330	0.174	38,312	75,531	0.47	0.48
Tanzania	0.579	0.618	0.544	0.362	728,341	1,423,114	0.45	0.54
Ghana	0.540	0.488	0.285	0.081	441,878	1,081,589	0.51	0.58
Cameroon	0.485	0.467	0.281	0.070	220,554	482,901	0.54	0.59
Zambia	0.446	0.469	0.185	0.073	286,385	562,255	0.37	0.43
Kenya	0.429	0.508	0.113	0.030	562,592	1,098,923	0.46	0.52
Morocco	0.400	0.339	0.140	-0.093	381,085	1,155,402	0.88	0.92
Uganda	0.328	0.363	0.319	0.104	280,991	487,490	0.57	0.63
Rwanda	0.300	0.335	0.284	0.064	140,607	255,426	0.82	0.84
Senegal	0.251	0.246	0.252	0.019	124,596	336,600	0.80	0.84
Sierra Leone	0.238	0.236	-0.139	-0.252	38,245	94,108	0.71	0.76
Liberia	0.221	0.317	-0.032	-0.178	25,494	59,015	0.52	0.60
Guinea	0.213	0.197	0.092	-0.189	67,335	158,275	0.88	0.90
Mali	0.200	0.179	-0.037	-0.247	237,472	535,515	0.88	0.89
Burkina Faso	0.159	0.155	-0.038	-0.284	181,097	334,996	0.91	0.91
Malawi	0.133	0.209	-0.076	-0.231	195,803	333,869	0.70	0.71
Ethiopia	0.120	0.145	-0.138	-0.287	777,437	1,385,041	0.94	0.94
Sudan	0.108	0.195	-0.145	-0.301	441,159	986,776	0.87	0.88
Mozambique	0.099	0.152	-0.183	-0.295	241,611	461,936	0.86	0.87
South Sudan	0.036	0.076	-0.215	-0.417	41,841	91,326	0.88	0.90
mean	0.335	0.350	0.137	-0.058				

Columns (1)-(4) of this table show country-level estimates of IM (likelihood that children of illiterate parents become literate). Higher numbers → higher IM. "age range" indicates the range of ages for children in the sample. "fixed effects" indicates whether estimates are unconditional or conditional on census-year (y) and birth-decade for young and old (b) fixed effects. Countries sorted by column (1). "mean" is the mean of the country-level estimates. Columns (5) and (6) give numbers of observations for which parental education is observed and columns (7) and (8) give the share of obervations among (5) and (6) with illiterate parents.

Column (1) gives the simple unconditional estimate in the full sample (individuals aged 14 and older), while column (2) reports the corresponding estimate when we look only at young individuals aged 14-18 so as to account for cohabitation selection (equation 3.1). The two series are strongly correlated (correlation 0.978). In columns (3) and (4) we show the corresponding estimates when we condition on birth-decade fixed effects for the old and the young and census year fixed-effects (equation 3.2). The correlation of the two series is 0.92 and 0.93, respectively.

Table 4.2: Correlation table, country-level measures of IM

		no FE 14-18	no FE 14-100	y,b FE 14-18	y,b FE 14-100
no FE	14-18	1			
no FE	14-100	0.978	1		
y,b FE	14-18	0.916	0.889	1	
y,b FE	14-100	0.931	0.936	0.974	1

Social mobility is on average quite low. Only a third of Africans born to parents without completed primary education have managed to complete primary education (or higher). But social mobility, as reflected by upward educational intergenerational mobility, varies widely across African countries. The likelihood that children of parents without education will manage to complete at least primary education ranges from below 10% in South Sudan to $74\% - 78\%$ in South Africa. There is an evident regional component to country differences in educational IM. We calculate the lowest IM in the Sahel, the region immediately South of the Sahara (Ethiopia, Morocco, Sudan, Burkina Faso and to a lesser extent in Mali and Senegal) and the highest in South Africa (Botswana, Zambia, and South Africa) with countries in Western and Eastern Africa being in the middle.

4.1.2 Literacy and Mobility across Countries

We examine the association between IM and the average education of the old. Table 4.3 gives the simple cross-country correlations. Figure 4.1 plots the relationship between country-level social mobility on the vertical axis and the share of literacy of the old generation on the horizontal axis (both estimated net of birth-decade and census-year fixed effects). Figure 4.1 (a) shows the relationship for $14 - 18$ year olds, while figure 4.1 (b) shows the relationship for all individuals 14 and older.

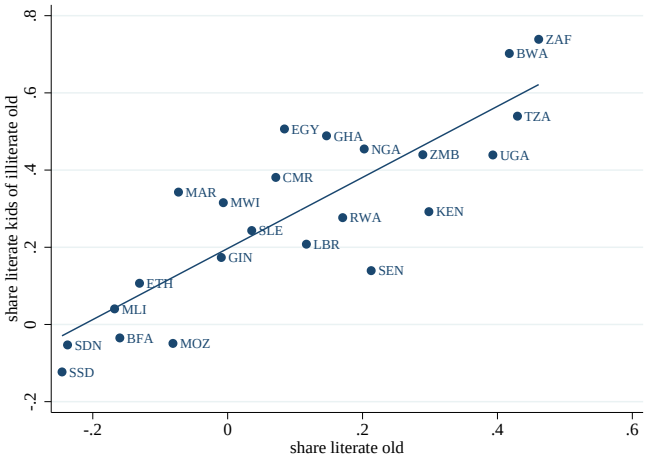
Table 4.3: Literacy and IM at the country-level

	(1) IM	(2) IM	(3) IM	(4) IM
share literate old	0.969***	1.125***	0.921***	1.112***
	(9.34)	(11.64)	(8.67)	(13.51)
R-squared	0.676	0.782	0.710	0.836
N	23	23	23	23
age-range	14-18	14-100	14-18	14-100
student correction	no	no	yes	yes

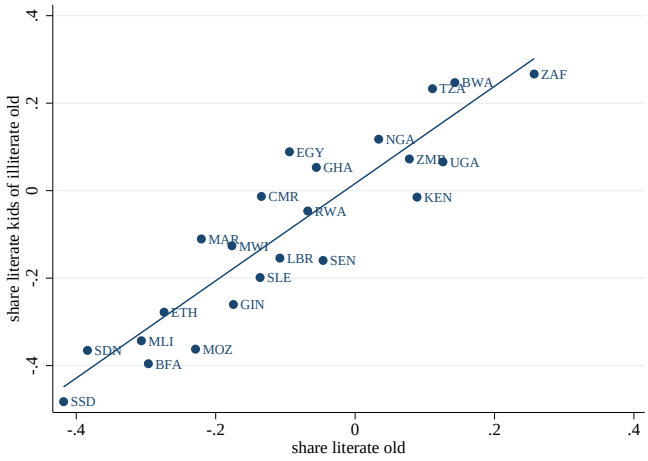
The dependent variable is the country-level share of literate kids of illiterate parents (estimated net of census year and old and young birth decade fixed effects). The independent variable is the country-level share of literate parents (also estimated net of fixed effects). t-statistics based on robust standard errors in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Figure 4.1: Literacy and IM at the country-level, with student correction

(a) year + birth-decade FEs, ages 14-18



(b) year + birth-decade FEs, ages 14-100



A strong positive association emerges between the education (literacy) of the old generation and absolute IM among the young. Referring back to the unconditional estimates reported in table 4.1, for example, in Ethiopia, Burkina Faso, North and South Sudan and Mozambique, where the share of literacy among the "old" generation is less than 20%, the likelihood that children from parents without schooling will complete primary is below or close to 20%. In contrast, the likelihood that children of parents without primary education will complete primary or higher education exceeds 60% in countries where the " old" generation is –on average– more educated, as, for example, in South Africa, and Nigeria.[11]

[11]The plots show associations conditional on birth-cohort fixed effects for both the young and the old and when we also add census-specific constants to account for trends, differential reporting, and other features. As we show in Appendix Figure C.1 and C.2, the patterns are similar when we do not condition on fixed effects as well as with and without the correction for individual student status.

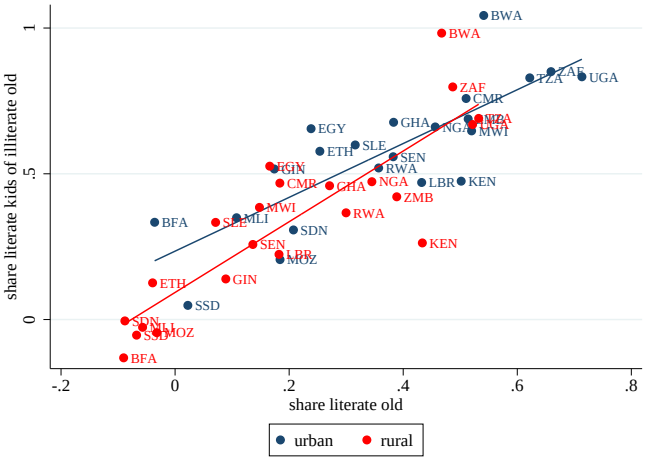
4.1.3 Heterogeneity

We exploit the richness of the census data to explore heterogeneity with respect to gender and the type of household residence.

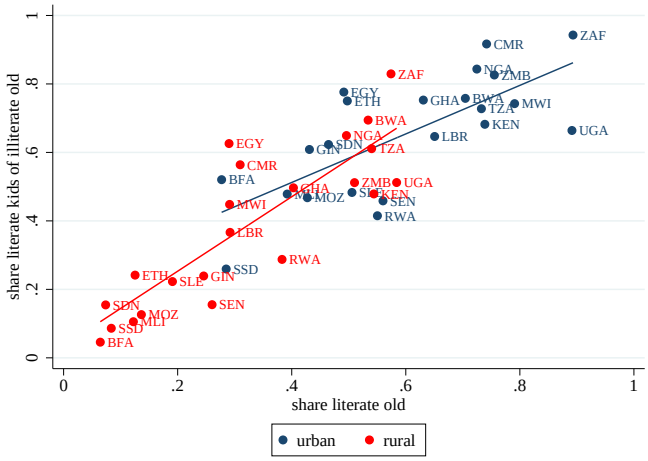
Rural-Urban In Figure 4.2 (a)-(b) and panels (A) and (B) of table 4.4 we distinguish between urban and rural households (using the IPUMS classification) looking again at the sample of individuals aged 14 and over and the sample of individuals aged $14-18$. As in the full sample, the likelihood that kids of illiterate parents will manage to complete at least primary education is positively related to the mean education of the "old" generation for both urban and rural households. In addition, there is clear rural-urban gap with the educational IM being lower for rural households. This pattern applies to all countries; it is the highest in countries with overall low levels of social mobility and literacy. For example, the gap is large in Ethiopia and Burkina Faso and quite limited in South Africa and Botswana. The positive association between absolute IM and the level of literacy of the old generation is quite steep for rural households, while for urban households the association is flatter. The correlation coefficient between literacy of the old and IM exceeds one for rural households (Panel A), while it is around 0.7 for urban households (Panel B). Moreover, the literacy of the old can explain around 70% of the variation of rural households IM, while the R^2 for urban households is around 0.5. This implies that inertia are especially strong in rural Africa, which could partly explain the exodus of millions of Africans from the countryside and the massive urban migration.

Figure 4.2: Literacy and IM at the country-level, urban/rural, male female with student correction

(a) year + birth-decade FEs, ages 14-18, urban/rural



(b) year + birth-decade FEs, ages 14-100, urban/rural

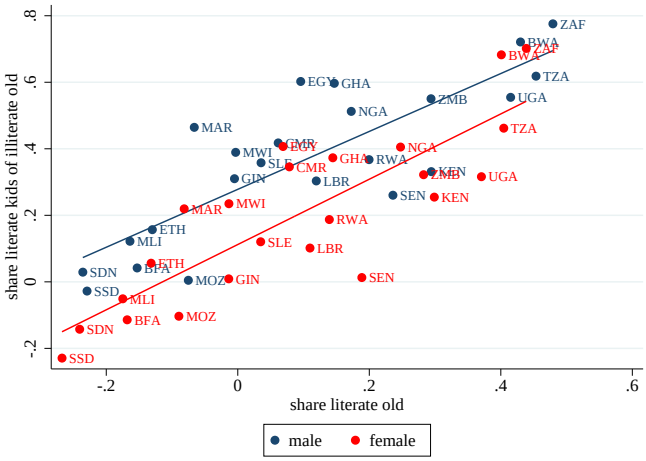


Gender In figure 4.3 (a)-(b) and panels (C) and (D) of table 4.4 we distinguish by gender of the children ("young" generation). The positive association between the likelihood of completing schooling for kids of parents without any education and the share of literate old generation is strong for both boys and girls. The R^2 is around 0.7 for both genders. The likelihood of exiting family illiteracy (absolute IM) is, however, higher for boys. Although the gender gap in IM is present in all countries, there is some

heterogeneity. The gap is relatively small in South Africa and Botswana and relatively large in Mali, Sudan, and Sierra Leone, countries with overall low levels literacy.

Figure 4.3: Literacy and IM at the country-level, urban/rural, male female with student correction

(a) year + birth-decade FEs, ages 14-18, male/female



(b) year + birth-decade FEs, ages 14-100, male/female

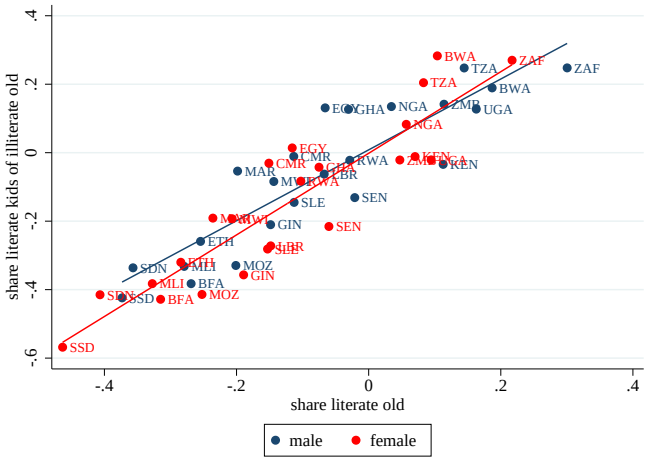


Table 4.4: Literacy and IM at the country-level, rural/urban, female/male heterogeneity

	(1) IM	(2) IM	(3) IM	(4) IM	(5) IM	(6) IM	(7) IM	(8) IM
	Panel A: rural subsample				**Panel C: female subsample**			
share literate old	1.140***	1.067***	1.213***	1.088***	1.019***	1.198***	0.981***	1.192***
	(6.80)	(8.56)	(7.71)	(8.68)	(8.47)	(10.25)	(8.22)	(12.38)
R-squared	0.655	0.730	0.759	0.732	0.690	0.776	0.713	0.828
N	22	22	22	22	23	23	23	23
	Panel B: urban subsample				**Panel D: male subsample**			
share literate old	0.849***	0.658***	0.923***	0.709***	0.926***	1.053***	0.870***	1.035***
	(4.36)	(3.86)	(6.17)	(4.56)	(9.80)	(10.81)	(8.78)	(11.70)
R-squared	0.457	0.449	0.662	0.528	0.648	0.757	0.684	0.799
N	22	22	22	22	23	23	23	23
age-range	14-18	14-100	14-18	14-100	14-18	14-100	14-18	14-100
student correction	no	no	yes	yes	no	no	yes	yes

The dependent variable is the country-level share of literate kids of illiterate parents (estimated net of census year and old and young birth decade fixed effects). The independent variable is the country-level share of literate parents (also estimated net of fixed effects). t-statistics based on robust standard errors in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

4.2 Regional Statistics and Patterns

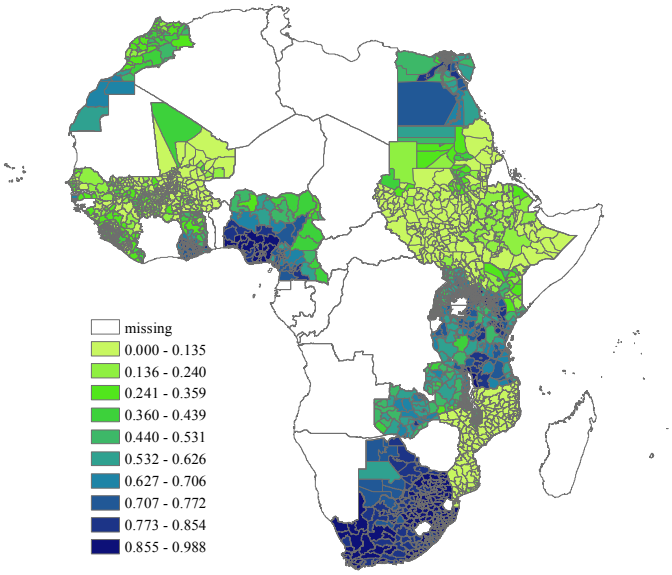
4.2.1 Where is the Land of Opportunity?

The richness of the Census data allows us estimating social mobility indicators at the region level.

Figure 4.4 (a) shows the distribution of absolute IM (the likelihood that children of illiterate parents manage to complete at least primary education across 2,444 African (admin-2) districts using the full sample (children aged 14 and older). Figure 4.4 (b) shows the distribution of absolute mobility for children aged 14-18. The maps give the unconditional likelihood that kids of parents without completed primary will complete at least primary schooling (equation (3.3) estimated country-by-country).

Figure 4.4: Pan-Africa: District-level estimates of IM

(a) Fraction of literate children of illiterate old; darker colors → higher IM, age range 14-18



(b) Fraction of literate children of illiterate old; darker colors → higher IM, age range 14-100

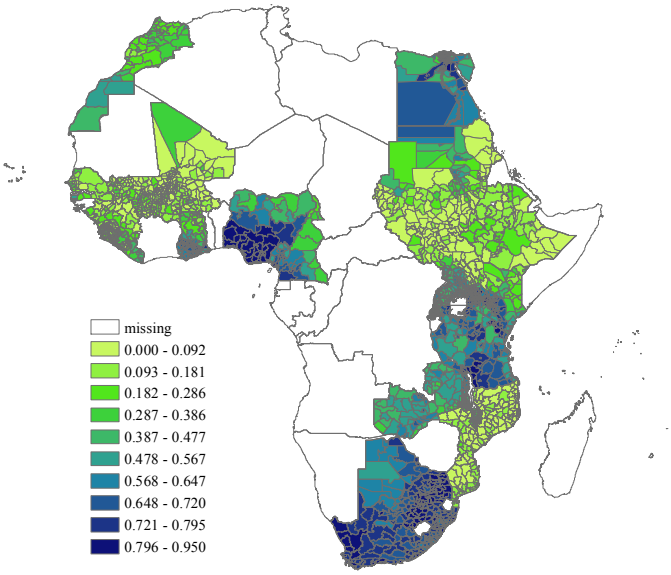


Table 4.5 gives summary statistics (mean, median, and range) by country. The cross-country average and the median are around 0.38. The percentage of the variance explained by the country constants is 0.70 (for the estimates conditional on year and birth-cohort fixed effects, the R^2 drops to 48%). The Table also gives the range of IM estimates for each country. There is considerable variability on IM across regions in a given country.

For example, Figures 4.5 (a) and (b) portray the variability in absolute IM across 102 admin-2 units in Ghana. While average IM in Ghana is 0.57 for 14-18 year olds, regional IM estimates range from 0.173 to 0.820. IM rates are quite low in the Northern and Central regions (below 0.5) and considerably higher in the Southern regions (higher than 0.7).

As Table 4.5 shows, this within-country variation in IM applies to almost all countries. For example, in Burkina Faso the average IM estimate of 0.123 for 14-18 year olds masks huge variability with regional IM estimates ranging from 0.024 to 0.49. In Uganda the range in IM across regions is even wider [0.013-0.666].

In South Africa and in Botswana not only IM is higher, but regional differences are less severe.

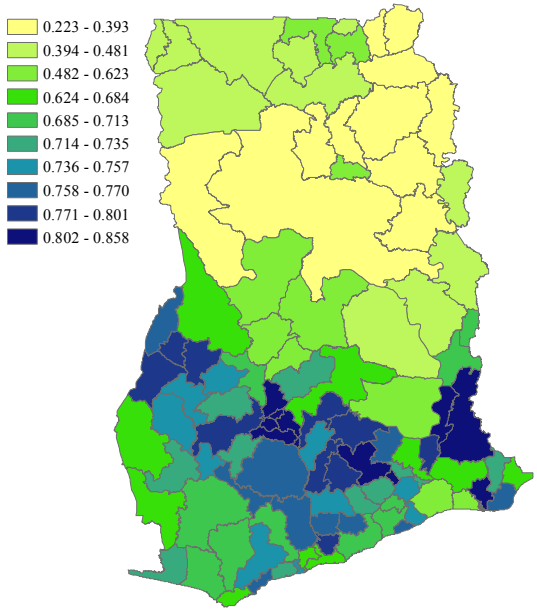
Table 4.5: Country-by-country summary statistics of unconditional district-level estimates of IM

	ages 14-18					ages 14-100				
country	mean	median	stdev	min	max	mean	median	stdev	min	max
Burkina Faso	0.123	0.116	0.074	0.024	0.490	0.112	0.107	0.072	0.017	0.476
Botswana	0.697	0.691	0.068	0.567	0.798	0.648	0.623	0.083	0.501	0.782
Cameroon	0.564	0.618	0.170	0.198	0.800	0.539	0.587	0.174	0.161	0.828
Egypt	0.665	0.676	0.104	0.391	0.913	0.626	0.633	0.101	0.368	0.900
Ethiopia	0.109	0.089	0.120	0.007	0.711	0.124	0.093	0.131	0.011	0.750
Ghana	0.570	0.625	0.158	0.173	0.820	0.516	0.591	0.169	0.122	0.754
Guinea	0.175	0.163	0.080	0.069	0.481	0.150	0.138	0.088	0.050	0.532
Kenya	0.496	0.517	0.189	0.046	0.875	0.557	0.617	0.203	0.051	0.893
Liberia	0.187	0.193	0.081	0.041	0.353	0.265	0.265	0.095	0.054	0.506
Morocco	0.419	0.400	0.139	0.156	0.697	0.336	0.325	0.122	0.112	0.615
Mali	0.142	0.126	0.093	0.014	0.555	0.115	0.098	0.080	0.011	0.527
Mozambique	0.084	0.061	0.067	0.016	0.321	0.118	0.082	0.097	0.025	0.511
Malawi	0.172	0.136	0.099	0.053	0.513	0.256	0.211	0.131	0.074	0.735
Nigeria	0.696	0.740	0.204	0.326	0.962	0.711	0.777	0.202	0.317	0.940
Rwanda	0.299	0.293	0.073	0.077	0.510	0.338	0.333	0.070	0.162	0.570
Sudan	0.148	0.096	0.139	0.001	0.600	0.214	0.141	0.193	0.003	0.763
Senegal	0.263	0.198	0.153	0.075	0.587	0.244	0.170	0.160	0.067	0.598
Sierra Leone	0.196	0.177	0.104	0.028	0.576	0.190	0.167	0.098	0.031	0.579
South Sudan	0.040	0.022	0.053	0.000	0.324	0.084	0.054	0.088	0.000	0.463
Tanzania	0.600	0.601	0.098	0.363	0.836	0.635	0.635	0.103	0.378	0.858
Uganda	0.357	0.353	0.122	0.013	0.666	0.392	0.393	0.121	0.017	0.714
South Africa	0.774	0.789	0.075	0.548	0.895	0.726	0.745	0.071	0.511	0.839
Zambia	0.450	0.440	0.123	0.261	0.758	0.460	0.455	0.118	0.277	0.763
overall	0.369	0.317	0.264	0.000	0.962	0.379	0.349	0.251	0.000	0.940

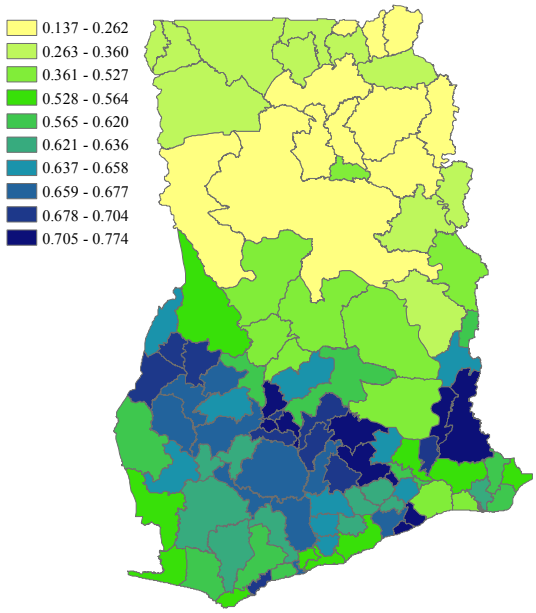
This table shows summary statistics for district level esimates of IM (estimated without fixed effects). The row "overall" shows the overall summary statistics for all districts in the sample.

Figure 4.5: Ghana: District-level estimates of IM

(a) Fraction of literate children of illiterate old; darker colors → higher IM, age range 14-18



(b) Fraction of literate children of illiterate old; darker colors → higher IM, age range 14-100

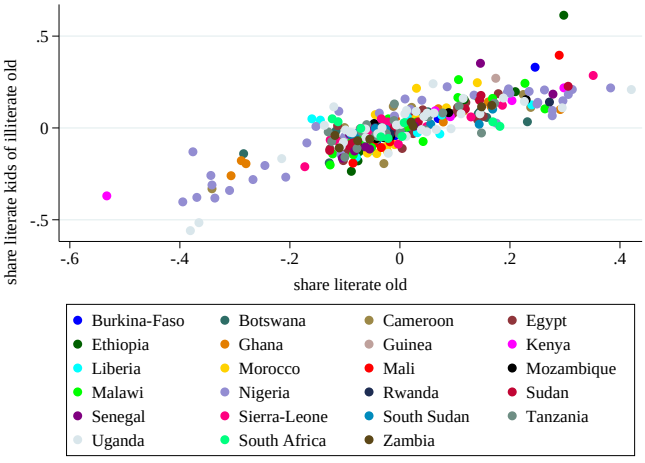


4.2.2 Literacy and Mobility at the Regional level

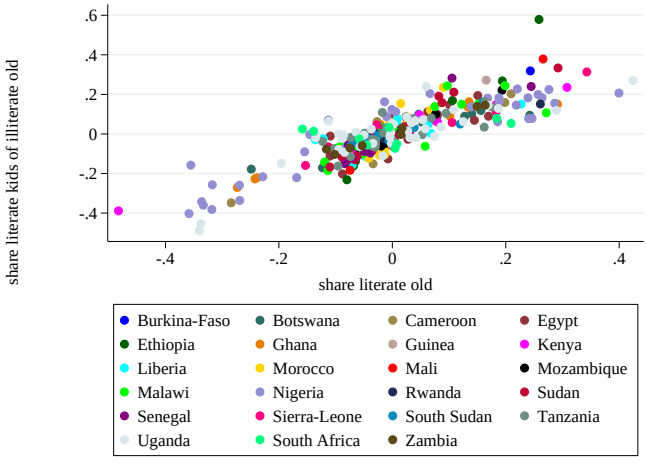
Following the cross-country analysis, Figures 4.6 (a) and (b) plot the association between IM and mean literacy rates across 345 admin-1 units (using different colors for provinces in different countries).

Figure 4.6: Literacy and IM at the admin-1 province-level

(a) y + b FEs, ages 14-18, scatter with country FEs



(b) y + b FEs, ages 14-100, scatter with country FEs



There is an evident positive association between the literacy of the " old" generation and absolute IM, a pattern that echoes the cross-country pattern. Table 4.6 reports the corresponding within-country across province associations.

Table 4.6: Literacy and IM at the province-level

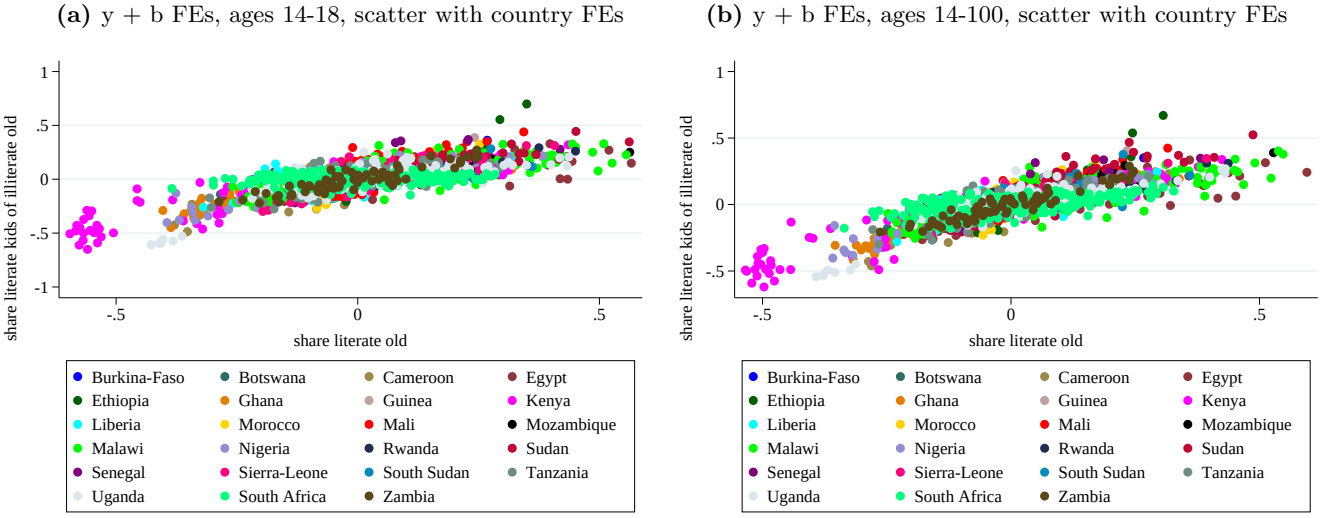
	(1) IM	(2) IM	(3) IM	(4) IM
share literate old	0.733***	0.857***	0.790***	0.889***
	(20.77)	(17.49)	(19.65)	(17.58)
R-squared	0.928	0.960	0.926	0.949
within-R-squared	0.725	0.770	0.698	0.750
N	345	345	345	345
country-FEs	yes	yes	yes	yes
age-range	14-18	14-100	14-18	14-100
student correction	no	no	yes	yes

The dependent variable is the admin-1 (province) -level share of literate kids of illiterate parents (estimated country-by-country net of census year and old and young birth decade fixed effects.) The independent variable is the province-level share of literate parents (also estimated country-by-country, net of fixed effects). All correlations net of country fixed effects. t-statistics based on standard errors clustered at the country-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

A 10 percentage points higher literacy rate is a associated with an increase of $7.3 - 8.9$ percentage points in the likelihood that kids of illiterate parents will manage to complete at least primary schooling. The within-R^2 is $0.7 - 0.75$, further showing the strength of the association.

In figures 4.7 (a) and (b) and Table 4.7 we report the correlation between literacy and educational mobility across $2,444$ admin-2/3 districts (see Figures 4.4 above).

Figure 4.7: Literacy and IM at the admin-2 district-level



The within-country correlation retains its significance, though the coefficient somewhat drops. A 10 percentage points increase in literacy of the "old" generation in the district is associated with a $6 - 7$ percentage point higher likelihood that the children of parents without any education will manage to

complete primary schooling or higher.[12] This pattern is pervasive, as it applies to all countries.

Table 4.7: Literacy and IM at the district-level

	(1)	(2)	(3)	(4)
	IM	IM	IM	IM
share literate old	0.603***	0.736***	0.663***	0.770***
	(11.57)	(11.28)	(11.79)	(11.46)
R-squared	0.917	0.956	0.924	0.948
within-R-squared	0.633	0.704	0.617	0.688
N	2440	2440	2440	2440
country-FEs	yes	yes	yes	yes
age-range	14-18	14-100	14-18	14-100
student correction	no	no	yes	yes

The dependent variable is the admin-2 (district) -level share of literate kids of illiterate parents (estimated country-by-country net of census year and old and young birth decade fixed effects.) The independent variable is the district-level share of literate parents (also estimated country-by-country, net of fixed effects). All correlations net of country fixed effects. t-statistics based on standard errors clustered at the country-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

4.2.3 Heterogeneity at the district-level

In panels (A) and (B) of table 4.8 we distinguish between rural (Panel A) and urban (Panel B) households. The correlation between educational IM and literacy of the "old" generation in the district is highly significant for both sets of households. But it is stronger for rural households.

In panels (C) and (D) of table 4.8 we estimate separate specifications for girls (Panel C) and boys (Panel D). While the link between the literacy of the "old" generation in the district and IM is strong for both genders, the correlation is somewhat stronger for girls; this suggests that the regional features, in this case the low levels of literacy, are especially detrimental for girls.

Table 4.5 shows the country-by-country relationship between IM and literacy at the admin-2 district level for the overall sample as well as the rural/urban and female/male sub-samples.

[12] For the admin-2 districts, the numbers are R^2 of 0.71 for simple OLS, 0.88 for country FEs (66% within R2), coefficient of -1 for OLS, -0.78 for country FEs.

Table 4.8: Literacy and IM at the district-level, rural/urban, female/male heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IM	IM	IM	IM	IM	IM	IM	IM
	Panel A: rural subsample				**Panel C: female subsample**			
share literate old	0.860***	1.057***	1.066***	1.138***	0.929***	1.056***	1.102***	1.210***
	(40.64)	(55.72)	(44.98)	(74.82)	(71.26)	(71.80)	(67.67)	(115.69)
R-squared	0.558	0.629	0.608	0.784	0.654	0.678	0.608	0.842
N	2365	2365	2365	2365	2440	2440	2440	2440
	Panel B: urban subsample				**Panel D: male subsample**			
share literate old	0.826***	0.762***	0.799***	1.051***	0.721***	0.936***	0.953***	1.068***
	(34.88)	(33.58)	(33.50)	(52.40)	(39.15)	(48.88)	(53.52)	(74.25)
R-squared	0.359	0.301	0.338	0.654	0.430	0.463	0.558	0.715
N	2375	2375	2375	2375	2440	2440	2440	2440
age-range	14-18	14-100	14-18	14-100	14-18	14-100	14-18	14-100
student correction	no	no	yes	yes	no	no	yes	yes

The dependent variable is the district-level share of literate kids of illiterate parents (estimated net of census year and old and young birth decade fixed effects). The independent variable is the district-level share of literate parents (also estimated net of fixed effects). t-statistics based on country-clustered standard errors in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table 4.9: District-level IM-on-literacy-coefficients by country

	ages 14-18					ages 14-100				
country	overall	female	male	rural	urban	overall	female	male	rural	urban
Burkina Faso	1.271	1.179	1.350	1.703	0.973	1.277	1.206	1.354	1.733	0.952
Botswana	0.503	0.430	0.573	0.552	0.511	0.720	0.717	0.734	0.755	0.448
Cameroon	0.822	0.955	0.725	0.574	0.556	1.100	1.193	1.046	1.051	0.753
Egypt	0.558	0.799	0.360	0.714	0.361	0.614	0.893	0.430	0.795	0.369
Ethiopia	1.784	1.795	1.767	3.845	1.495	2.066	1.987	2.161	4.749	1.595
Ghana	0.770	0.739	0.804	0.606	0.345	0.958	0.959	0.995	0.821	0.568
Guinea	1.536	1.268	1.807	2.485	0.322	1.757	1.416	2.040	3.347	0.797
Kenya	0.747	0.827	0.683	0.773	0.395	0.847	0.963	0.766	0.824	0.566
Liberia	0.483	0.403	0.497	0.438	0.028	0.707	0.636	0.730	0.766	0.127
Morocco	1.709	2.002	1.331			2.303	2.578	1.989		
Mali	1.390	1.326	1.410	1.471	1.073	1.467	1.355	1.534	1.135	1.152
Mozambique	0.594	0.617	0.567	0.988	0.710	0.976	0.945	0.998	1.680	1.077
Malawi	0.584	0.604	0.556	0.427	0.499	0.629	0.672	0.593	0.568	0.579
Nigeria	0.753	0.838	0.721	0.755	0.713	0.803	0.998	0.745	0.810	0.761
Rwanda	0.570	0.760	0.369	0.748	0.288	0.663	0.798	0.522	0.622	0.428
Sudan	0.841	0.919	0.768	1.044	0.756	1.398	1.446	1.356	1.862	1.161
Senegal	1.031	0.987	1.067	1.712	0.830	1.261	1.147	1.358	2.012	0.990
Sierra Leone	0.807	0.821	0.773	0.686	0.404	0.863	0.831	0.903	0.672	0.697
South Sudan	0.324	0.294	0.353	0.320	0.477	0.693	0.628	0.751	0.699	0.881
Tanzania	0.627	0.682	0.574	0.660	0.747	0.770	0.884	0.675	0.807	0.876
Uganda	0.751	0.872	0.635	0.935	0.362	0.778	0.926	0.649	0.924	0.490
South Africa	0.065	0.072	0.060	0.124	0.046	0.270	0.303	0.232	0.269	0.143
Zambia	0.826	0.889	0.768	0.563	0.057	0.912	1.000	0.822	0.606	-0.108
unweighted mean	0.841	0.873	0.805	1.006	0.543	1.036	1.064	1.017	1.250	0.696
overall	0.663	0.731	0.602	0.653	0.493	0.770	0.862	0.707	0.769	0.639

This table shows coefficients on district-level literacy of the "old" in country-by-country regressions with district-level IM on the LHS (both old literacy and IM estimated net of census-year and birth decade of young and old fixed effects). The row "unweighted mean" shows the simple average of coefficients in each column and the row "overall" shows the overall coefficients for all districts net of country fixed effects.

As for the level of mobility, there is significant variation in the association between parental literacy and IM at the district level across countries: 1-percent difference in parental literacy is associated with greater than 1 percent difference in Burkina Faso, Guinea, Ethiopia, Morocco, Mali, and Senegal. The association is much weaker in Botswana, Egypt, Liberia, Mozambique, Malawi, Rwanda, South Sudan, and particularly in South Africa. This indicates that the association between literacy and IM is not simply a result of the overall level of development of the country. A second result worth noting is that literacy is more predictive of differences in IM in rural areas than in urban areas almost everywhere except Malawi, South Sudan, and Tanzania. Male-female differences, on the other hand, show no clear pattern: for some countries, the association is stronger for girls, for others, for boys.

5 Identifying Regional Exposure Effects

To what extent does the literacy of the previous generation "cause" IM? We move a step towards testing this more stringent interpretation of the data by looking at migrants. We use two different strategies to identify location effects on educational outcomes; both approaches exploit variation among individuals who have moved between different places within African countries. First, we employ a relatively straightforward household fixed-effects strategy that exploits variation from multi-children families who have moved over time, subjecting therefore children in their early years to different regional features. Second, we follow the approach of Chetty and Hendren (2018a) that also exploits differences on the timing of moving of multi-kid families to identify regional exposure effects.[13]

5.1 Within-Family Estimates

We define as a migrant someone who is not born in his or her current place of residence. We focus on individuals in households in which there are at least two children and where at least one child is a migrant and at least one who is not a migrant. Given our focus on upward absolute social mobility, among those individuals, we focus on those who also have illiterate parents. Using this sample, we estimate the following specification:

$$\mathbb{I}^{lit}_{1,ihtbcr} = [\alpha_h +] \ \gamma_b^{\text{old}} + \delta_b^{\text{young}} + \theta_{ct} + \lambda \times \text{sex}_{ihtbrc} + \beta \times \overline{\text{EDOPP}}_r^{nm,ybfe} + \epsilon_{ihtbcr}. \tag{5.1}$$

$\mathbb{I}^{lit}_{1,ihtbcr}$ is an indicator that equals one if individual i born to illiterate parents in household h in birth decade b, in birth region r in country c and observed in census-year t is literate and zero otherwise. Crucially, given our focus on multi-children families, α_h is a household fixed effect that accounts for hard-to-observe family-specific features, related among others to ethnicity, religion, social background, aspirations, etc. γ_b^{old} $\delta_b^{\text{young}}$are birth-decade fixed-effects for parents (old) and children (young), respectively, while θ_{ct} denotes country-census-year fixed effects. $\overline{\text{EDOPP}}_r^{nm,ybfe}$ is a location (r specific) effect that captures the "opportunities" of the birth-place of individuals. We experiment with two types of location effects:

(a) $\overline{\text{EDOPP}}_r^{nm,ybfe} = \overline{\text{LIT}}_r^{nm,old,ybfe}$: The share of literate parents of individuals, both non-migrants, in region r estimated net of census-year and birth decade (for young and old) fixed effects.

(b) $\overline{\text{EDOPP}}_r^{nm,ybfe} = \overline{\text{IM}}_r^{nm,ybfe}$: The share of literate individuals of illiterate parents, both non-migrants, in region r, again estimated net of census-year and birth-decade fixed effects.

[13] For all countries, except Nigeria and Morocco, IPUMS records an individual's "birth place" in addition to the individual's place of residence (admin-1 or admin-2 region) at the time of the census. For many countries, these birth-places are not recorded at the same level of disaggregation as the variables which we use to define admin-2 districts. In some cases, birth places are recorded at admin-1 level whereas the current region of residence is recorded at admin-2 level. In other cases, both region of residence and birth place are available at the same level of aggregation. To assess migration status for all individuals, the two sets of regions – region of residence and region of birth – have to have the same set of values. Hence, even if places of residence have admin-2 level detail, we can only use admin-1 detail to assess migration status if birth-regions are recorded at admin-1 level. We choose the finest possible level of aggregation for all countries and end up with 1194 "birth-regions" (as compared to 2,444 districts before).

The coefficient of interest, β, is identified purely from within-household variation in birth place. A positive coefficient on $\overline{\text{LIT}}_r^{nm,old,ybfe}$ in equation (5.1) would indicate that comparing siblings within the same household of illiterate parents, the one(s) born in a region with higher literacy among the old is also more likely to be literate. A positive coefficient on $\overline{\text{IM}}_r^{nm,ybfe}$ would indicate that within the same household of illiterate parents children are more likely to become literate in places with greater absolute IM among the non-migrant permanent residents.

Table 5.1 presents several regressions with the different types of the "location effects". Even-numbered specifications report household fixed effects estimates, while odd-numbered columns report otherwise identical regressions but without the household fixed-effects. The comparison of the specifications allows gauging the role of selection.

Table 5.1: Household fixed effects estimates of location-of-birth effects with 2+children migrant families, maximum age = 21, 100, minimum age gap = 5

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IM	IM	IM	IM	IM	IM	IM	IM
non-migrant parental literacy, kids aged 14-21	0.522***	0.0646*						
	(11.86)	(1.77)						
non-migrant parental literacy, kids aged 14-100			0.729***	0.183***				
			(22.75)	(10.25)				
non-migrant IM, kids aged 14-21					0.693***	0.111**		
					(17.11)	(2.35)		
non-migrant IM, kids aged 14-100							0.745***	0.176***
							(31.17)	(9.00)
R-squared	0.177	0.705	0.184	0.668	0.186	0.705	0.192	0.668
within R-squared	0.029	0.008	0.054	0.021	0.040	0.008	0.063	0.021
N	18215	18215	265232	265232	18215	18215	265232	265232
number of birth regions	986	986	1185	1185	986	986	1185	1185
country-year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
y+o cohort FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
household FEs	No	Yes	No	Yes	No	Yes	No	Yes
maximum age	21	21	100	100	21	21	100	100
minimum age gap	5	5	5	5	5	5	5	5

The dependent variable is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$

While we do not observe the exact time when individuals moved between locations (see section 5.2 for an identification strategy that exploits this information), we focus on individuals in the same household born at least 5-years apart in age to increase the chances that they grew up and were educated in different environments[14]. Imposing the 5-year age-gap means that we cannot restrict the sample to individuals aged $14-18$. Instead, in columns (1), (2), (5), and (6) we look at individuals aged $14-21$. For completeness

[14]To be precise, we first find the youngest migrant and oldest non-migrant within each household × generation. If the age-gap between them (age of youngest migrant minus age of oldest non-migrant) is smaller than (positive) 5, we discard the oldest non-migrant and move to the next youngest individual within the household etc until we are left with non-migrant individuals at least 5-years younger than the youngest migrant. We similarly compare individuals to the second-youngest, etc. until the oldest migrant. We then compare which comparison leaves the greatest number of individuals per household and choose those individuals such that the oldest non-migrant is always at least 5-years younger than the youngest migrant.

in columns (3), (4), (7), and (8) we use all individuals aged 14 and older, as this allows maximizing the sample and investigating the potential cohabitation bias.

Let us start with the baseline specifications (young generation aged $14-21$). There is a strong cross-sectional correlation between the literacy rates of the "old" generation and absolute mobility (estimate of 0.5, column (1)). Likewise, absolute mobility of moving household children correlates strongly with educational mobility in the place of birth (estimate of 0.7, column (5)). When we add household fixed-effects to exploit within-family variation, the estimates fall, 0.065 in (2) and 0.11 in (6), but they retain statistical significance. This shows that the characteristics of the place of birth, literacy and social mobility, have an effect of the chance of children, whose parents do not have any schooling, to acquire literacy.

The inclusion of household fixed effects increases the model fit considerably (the R^2 jumps by a factor of three-four). And the point estimates, while remaining significant, fall considerably. These patterns suggests (a) that – unsurprisingly – unobservable household characteristics are an important determinant of social mobility and (b) that those household characteristics are correlated with the educational opportunities of birth regions. Illiterate parents who are educating their children sort themselves into higher literacy and higher IM environments; however, sorting does not completely explain the association between individual mobility and location-characteristics.

In the Appendix we tabulate country-specific tables (D.1) - (D.4) that allow examining whether some countries drive this association and exploring heterogeneity. These results are not driven by a few countries but apply quite generally to most countries in our sample.

Obviously these results are not easily generalizable to the overall population since families that have moved may be different from families that have not.

5.2 Childhood Exposure Effects (Chetty and Hendren (2018a))

To further identify childhood exposure effects, we adapt the method of Chetty and Hendren (2018a) that also exploits variation among migrants only, but allows for differences on the time of exposure. Specifically, we focus on children of illiterate parents who have moved to their current region of residence sometime between age 1 and age 18. Letting "o" denote origins and "d" denote destinations, we construct for every origin-destination pair two measures of the "educational opportunity gap" between destination and origin:

$$\Delta_{od} = \begin{cases} \text{either} & \overline{\text{LIT}^{\text{nm,old,ybfe}}}_d - \overline{\text{LIT}^{\text{nm,old,ybfe}}}_o \\ \text{or} & \overline{\text{IM}^{\text{nm,old,ybfe}}}_d - \overline{\text{IM}^{\text{nm,old,ybfe}}}_o. \end{cases} \tag{5.2}$$

Δ_{od} captures the difference between the destination and the origin region either in terms of the share of uneducated permanent resident old or in terms of IM of the permanent resident young. The educational opportunity gaps then enter on the RHS of the following regression:

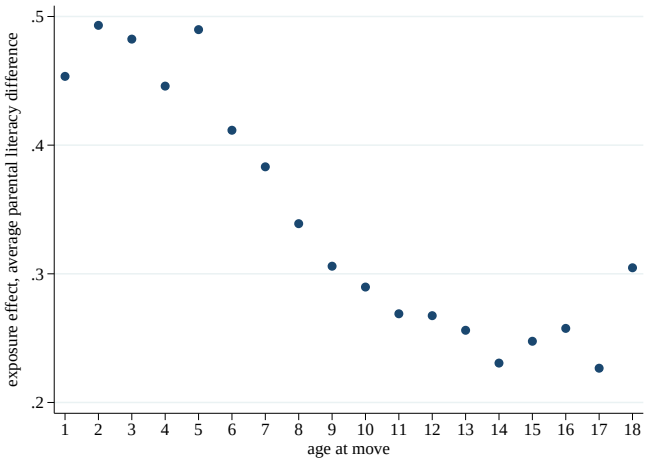
$$\mathbb{I}^{lit,illit}_{1,itbcod} = \alpha_o + \gamma^o_b + \delta^o_b + \theta_t + \sum_{m=1}^{18} \beta_m \times \mathbb{I}(m_i = m) \times \Delta_{od} + \epsilon_{i,itbcod}. \tag{5.3}$$

The idea behind equation (5.3) is that if individuals move from a place with worse to a place with better educational opportunities ($\Delta_{od} > 0$), and exposure effects matter for educational outcomes, the earlier on in life the move occurs, the greater the effect on the eventual outcome (Chetty and Hendren (2018a)). Note that the age-specific slopes, β_m in equation 5.3, are identified even in the presence of sorting; i.e., illiterate parents with higher latent propensity to educate their children are more likely moving to higher opportunity environments. The identifying assumption of the above regression equation is that the *timing* of the move between o and d is not correlated with this latent propensity. In other words, illiterate parents who were always going to educate their children are allowed to move from worse to better environments *on average* compared to illiterate parents who were not going to do so, but the ambitious parents should not move to better environments *earlier rather than later*.

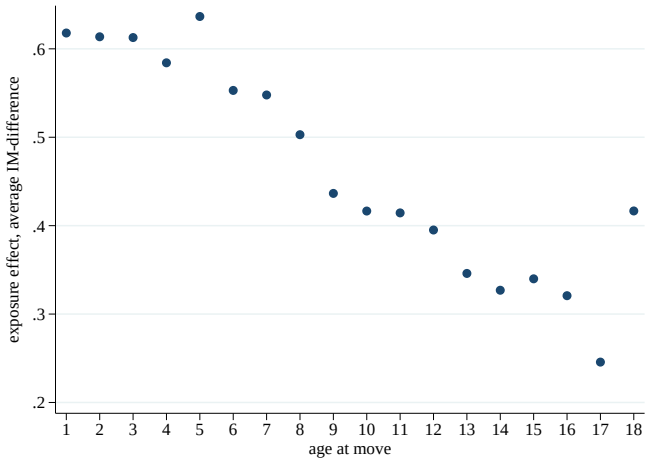
Figure 5.1 plots estimated exposure effects $\hat{\beta}_m$ against the age at move.

Figure 5.1: Average exposure effects

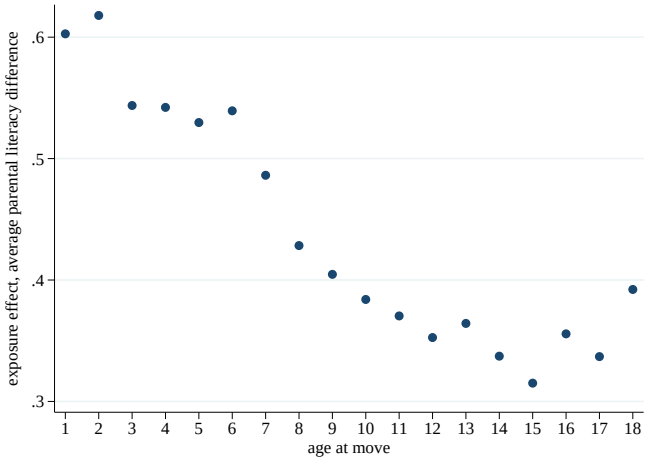
(a) exposure = Δ_{od} parental literacy, child ages 14-18



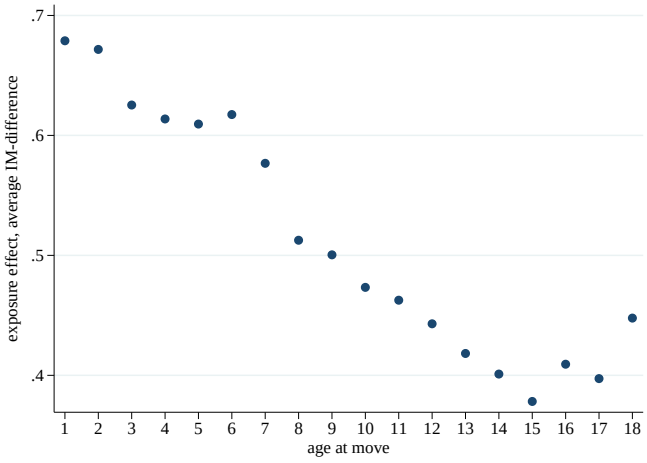
(b) exposure = Δ_{od} average IM, child ages 14-18



(c) exposure = Δ_{od} parental literacy, child ages 14-100



(d) exposure = Δ_{od} average IM, child ages 14-100



The figures indicate that exposure effects decline from age 5-6 to age 12-13, exactly the age-span relevant for primary education. Before age 6 and after age 13, coefficient estimates look constant on average, though there is some variation. This is what we would expect since if the environment matters for primary attainment, it should matter during the time individuals attend primary school – not much before or after. Hence, an individual moving at age 2 from a worse to a better environment should see roughly the same beneficial effect on her outcome as an individual who arrived in the better environment only at age 6, when primary school begins. Similarly, there should be no further systematic effect on primary attainment for moves that occur after the age at which most individuals have completed primary education. Following Chetty and Hendren (2018a), we interpret the average exposure effect after age 13 as an estimate of the selection effect.

6 Correlates of Intergenerational Mobility

6.1 Specification

We now examine the correlates of regional IM. We do not aim to identify causal effects, but simply uncover its main correlates. We run simple univariate specifications linking the proxies of intergenerational mobility ($IM_{r,c}$) with geographical, historical, and at independence population characteristics among non-migrants, accounting for country fixed effects (θ_c). The specification is:

$$IM_{r,c} = \theta_c + G_{r,c}\Phi + H_{r,c}\Gamma + Z_{r,c}\Psi \ [+\lambda L^o_{r,c}] \ + \zeta_{r,c}. \tag{6.1}$$

$G_{r,c}$ are geographic features of region r in country c; $H_{r,c}$ denotes historical, colonial and pre-colonial, characteristics, and $Z_{r,c}$ are variables computed as averages from census data for individuals born before 1960. Since the estimates do not have a causal interpretation, we estimate univariate specifications, adding each explanatory variable one by one (as Chetty *et al.* 2014). And, as we have shown that the literacy of the old generation is a strong correlate of IM, we also report specifications controlling for it, $L^o_{r,c}$. The appendix provides definitions and sources for all variables used in the regional analysis and gives summary statistics. We have also estimated specifications replacing the country constants with province (admin-1 units) fixed-effects so as to better account for unobservable features. Since the regressions do not aim at identifying causal effects, we report the admin-1 unit fixed-effect specifications in the appendix and briefly comment on them below.

Table 6.1 reports the country fixed effects estimates. Panel A looks at the role of geographic features. Panel B examines the association between intergenerational mobility in education and historical, colonial and pre-colonial, features, while Panel C looks at contemporary correlates of mobility. In columns (1)-(3) we examine the correlates of the IM estimate that looks at children aged $14-18$ and applying the "student" correction for kids that are about to complete primary schooling. Columns (4)-(6) investigate the correlates of district-level mobility looking at the sample of migrant households with more than two children born in different districts. In particular, we re-estimate equation (5.1) with a birth-region fixed effect instead of a birth-region specific measure of educational opportunities (which are invariant across individuals born in the same birth-region) and extract the estimated birth-region fixed effects. This allows looking at the correlates of social mobility netting out migration. The table reports three specifications. In columns (1) and (4) we examine the role of the various geographic variables in explaining variation of illiteracy among the "old". Columns (2) and (5) associates the proxies of IM with various geographic/locational/ecological features simply conditioning on country fixed effects. Columns (3) and (6) repeats estimation on the correlates of IM conditioning also on the share of the old generation without completed primary education ("illiterate old") that correlates strongly with IM.

6.2 Geography

Panel A explores the geographic correlates of IM.

Natural Resources A large literature on the "natural resource curse" has linked conflict and other aspects of underdevelopment to the presence of oil, diamonds, and other precious metals. [See, among others, Ross (2004), Berman *et al.* (2017), Guidolin and La Ferrara (2007).] In recent work Hohmann (2018) shows that across African regions natural resource shocks are associated with higher education and structural transformation. We thus associated IM with dummy variables indicating whether there are diamond mine or oil fields in the district. Rows (1) and (2) give the results. There is no significant association between natural resources and IM. Natural resources are somewhat related to literacy (as in Hohmann (2018), but the association with mobility is weak. We also examined whether IM is related to proximity to other mineral sites (like silver or platinum mines), without detecting any significant correlation.

Distance to the Capital Much evidence documents the limited capacity of modern African states to exercise control far from the capitals. [See, among others, Herbst (2000) and Michalopoulos and Papaioannou (2014)]. Development, and as columns (1) and (4) show literacy of the "old", are higher in proximate to capital cities regions. The unconditional correlation in columns (2) and (5) suggests a significant association between proximity to capitals and social mobility. The coefficient on distance to the capita, however, drops once considerably we condition on literacy of the "old" generation (in (3) and (6)). The estimates do not pass standard significance level thresholds both when we use the simple proxy of IM (in (3)) and when we use the one from the multi-children migrant household (in (6)).

Distance to the Border Border areas in many parts of Africa appear unruly and a sizeable part of conflict takes place in areas close to the national border, which often split ethnic groups between two or more modern states.[15] The association between distance to the border and literacy of the "old" is weak as the estimate does not pass significance levels (in (1) and (4)). Likewise social mobility does not correlate with distance to the border (in (2) and (5)). And even though the estimate on distance to the border is significant with the baseline IM index (in (3)), it is statistically indistinguishable from zero in (6) where we use the household fixed-effect regional proxy of IM. Overall there is little evidence that social mobility is linked with proximity to the border.

Distance to the Coast African development is concentrated in the coastline (Henderson et al. (2018)). The distance to the coast is linked to the presence of Europeans during colonization that mostly settled in coastal areas and towns. As the level specification in columns (1) and (4) show distance to the coast is correlated with illiteracy, reflecting, among other things, the relatively higher levels of development in coastal areas (Henderson *et al.* (2018)). IM is also lower in areas proximate to the coast (columns (2) and (5)). The coefficient on distance to coast retains significance even when we condition on the stock of literacy of the "old" in (3) and (6). This implies that social mobility appears higher in coastal areas, even when one nets the higher initial levels of human capital in close-to-the-coast areas.

[15] See Alesina, et al. (2011) and Michalopoulos and Papaioannou (2016) for evidence linking border artificiality and ethnic partitioning to underdevelopment and conflict.

Malaria Malaria has been invariably linked to Africa's underdevelopment (see, among others, Gallup and Sachs (2001), Sachs (2003), Cervellati and Sunde (2015), Weil (2017), Cervellati et al. (2016)). We thus associated the proxies of intergenerational mobility (and the literacy of the "old") with an index reflecting the ecology for malaria. In line with earlier works, the level specifications in columns (1) and (4) show that illiteracy is higher in places with ecological conditions favourable to malaria. As columns (2) and (5) show, educational IM is significantly lower in places favourable to the spread of malaria. However, when we condition on literacy of the old generation in the region, the coefficient on the malaria suitability index falls in absolute value and becomes statistically indistinguishable from zero.

Other Geographic Features [Land Suitability, Elevation, and Ruggedness] We also linked IM to various measures reflecting geographic/land endowments, namely soil quality (land suitability) for agriculture, elevation, and terrain ruggedness, as works have linked these features to regional development. The link between land suitability and social mobility is not particularly strong. Likewise, the association between elevation with social mobility (and literacy) is weak and does not pass standard significance levels.

In contrast, there is a positive and significant association between terrain ruggedness and social mobility; this applies both in the unconditional estimates (in (2) and (5)) and also when we condition on the literacy rates of the old generation (in (3) and (6)) that is also systematically linked to ruggedness. These findings add to Nunn and Puga (2012), who argue that regions with rugged terrain were shielded from Africa's slave trades that have been detrimental to long-run African development (Nunn (2008)).

6.3 Historical Traits

Table 6.1 Panel *B* reports specifications associating the two IM proxies with various historical variables. In each row we report three and three specifications that explore the role of the specific historical trait on IM and on the illiteracy rate of the " old cohort" (in columns (1) and (4)).

Development at independence We start our analysis linking IM to the (log of) population density in 1950, which for most countries in our sample corresponds to the period just before independence. As Africans at the time were almost exclusively employed in agriculture, population density serves as a good proxy of local development. There is a significantly positive association between log population density at independence and literacy rates of the old (columns (1) and (4)). Likewise, population density correlates strongly with social mobility, both when we use the full sample of 14-18 children in the region (in (2)) and when we look at multi-children moving families (in (5)). The within-country correlation between IM and log population density retains statistical significance, once we control for the share of literacy among the old generation (columns (3) and (6)). These patterns show strong inertia in development; IM is positively and strongly related to the level of development in the end of colonization, even when we control for the higher literacy rates in more densely populated places.

Colonial Infrastructure Investments Colonial investments in railroads and roads seem to have played a crucial role in shaping African countries' post-independence development path (See Kerby, Jedwab

and Moradi (2017), and Jedwab and Moradi (2016)). We thus examined the correlation between IM and distance to colonial railroads and roads.[16] In line with earlier works linking roads-railroads to colonial-era development, we find a strong positive association between proximity to railroads and literacy among the "old" (columns (1) and (4)). Log distance to colonial railroads is significantly related also to the share of literacy for children whose parents have not completed primary education (column (2) and (5)).[17] The correlation between proximity to colonial railroads and IM retains significance when we control for the literacy share of the "old" generation. Moreover, the correlation is robust to replacing the country constants with province fixed effects (reported in the Appendix). These results suggest that colonial railroads not only had an impact on development at independence (as reflected on the education level of the " old"), but also on the intergenerational transmission of education beyond any initial effect.

There is also a strong positive association between proximity to colonial roads and IM, as well as the literacy of the old generation. However, the correlation loses significance once we control for the share of literacy of the "old" (in columns (3) and (6)).

Colonial Missions A considerable body of research has uncovered positive local effects of Christian, especially Protestant, missions. We thus examined the correlation between IM and proximity to colonial missions using digitized data from Nunn (2012) and Cage and Rueda (2016). Overall there are 1,321 (361 Catholic, 933 Protestant, 27 British and Foreign Bible Society) and 723 (Protestant only) missions in these data sets. We find a strong within-country positive association between proximity to Catholic and Protestant Christian missions and literacy rates of the " old" . Likewise there is a significantly positive correlation between proximity to Catholic and Protestant missions with social mobility, both when we use the simple proxy of absolute mobility (in column (2)) and when we look at moving households (in column (5)). When we condition on the literacy rates of the "old" generation, the coefficient on log distance to Catholic missions loses significance, while the estimate on Protestant missions retains statistical significance. While data on Christian missions are incomplete and there may be systematic biases (Jedwab et al. (2018)) the analysis shows that early colonial schooling and health investments in Christian missions had lasting effects, both by shaping literacy which in turn is linked to educational mobility and also by shaping subsequent social mobility beyond its influence on the initial levels of literacy.

Pre-colonial Political Centralization and Early Statehood As recent works have traced Africa's post-independence development path to pre-colonial features, we also explored the correlation between IM and pre-colonial political centralization that recent works have linked to contemporary development (see among others Michalopoulos and Papaioannou (2013, 2015), Gennaioli and Rainer (2006, 2007), Alsan (2016), and Depetris-Chauvin (2017).). We correlate IM (and the share of literacy among the old cohort) with the distance to the centroid of the nearest large kingdom or empire using data from Brecke (1999), as geo-coded by Besley and Reynal-Querol (2015) and log distance to pre-colonial states

[16] Data on colonial roads come from Jedwab, Moradi, and Kerby (2016) and cover all Sub-Saharan African countries, but South Africa. So, in these specifications we drop South African regions.

[17] The estimate implies that in districts very close to colonial railroads (log distance of 0) the percentage of literacy for individuals born to illiterate parents is around 18.5 points higher compared to districts being 150 km far from colonial roads ($ln(150) \times 0.037 \approx 0.185$).

using Murdock's data (1959, 1967) though data is missing for some parts of the continent. There is no systematic link between distance to pre-colonial states and IM with the level of literacy (or mean years of schooling) for the old (in (1) and (4)), a result in line with Michalopoulos and Papaioannou (2013) who also failed to detect a robust positive relationship between pre-colonial statehood and education. Distance to pre-colonial states-kingdoms is negatively related to social mobility, though the estimate is positive only with the Murdock (1959, 1967) data. Interestingly, the coefficient on log distance to politically centralized ethnicities during pre-colonial times retains significance when we condition on the share of literacy of the old generation; this hints that these regions may have gained from educational policies post-independence.

6.4 Contemporary Correlates

In Panel *C* of table 6.1 we explore the contemporary correlates of educational mobility.

Industrial Specialization We then examines the role of industrial specialization, associating the proxies of IM and the stock of literacy with the share of employment in agriculture, manufacturing, and services. Literacy is significantly higher in regions with a relatively higher (lower) employment share in manufacturing and services (agriculture). At the same time, educational IM is also higher in regions specializing in the "modern" sectors (manufacturing and services) as compared to the "traditional" sector (agriculture). These patterns apply both with the simple absolute mobility measure estimated among 14-18 aged children (in (1)-(3)) and when we look at children of migrant households (in (4)-(6)). However, when we condition on the literacy rate of the "old" generation, the coefficients on the share of agriculture, manufacturing, and services drop in absolute value and become statistically indistinguishable from zero.

The patterns are similar when we look at the urban-rural regional share. The share of urban households in the region correlates positively with literacy and IM, but the correlation becomes zero once we condition on the literacy share of the "old".

Ethnic Composition Following Easterly and Levine (1997), a large literature has traced Africa's underdevelopment to ethnic fractionalization (see Alesina and La Ferrara (2005) for an overview). Ethnic polarization is strongly related to conflict (Montalvo and Reynal-Querol (2005), Esteban, Mayoral, and Ray (2012)), while ethnic-linguistic fragmentation is inversely related to public goods provision across countries (Desmet et al. (2012)). However, when we exploit within-country variation, there seems to be a significantly positive correlation between ethnic fragmentation or polarization and literacy (in (1) and (4)), a result that appears in line with recent within-country works (Montalvo and Reynal-Querol (2017) and Desmet et al. (2018)). There is also a significantly positive correlation between ethnic fragmentation and IM, reflecting the higher levels of social mobility in capitals where diversity is higher. Once we condition on the share of illiteracy of the "old" generation the coefficient on ethnic fragmentation and ethnic polarization turns negative though it is not significant.

6.5 Summary

The correlation analysis shows that geographic and historical, mostly colonial era, features are related to social mobility. In particular, proximity to the coast and the capital cities is related not only to higher levels of literacy among "old" generation Africans, but also to much higher upward mobility. The likelihood that children of parents without any education will manage to complete at least primary schooling is systematically higher in proximate to the capital and the coast regions, even when one nets their "indirect" role in shaping literacy of the old, which is the strongest correlate of upward mobility. Educational mobility is not much related to natural resources or other geographic traits, with the exception of terrain ruggedness that is positively correlated to social mobility.

The univariate regression analysis further showed that at-independence development, related to population density during late colonization and proximity to colonial railroads and Christian missions that provided education and basic health are also linked to higher levels of social mobility even when we condition on literacy rates.

In contrast, while industrial specialization and "structural transformation" features are strongly correlated with the education "stock", the literacy of the old generation, and unconditionally with inter-generational mobility, the share of employment in agriculture, services, and manufacturing is unrelated to mobility once we condition on the literacy rates of the "old" generation.

While these correlations do not have a causal interpretation, they suggests that early development, linked to geography-ecology and colonial investments have shaped contemporary African development via two main channels. First, given strong inertia in education small differences in educational attainment at independence had sizeable effects, by shaping educational opportunity. Second, some geographic and historical aspects tend to exert an influence on social mobility, even conditioning on literacy, suggesting that early differences get magnified over time.

Table 6.1: District and birth-region-level correlates of the share of literate old and IM, ages 14-18, country fixed effects, with student correction

	observational estimates				within-migrant-household estimates			
	(1)	(2)	(3)		(4)	(5)	(6)	
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
Panel A: geography								
DCAP	-0.314***	-0.184***	-0.027	2437	-0.497***	-0.286***	-0.004	1194
	(0.036)	(0.022)	(0.019)		(0.094)	(0.061)	(0.025)	
DBORD	0.028	-0.012	-0.026***	2437	0.066	0.011	-0.027	1194
	(0.029)	(0.016)	(0.009)		(0.096)	(0.057)	(0.020)	
DCOAST	-0.171***	-0.123***	-0.037**	2437	-0.371***	-0.288***	-0.086***	1194
	(0.045)	(0.025)	(0.017)		(0.128)	(0.085)	(0.024)	
MALTR	-0.300***	-0.172***	-0.021	2436	-0.391***	-0.243***	-0.023	1193
	(0.069)	(0.047)	(0.019)		(0.101)	(0.066)	(0.035)	
AGSUIT	0.012	0.062*	0.056***	2407	-0.039	-0.013	0.010	1183
	(0.057)	(0.035)	(0.018)		(0.151)	(0.087)	(0.030)	
ELEV	0.035	-0.008	-0.025	2428	0.064	0.043	0.007	1193
	(0.060)	(0.032)	(0.018)		(0.111)	(0.069)	(0.028)	
TERRUG	0.125**	0.096***	0.033*	2435	0.230**	0.202***	0.075***	1193
	(0.053)	(0.031)	(0.018)		(0.101)	(0.055)	(0.028)	
OILDUM	0.003	0.010	0.008	2437	0.105*	0.066*	0.006	1194
	(0.023)	(0.015)	(0.009)		(0.063)	(0.035)	(0.013)	
DIADUM	-0.015	-0.017**	-0.009	2437	-0.019	-0.013	-0.002	1194
	(0.015)	(0.008)	(0.008)		(0.017)	(0.011)	(0.008)	
Panel B: history								
POPD	0.266***	0.178***	0.046***	2435	0.387***	0.286***	0.080***	1193
	(0.063)	(0.038)	(0.015)		(0.072)	(0.042)	(0.021)	
DRRD	-0.348***	-0.209***	-0.030*	1932	-0.542***	-0.337***	-0.049**	1156
	(0.054)	(0.031)	(0.016)		(0.087)	(0.056)	(0.019)	
DROAD	-0.318***	-0.171***	-0.015	2148	-0.364***	-0.221***	-0.019	1165
	(0.038)	(0.022)	(0.011)		(0.061)	(0.038)	(0.020)	
DCMISS	-0.404***	-0.216***	-0.010	2437	-0.512***	-0.321***	-0.037	1194
	(0.070)	(0.046)	(0.023)		(0.113)	(0.076)	(0.027)	
DPMISS	-0.395***	-0.240***	-0.045***	2437	-0.596***	-0.387***	-0.067***	1194
	(0.052)	(0.031)	(0.017)		(0.100)	(0.062)	(0.022)	
DPCEM	0.019	-0.019	-0.029	2437	-0.052	-0.047	-0.017	1194
	(0.053)	(0.027)	(0.022)		(0.072)	(0.047)	(0.023)	
DPCST	-0.019	-0.045**	-0.036**	2437	-0.134	-0.151***	-0.077***	1194
	(0.045)	(0.023)	(0.015)		(0.088)	(0.050)	(0.024)	
Panel C: contemporary								
URBSHR	0.423***	0.168***	-0.069***	2386	0.386***	0.231***	0.014	1170
	(0.024)	(0.016)	(0.014)		(0.036)	(0.027)	(0.031)	
AGSHR	-0.686***	-0.341***	-0.016	2264	-0.544***	-0.323***	-0.052	1006
	(0.028)	(0.015)	(0.036)		(0.064)	(0.042)	(0.048)	
MANSHR	0.267***	0.109***	-0.023	2264	0.344***	0.174***	-0.009	1006
	(0.045)	(0.027)	(0.016)		(0.079)	(0.049)	(0.024)	
SERSHR	0.635***	0.329***	0.039	2264	0.518***	0.315***	0.057	1006
	(0.032)	(0.016)	(0.032)		(0.064)	(0.040)	(0.051)	
ETHFRAG	0.210***	0.057**	-0.045***	1504	0.135***	0.049**	-0.014	567
	(0.042)	(0.024)	(0.016)		(0.041)	(0.025)	(0.014)	
ETHPOL	0.090***	0.008	-0.035**	1504	0.069**	0.017	-0.015	567
	(0.031)	(0.019)	(0.016)		(0.029)	(0.021)	(0.013)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include country fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. ■ ■ lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

7 Conclusion

We have conducted the first systematic exploration of intergenerational mobility in education across African countries and regions using census data covering 20 million Africans, who cohabitate with their parent(s).

In the first part of the analysis, we construct and describe absolute educational mobility across 23 African countries and 2,444 regions. Our analysis uncovers large differences across and within countries in the likelihood that children of illiterate parents complete primary schooling (or more). While some parts of Africa, mostly in the South, reflect a "hopeful continent", there are also places, mostly in the Sahel and the Sahara, that hope is still a mirage. Escaping a family legacy of illiteracy varies sizeably also within countries. For example, educational mobility is quite high in the coastal parts of Ghana (comparable to South Africa and Botswana), but abysmally low in the Northern regions that look more similar to Burkina Faso and Sudan. However, half of the variation in intergenerational mobility is explained by a single variable: the literacy of the "old" generation in the country and region. This pattern applies both to boys and girls and it is especially pronounced for rural households. This finding suggests huge inertia that can explain the sizeable persistence in historical development that a growing strand of research documents.

In the second part, we exploit variation from moving households, as in this case children of the same family were subject to different regional educational exposure. The analysis uncovers that while selection is sizeable, there are considerable early childhood exposure effects. These results suggest a causal link between literacy of the old and intergenerational mobility.

In the third part of the paper, we have conducted a preliminary exploration of the correlates of intergenerational mobility in education, assessing the role of geographic-ecological features, historical factors, and contemporary features. We find that mobility is positively correlated with colonial infrastructure investments and some geographic characteristics, especially distance form the coast (negatively) and ruggedness (positively).

Our exploration of social mobility calls for future research. Motivated by the salience of ethnic and religious cleavages across the African continent, one avenue we are currently pursuing is the construction of ethnic and religion-specific measures of social mobility within African countries and exploring their correlates. Another avenue is examining the role of post-independence educational policies and schooling investments on social mobility.

Social Mobility in Africa
Supplementary Online Appendix*

Alberto Alesina
Harvard University, CEPR and NBER

Sebastian Hohmann
London Business School

Stelios Michalopoulos
Brown University, CEPR and NBER

Elias Papaioannou
London Business School and CEPR

This draft: July 2018

*Alberto Alesina Harvard Univerity and IGIER Bocconi, Sebatian Hohmnn , London Busienss Schoiol, Stelios Michalopoulos. Brown University, Elias Papaioannou. London Business School.

A Correlation between schooling and household wealth with DHS and Afrobarometer

A.1 DHS

A.1.1 Household wealth

Table A.1: Household wealth quintile and years of schooling

	(1) wealth quintile	(2) wealth quintile	(3) wealth quintile	(4) wealth quintile	(5) wealth quintile	(6) wealth quintile
years of schooling	0.123***	0.0815***	0.0994***	0.0857***	0.0857***	0.0791***
	(39.02)	(19.38)	(31.43)	(33.74)	(34.31)	(33.42)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.175	0.402	0.459	0.520	0.525	0.557
marginal R-squared	0.175	0.06	0.073	0.05	0.052	0.042
within R-squared		0.399	0.441	0.325	0.339	0.274
N	3516848	3509051	3509051	3509051	2823745	2823745

This table shows regression results of household wealth on years of schooling for individuals aged 18+. The dependent variable in all columns is the DHS household wealth quintile (computed for each survey, i.e. country-year) separately based on the DHS-computed wealth index). Individual controls are age, age squared, dummies for male individuals, male household head, urban residence, the log of the number of household members, and individual birth decade dummies. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GDS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

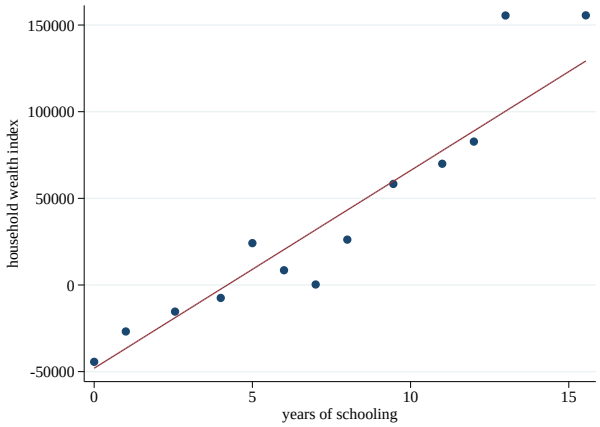
Table A.2: Household wealth index and years of schooling

	(1) wealth index	(2) wealth index	(3) wealth index	(4) wealth index	(5) wealth index	(6) wealth index
years of schooling	11409.6***	7100.9***	8957.8***	8140.5***	8320.5***	7940.2***
	(7.69)	(8.73)	(8.85)	(8.01)	(7.45)	(7.10)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.049	0.123	0.135	0.330	0.250	0.287
marginal R-squared	0.049	0.013	0.017	0.025	0.025	0.026
within R-squared		0.121	0.132	0.101	0.097	0.079
N	3516854	3509057	3509057	3509057	2823751	2823751

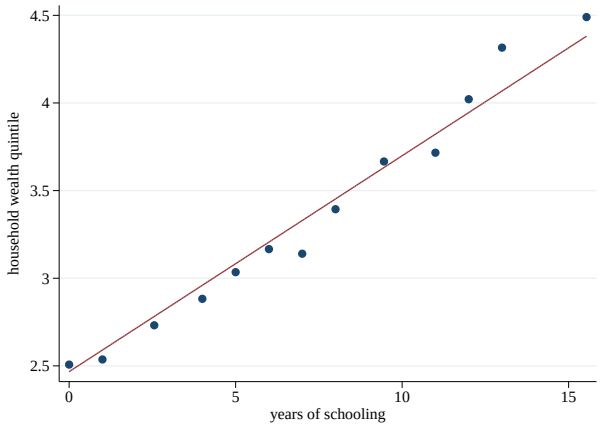
This table shows regression results of household wealth on years of schooling for individuals aged 18+. The dependent variable in all columns is the DHS household wealth index (computed for each survey, i.e. country-year) separately as the principal component of a variety of variables capturing asset ownership, health etc.). Individual controls are age, age squared, dummies for male individuals, male household head, urban residence, the log of the number of household members, and individual birth decade dummies. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GDS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.1: Binned scatter plots

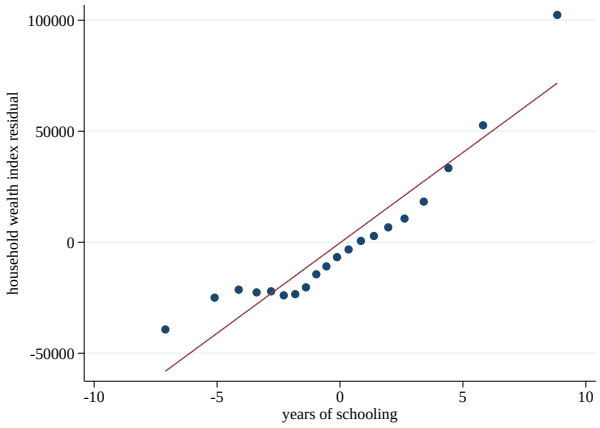
(a) wealth index, unconditional



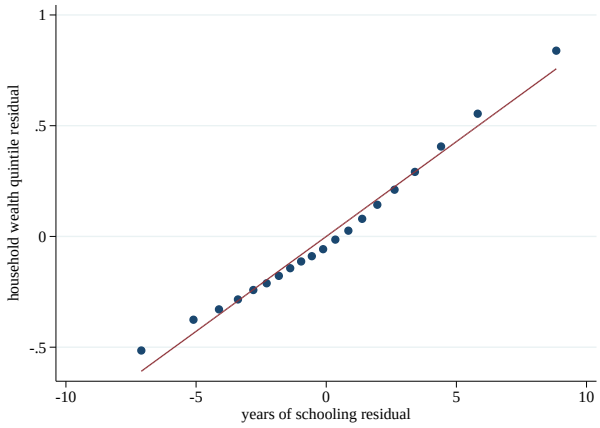
(b) wealth quintile, unconditional



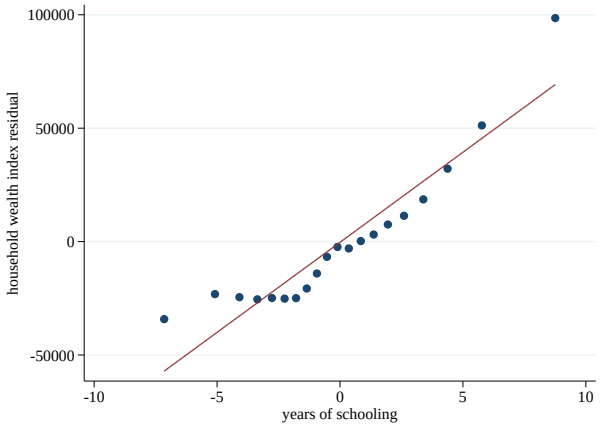
(c) wealth index, conditional on controls and region FE



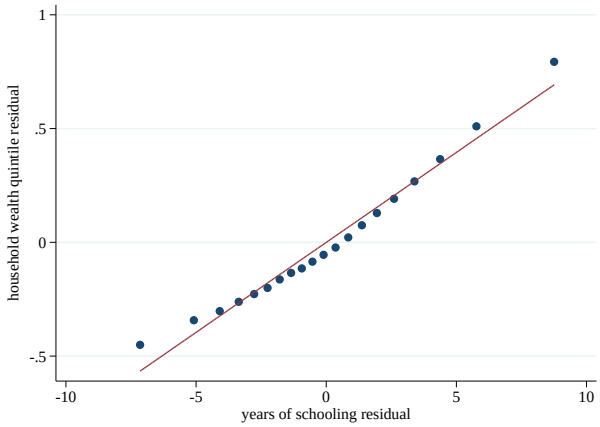
(d) wealth quintile, conditional on controls and region FEs



(e) wealth index, conditional on controls and admin-2 FE



(f) wealth quintile, conditional on controls and admin-2 FE



A.1.2 Child mortality

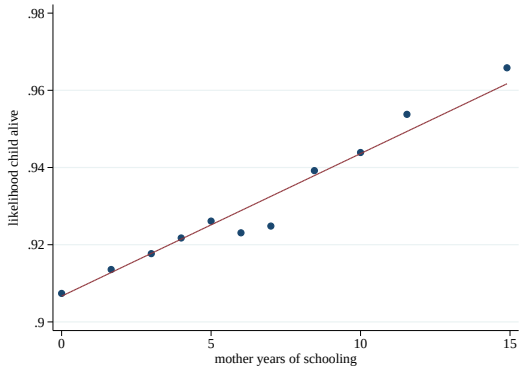
Table A.3: Probability that child survives and years of schooling

	(1) 𝕀(child alive)	(2) 𝕀(child alive)	(3) 𝕀(child alive)	(4) 𝕀(child alive)	(5) 𝕀(child alive)	(6) 𝕀(child alive)
years of schooling	0.00369***	0.00313***	0.00208***	0.00170***	0.00165***	0.00154***
	(12.51)	(12.82)	(8.97)	(12.08)	(10.90)	(10.71)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.003	0.058	0.066	0.068	0.068	0.070
marginal R-squared	0.003	0.002	0.001	0	0	0
within R-squared		0.055	0.052	0.052	0.052	0.051
N	1239858	1172339	1172339	1172339	923261	923260

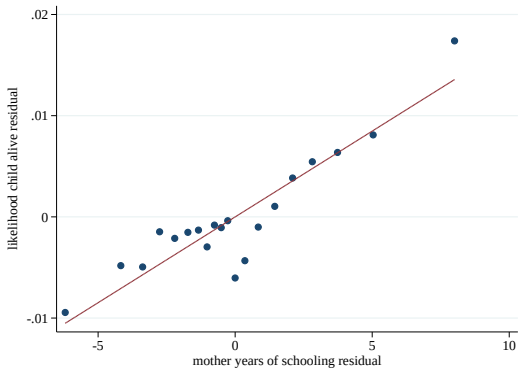
This table shows regression results for child mortality on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to 1 if a child is alive and zero otherwise. Individual controls are mother age, age squared, dummies for children born as twins, child-birth-year dummies, a dummy for the number a child occupies in the birth sequence of the mother, the number of births of the mother, dummies for male household head, urban residence, the log of the number of household members, and individual birth decade dummies. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GDS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.2: Binned scatter plots

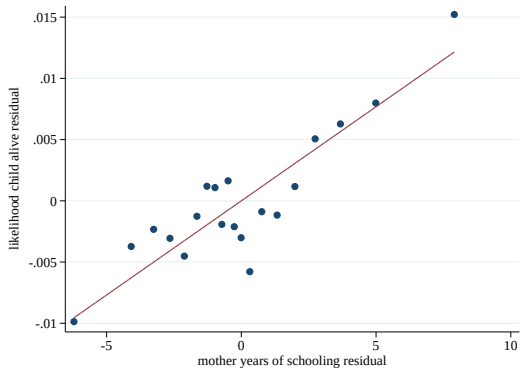
(a) 𝕀(child alive), unconditional



(b) 𝕀(child alive), conditional on controls and region FE



(c) 𝕀(child alive), conditional on controls and admin-2 FE



A.1.3 Bargaining power

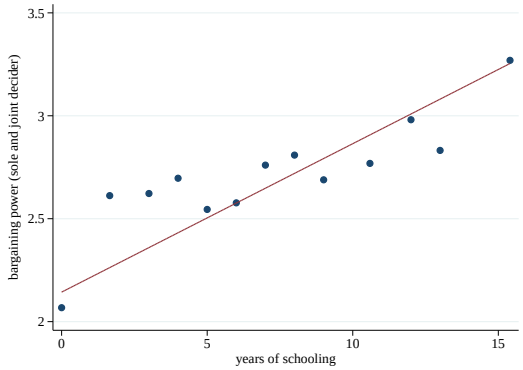
Table A.4: Bargaining power (sole and joint decider) on years of schooling

	(1) bargaining power	(2) bargaining power	(3) bargaining power	(4) bargaining power	(5) bargaining power	(6) bargaining power
years of schooling	0.0721***	0.0698***	0.0442***	0.0296***	0.0300***	0.0275***
	(7.10)	(7.52)	(5.98)	(7.89)	(9.36)	(8.87)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.041	0.126	0.288	0.322	0.326	0.340
marginal R-squared	0.041	0.031	0.01	0.004	0.004	0.003
within R-squared		0.1	0.057	0.043	0.041	0.039
N	615205	614634	614634	614634	534752	534751

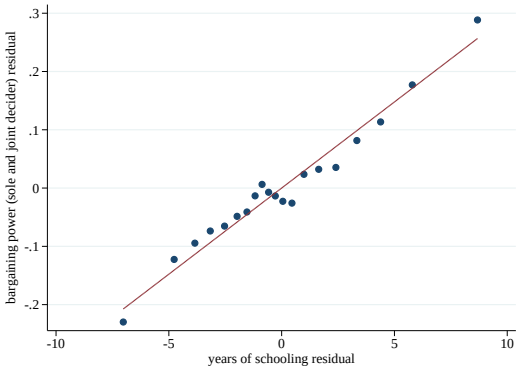
This table shows regression results for individual bargaining power on years of schooling for individuals aged 18+. The dependent variable in all columns is a measure of individual bargaining power. This measure is constructed as the sum of six indicators equal to 1 if an individual takes part (either as sole or joint decision maker) in a particular decision: (a) decisions affecting the individual's health, (b) large household purchases, (c) daily needs household purchases, (d) visits of family relatives, (e) what to cook each day, (f) what is to be done with money earned by the spouse. Individual controls are age, age squared, dummies for male individuals, male household head, and urban residence, as well as the log of the number of household members, and individual birth decade dummies. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.3: Binned scatter plots

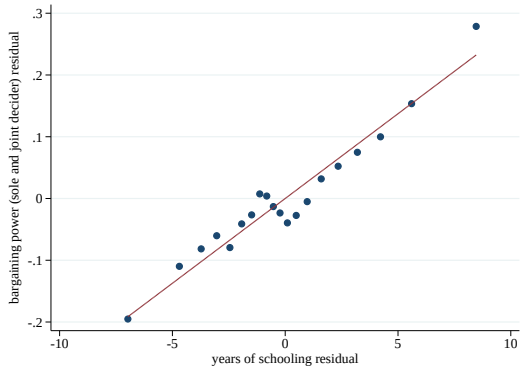
(a) Bargaining power, unconditional



(b) Bargaining power, conditional on controls and region FE



(c) Bargaining power, conditional on controls and admin-2 FE



A.1.4 Attitudes towards domestic violence

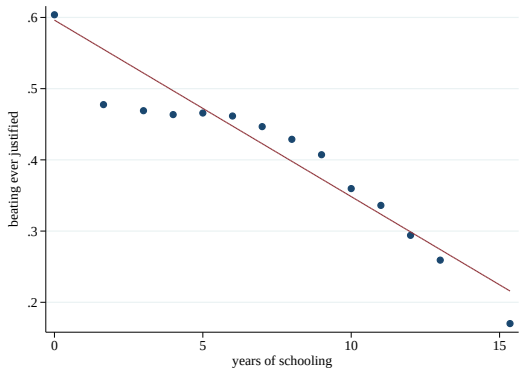
Table A.5: Attitudes towards domestic violence on years of schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	𝕀(beating justified)	𝕀(beating justified)	𝕀(beating justified)	𝕀(beating justified)	𝕀(beating justified)	𝕀(beating justified)
years of schooling	-0.0248***	-0.0196***	-0.0178***	-0.0170***	-0.0172***	-0.0168***
	(-11.01)	(-10.34)	(-14.02)	(-14.59)	(-12.84)	(-12.22)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.057	0.093	0.193	0.228	0.241	0.257
marginal R-squared	.057	.028	.019	.016	.016	.014
within R-squared		.09	.045	.029	.03	.025
N	766631	765884	765884	765884	666739	666739

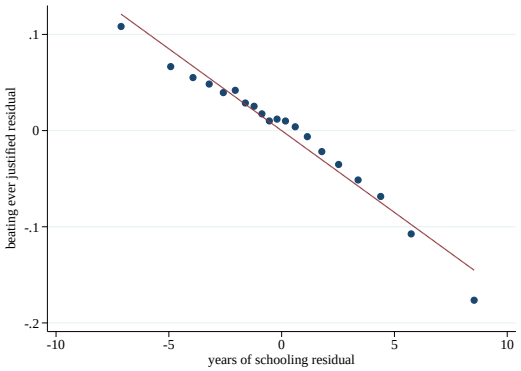
This table shows regression results for attitudes towards domestic violence on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to one if the respondent responds 'yes' to any of the questions of whether beating the wife is justified if she (a) goes out without telling the husband, (b) neglects the children, (c) argues with the husband, (d) refuses to have sex with the husband, (e) burns the food.. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.4: Binned scatter plots

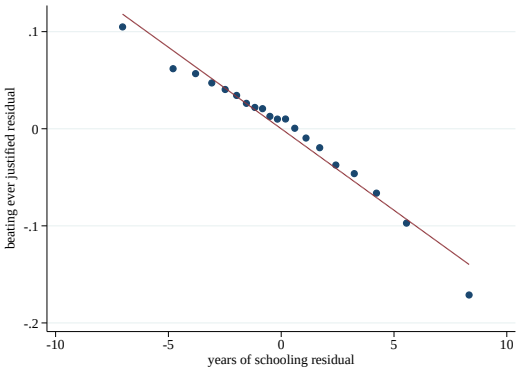
(a) Attitudes towards domestic violence, unconditional



(b) Attitudes towards domestic violence, conditional on controls and region FE



(c) Attitudes towards domestic violence, conditional on controls and admin-2 FE



A.1.5 Fertility

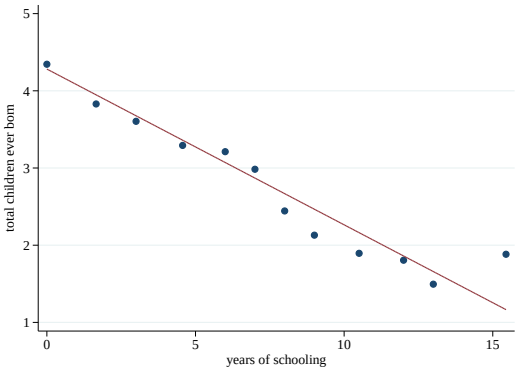
Table A.6: Fertility on years of schooling

	(1) # children	(2) # children	(3) # children	(4) # children	(5) # children	(6) # children
years of schooling	-0.202***	-0.0893***	-0.0970***	-0.0894***	-0.0880***	-0.0852***
	(-41.97)	(-25.56)	(-31.56)	(-30.05)	(-26.84)	(-26.14)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.096	0.578	0.597	0.603	0.603	0.606
marginal R-squared	.096	.015	.015	.011	.012	.01
within R-squared		.386	.264	.237	.24	.231
N	1923074	1856989	1856989	1856989	1491708	1491708

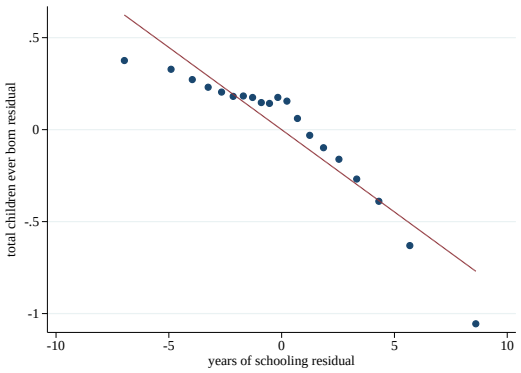
This table shows regression results for total number of children ever born on years of schooling for individuals aged 18+. The dependent variable in all columns is the total number of children ever born. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.5: Binned scatter plots

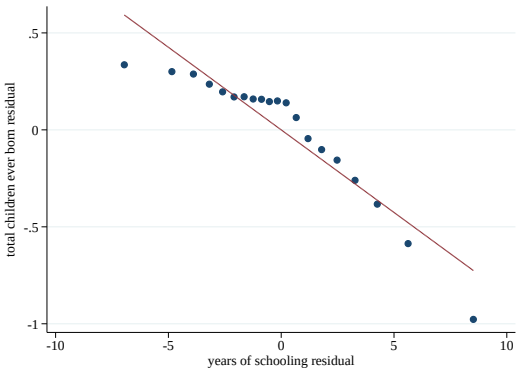
(a) Fertility, unconditional



(b) Fertility, conditional on controls and region FE



(c) Fertility, conditional on controls and admin-2 FE



A.1.6 Desired number of children

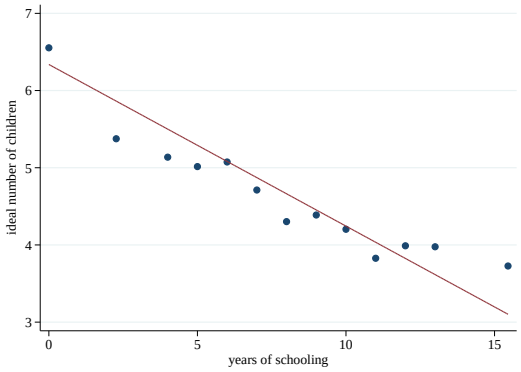
Table A.7: Desired number of children on years of schooling

	(1) desired # children	(2) desired # children	(3) desired # children	(4) desired # children	(5) desired # children	(6) desired # children
years of schooling	-0.209***	-0.183***	-0.142***	-0.109***	-0.101***	-0.0926***
	(-17.46)	(-14.49)	(-11.38)	(-16.03)	(-14.07)	(-14.27)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.083	0.162	0.291	0.341	0.328	0.342
marginal R-squared	.083	.051	.025	.014	.012	.01
within R-squared		.138	.097	.064	.062	.054
N	1549614	1495878	1495878	1495878	1192596	1192594

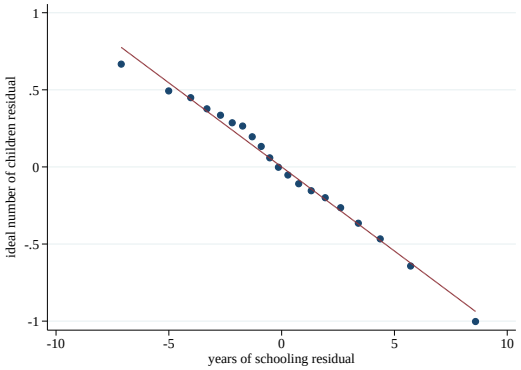
This table shows regression results for desired number of children on years of schooling for individuals aged 18+. The dependent variable in all columns is the individual's ideal desired number of children. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.6: Binned scatter plots

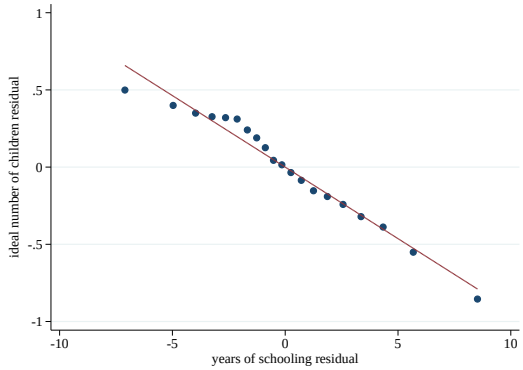
(a) Desired number of children, unconditional



(b) Desired number of children, conditional on controls and region FE



(c) Desired number of children, conditional on controls and admin-2 FE



A.1.7 Age at first marriage

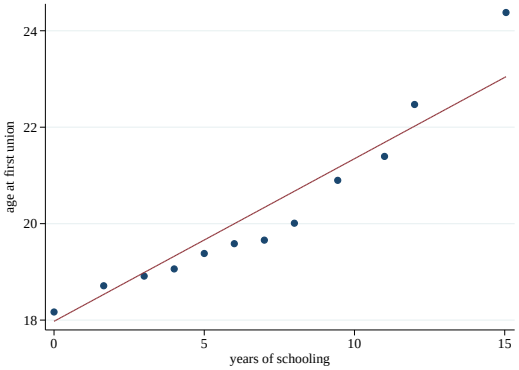
Table A.8: Age of first union on years of schooling

	(1) age first union	(2) age first union	(3) age first union	(4) age first union	(5) age first union	(6) age first union
years of schooling	0.337***	0.242***	0.259***	0.243***	0.243***	0.240***
	(32.35)	(24.96)	(30.91)	(34.71)	(30.13)	(30.25)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.094	0.328	0.357	0.369	0.371	0.375
marginal R-squared	.094	.04	.036	.029	.03	.028
within R-squared		.306	.262	.25	.251	.248
N	1449207	1389458	1389458	1389458	1106824	1106824

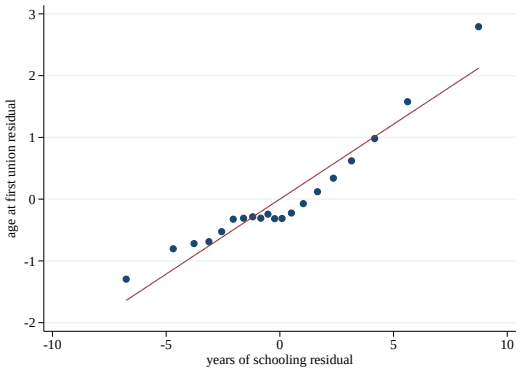
This table shows regression results for age at first union on years of schooling for individuals aged 18+. The dependent variable in all columns is the individual's age at first union / marriage. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.7: Binned scatter plots

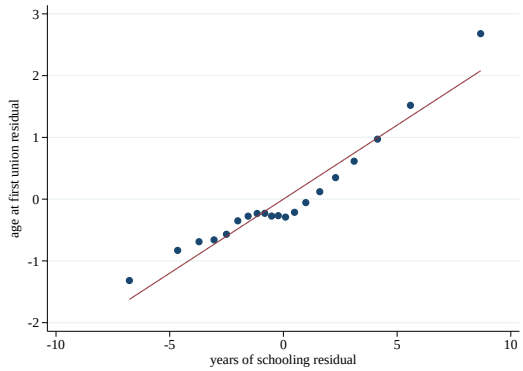
(a) Age at first marriage, unconditional



(b) Age at first marriage, conditional on controls and region FE



(c) Age at first marriage, conditional on controls and admin-2 FE



A.1.8 Age at first sexual intercourse

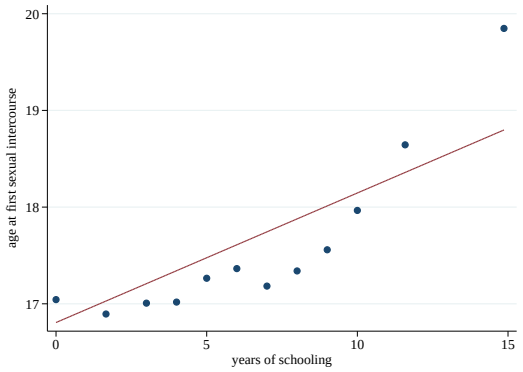
Table A.9: Age of first sexual intercourse on years of schooling

	(1) age first sex	(2) age first sex	(3) age first sex	(4) age first sex	(5) age first sex	(6) age first sex
years of schooling	0.134***	0.113***	0.141***	0.142***	0.144***	0.143***
	(9.95)	(8.11)	(18.13)	(17.71)	(15.05)	(15.18)
individual controls	no	yes	yes	yes	yes	yes
fixed effects	no	no	survey	survey, region	survey, admin-1	survey, admin2
R-squared	0.029	0.115	0.189	0.211	0.209	0.216
marginal R-squared	.029	.016	.02	.019	.019	.018
within R-squared		.101	.08	.075	.077	.074
N	1513798	1483235	1483235	1483235	1171074	1171074

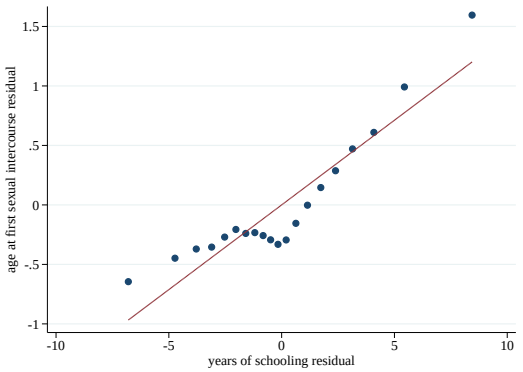
This table shows regression results for age at first sexual intercourse on years of schooling for individuals aged 18+. The dependent variable in all columns is the individual's age at first sexual intercourse. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by DHS) fixed effects. Columns (5) and (6) restrict attention only to the sample for which GPS co-ordinates are available and replaces the DHS region fixed effects with admin-1 (5) and admin-2 (6) region fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.8: Binned scatter plots

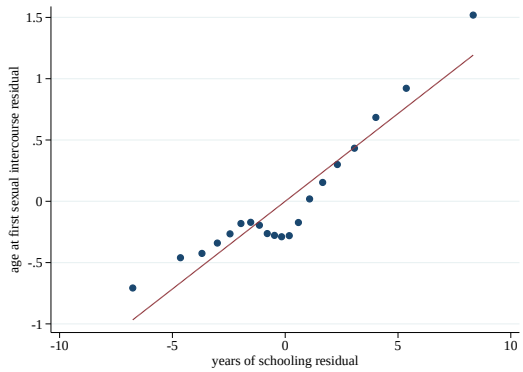
(a) Age at first sexual intercourse, unconditional



(b) Age at first sexual intercourse, conditional on controls and region FE



(c) Age at first sexual intercourse, conditional on controls and admin-2 FE



A.2 Afrobarometer

A.3 Living conditions

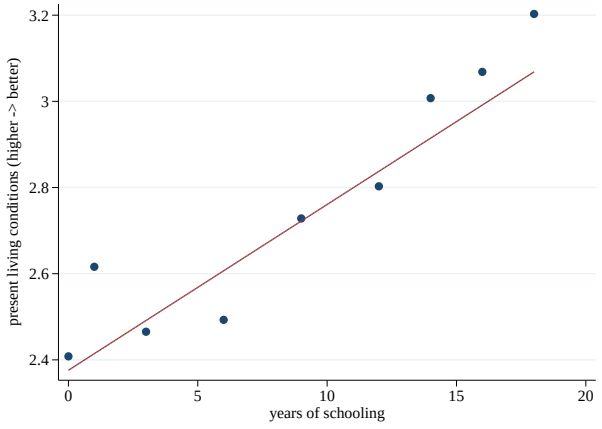
Table A.10: Present living conditions (higher → better) on years of schooling

	(1) living conds.	(2) living conds.	(3) living conds.	(4) living conds.
years of schooling	0.0385***	0.0320***	0.0325***	0.0334***
	(9.35)	(7.25)	(15.43)	(15.91)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.025	0.034	0.117	0.151
marginal R-squared	.025	.014	.012	.012
within R-squared		.024	.019	.019
N	104004	102977	102977	102977

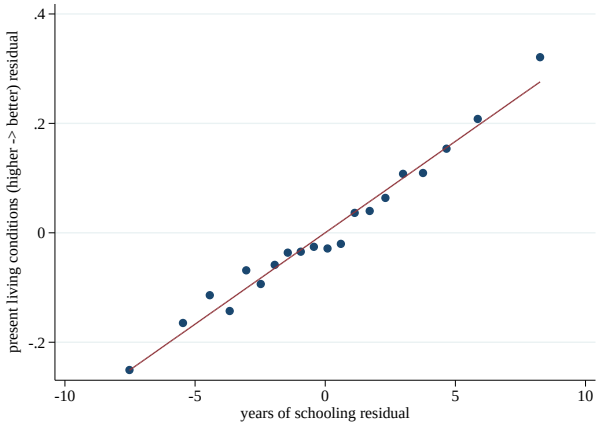
This table shows regression results for living conditions on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's present living conditions (higher → better). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.9: Binned scatter plots

(a) Living conditions, unconditional



(b) Living conditions, conditional on controls and region FE



A.4 Own living conditions vs. those of others

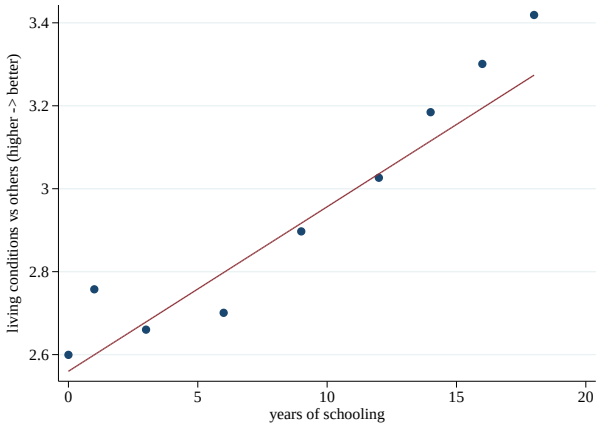
Table A.11: Living conditions vs others (higher → better) on years of schooling

	(1) rel living conds.	(2) rel living conds.	(3) rel living conds.	(4) rel living conds.
years of schooling	0.0397***	0.0337***	0.0368***	0.0366***
	(13.87)	(11.56)	(18.63)	(18.67)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.037	0.045	0.102	0.128
marginal R-squared	.037	.022	.022	.02
within R-squared		.04	.034	.029
N	100826	99854	99854	99854

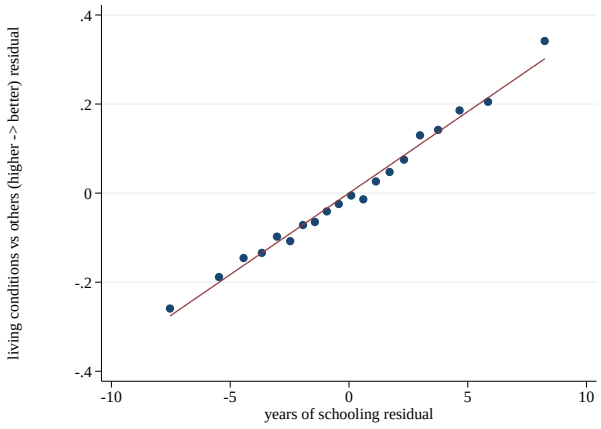
This table shows regression results for relative living conditions on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's living conditions vs how she perceives those of others (higher → better). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.10: Binned scatter plots

(a) Living conditions vs. others, unconditional



(b) Living conditions vs. others, conditional on controls and region FE



A.5 Own living conditions in 12 months

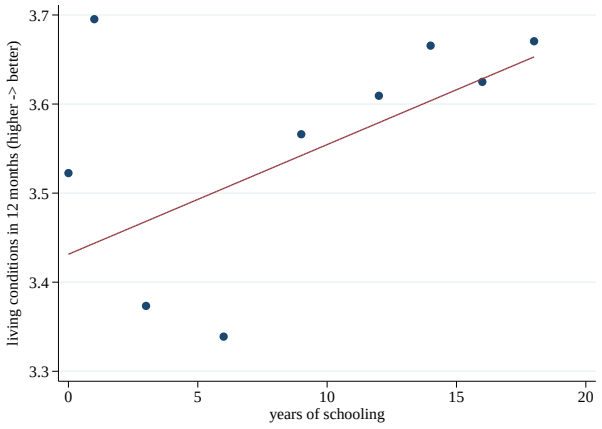
Table A.12: Living conditions in 12 months (higher → better) on years of schooling

	(1) living conds in 1 yr	(2) living conds in 1 yr	(3) living conds in 1 yr	(4) living conds in 1 yr
years of schooling	0.0123**	0.00194	0.0126***	0.0138***
	(2.29)	(0.35)	(5.89)	(7.38)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.003	0.015	0.205	0.236
marginal R-squared	.003	0	.002	.002
within R-squared		.005	.004	.004
N	92145	91398	91398	91398

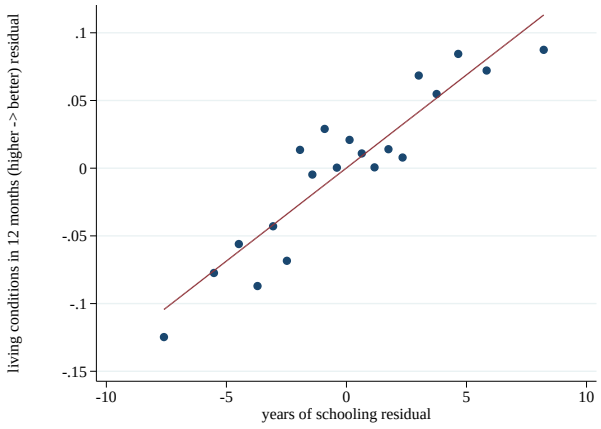
This table shows regression results for living conditions in 12 months on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's expected living conditions in 12 months (higher → better). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.11: Binned scatter plots

(a) Living conditions in 12 months, unconditional



(b) Living conditions in 12 months, conditional on controls and region FE



A.6 How often go without food

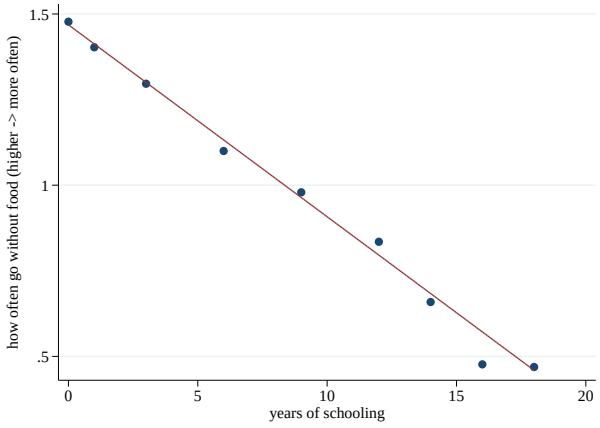
Table A.13: How often go without food (higher → more often) on years of schooling

	(1) freq. no food	(2) freq. no food	(3) freq. no food	(4) freq. no food
years of schooling	-0.0561***	-0.0462***	-0.0476***	-0.0474***
	(-12.96)	(-10.27)	(-15.68)	(-16.94)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.049	0.061	0.149	0.185
marginal R-squared	.049	.027	.024	.023
within R-squared		.057	.045	.037
N	104233	103187	103187	103187

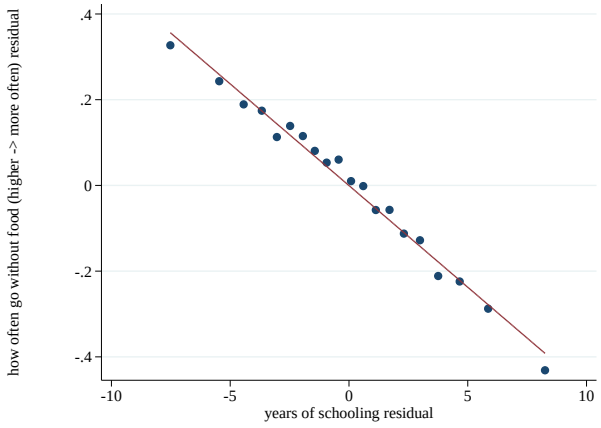
This table shows regression results for frequency of going without food on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without food (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.12: Binned scatter plots

(a) How often go without food, unconditional



(b) How often go without food, conditional on controls and region FE



A.7 How often go without water

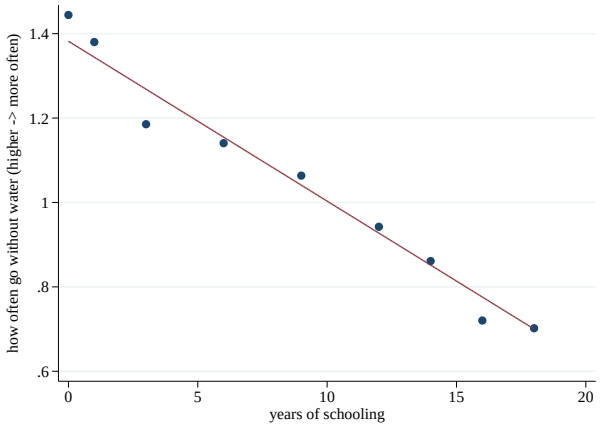
Table A.14: How often go without food (higher → more often) on years of schooling

	(1) freq. no food	(2) freq. no food	(3) freq. no food	(4) freq. no food
years of schooling	-0.0561***	-0.0462***	-0.0476***	-0.0474***
	(-12.96)	(-10.27)	(-15.68)	(-16.94)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.049	0.061	0.149	0.185
marginal R-squared	.049	.027	.024	.023
within R-squared		.057	.045	.037
N	104233	103187	103187	103187

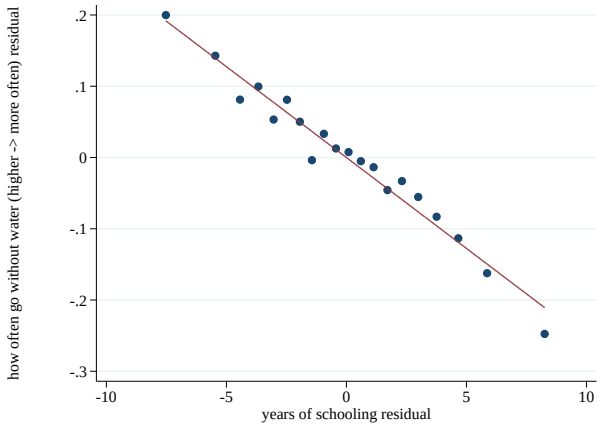
This table shows regression results for frequency of going without food on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without food (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.13: Binned scatter plots

(a) How often go without water, unconditional



(b) How often go without water, conditional on controls and region FE



A.8 Interest in public affairs

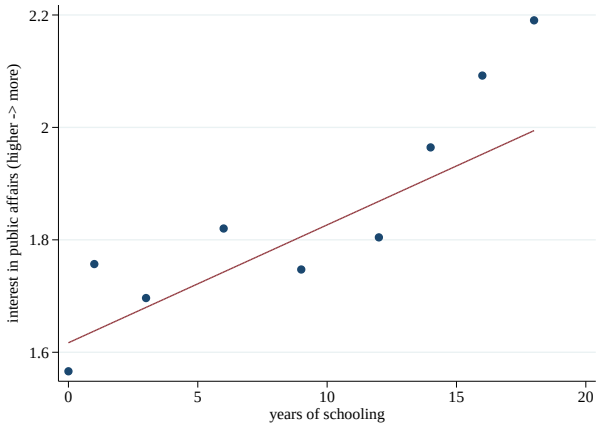
Table A.15: Interest in public affairs (higher → more) on years of schooling

	(1) int. public aff.	(2) int. public aff.	(3) int. public aff.	(4) int. public aff.
years of schooling	0.0210***	0.0247***	0.0329***	0.0340***
	(7.26)	(9.30)	(15.54)	(17.00)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.009	0.038	0.086	0.109
marginal R-squared	.009	.01	.015	.015
within R-squared		.033	.038	.04
N	103355	102364	102364	102364

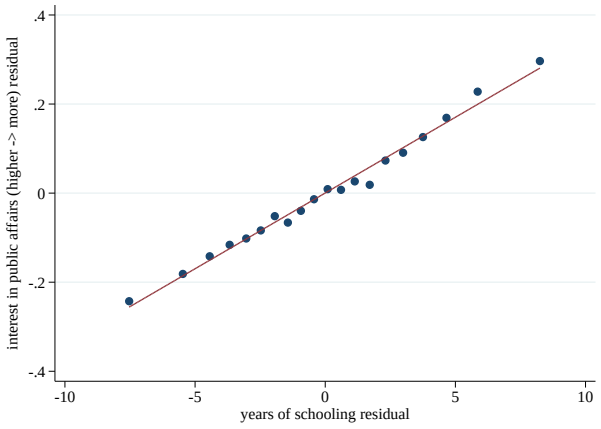
This table shows regression results for interest in public affairs on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's interest in public affairs (higher → more). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.14: Binned scatter plots

(a) Interest in public affairs, unconditional



(b) Interest in public affairs, conditional on controls and region FE



A.9 Frequency of discussing politics

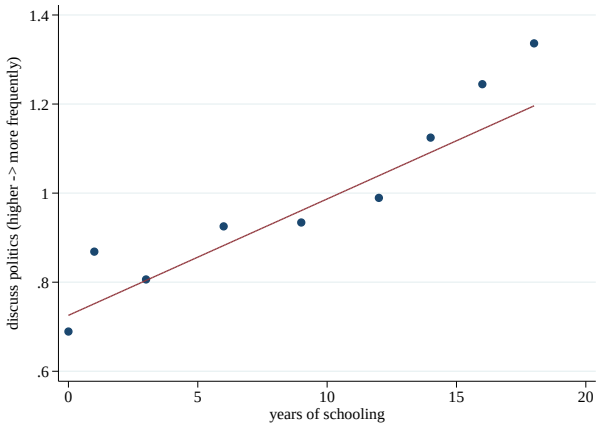
Table A.16: Discuss politics (higher → more frequently) on years of schooling

	(1) discuss pol	(2) discuss pol	(3) discuss pol	(4) discuss pol
years of schooling	0.0261***	0.0247***	0.0290***	0.0294***
	(13.94)	(13.69)	(19.66)	(21.52)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.031	0.060	0.101	0.119
marginal R-squared	.031	.023	.026	.025
within R-squared		.057	.063	.063
N	103467	102461	102461	102461

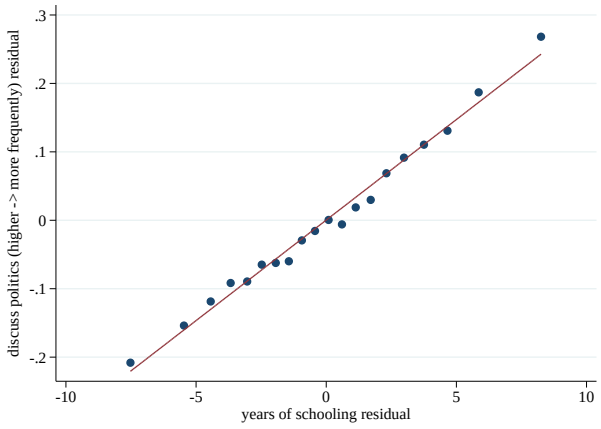
This table shows regression results for frequency of discussing politics on years of schooling for individuals aged 18+. The dependent variable in all columns is the frequency with which the respondent discusses politics (higher → more frequently). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.15: Binned scatter plots

(a) Frequency of discussing politics, unconditional



(b) Frequency of discussing politics, conditional on controls and region FE



A.10 Politics too complicated?

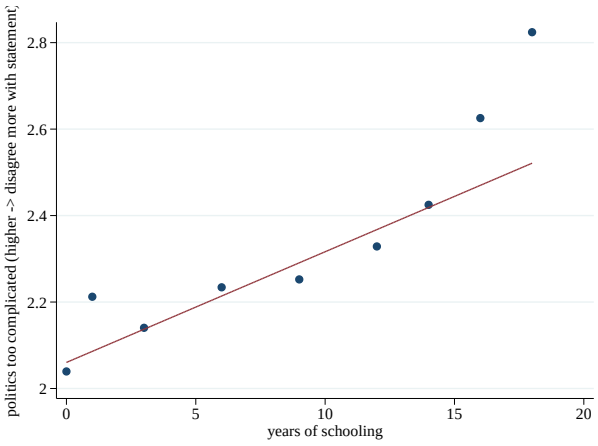
Table A.17: Politics too complicated (higher → disagree more with statement) on years of schooling

	(1) pol too compl	(2) pol too compl	(3) pol too compl	(4) pol too compl
years of schooling	0.0256***	0.0259***	0.0232***	0.0242***
	(6.80)	(6.64)	(7.68)	(9.06)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.010	0.013	0.038	0.069
marginal R-squared	.01	.008	.006	.006
within R-squared		.013	.01	.01
N	72403	71808	71808	71808

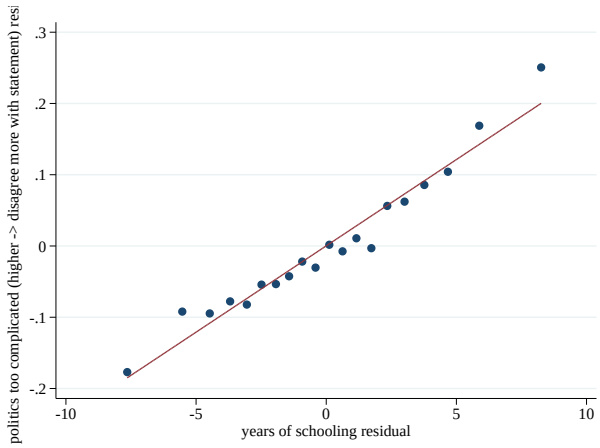
This table shows regression results for of whether find politics too complicated on years of schooling for individuals aged 18+. The dependent variable in all columns is whether the respondent disagrees with the statement that politics too complicated (higher → disagree more with statement). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.16: Binned scatter plots

(a) Politics too complicated?, unconditional



(b) Politics too complicated?, conditional on controls and region FE



A.11 Support for democracy

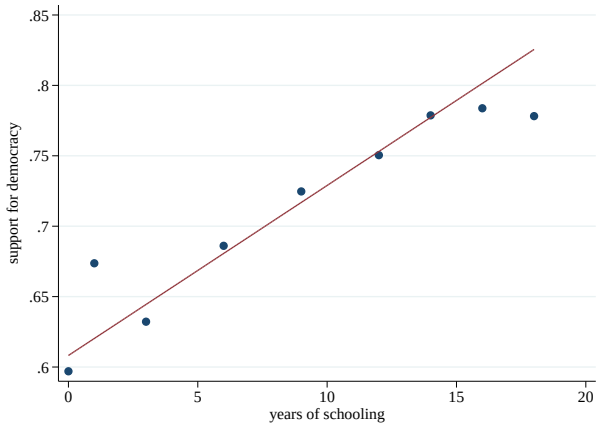
Table A.18: Support for democracy on years of schooling

	(1) support democ	(2) support democ	(3) support democ	(4) support democ
years of schooling	0.0121***	0.0109***	0.0133***	0.0137***
	(8.13)	(7.53)	(10.88)	(11.08)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.016	0.026	0.089	0.109
marginal R-squared	.016	.011	.013	.013
within R-squared		.025	.026	.026
N	104435	103383	103383	103383

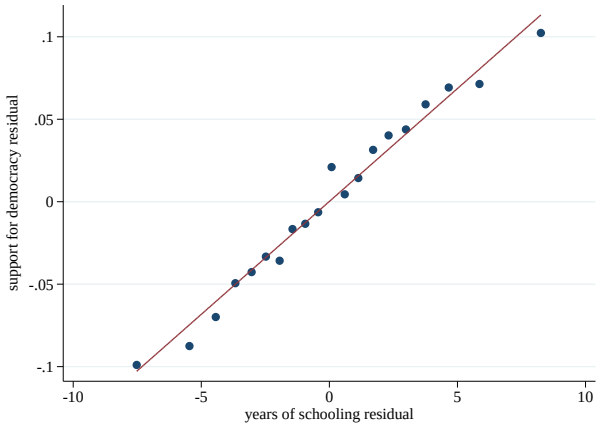
This table shows regression results for support for democracy on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's support for democracy. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the relationship conditional on individual controls without fixed effects. Column (3) adds survey fixed, column (4) adds region (defined by Afro) fixed effects. t-statistics based on standard errors clustered at the survey-level in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure A.17: Binned scatter plots

(a) Support for democracy, unconditional



(b) Support for democracy, conditional on controls and region FE



B Details on sample coverage and onstruction

B.1 Observations per census

Table B.1: Number of usable observations per census

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
country	**fraction**	**year**	**raw IPUMS**	**age data**	**schooling data**	**14+**	**observe "old"**	**14-18, observe "old"**	**illiterate old**	**14-18, illiterate old**	**nP**	**nD**	**have student**	**have urban**
Botswana	10	1981	97238	96187	72951	50399	14132	5802	12045	4657	21	21	yes	no
Botswana	10	1991	132623	132623	113172	78814	22878	10117	16394	6410	21	21	yes	yes
Botswana	10	2001	168676	168134	159257	109649	36006	14150	20412	6168	21	21	yes	no
Botswana	10	2011	201752	201235	190212	138375	40463	12794	16040	2786	21	21	yes	no
Burkina Faso	10	1985	884797	883447	484384	410398	0	0	0	0	13	45	no	no
Burkina Faso	10	1996	1081046	1075824	803264	552402	156495	77238	147669	71934	13	45	no	no
Burkina Faso	10	2006	1417824	1410123	1244291	770161	178501	103859	155898	92855	13	45	yes	yes
Cameroon	10	1976	736514	736320	605749	413814	78692	36652	71802	31786	7	39	yes	no
Cameroon	10	1987	897211	896649	763652	481727	93198	45721	72168	31980	7	39	no	yes
Cameroon	10	2005	1772359	1772359	1542200	1018632	311011	138181	141740	56272	7	39	yes	yes
Egypt	14.1	1986	6799093	6794386	5418332	4262426	1489558	649695	1187454	517672	24	235	no	yes
Egypt	10	1996	5902243	5901839	4453382	3810835	1331716	670174	1054871	516873	24	235	no	yes
Egypt	10	2006	7282434	7282434	5739722	5096618	1916007	753720	1239841	443837	24	235	yes	yes
Ethiopia	10	1984	3404306	3398027	2733575	1800650	379412	204664	368396	197342	12	63	yes	yes
Ethiopia	10	1994	5044598	5044597	4201616	2833214	793791	451168	749892	423495	12	63	yes	yes
Ethiopia	10	2007	7434086	7434086	1097614	744744	211838	121605	189209	107719	12	63	yes	yes
Ghana	10	1984	1309352	1309351	1050813	747642	271505	111672	204644	74998	10	102	yes	no
Ghana	10	2000	1894133	1894133	1730902	1152128	310913	129369	178820	63859	10	102	yes	yes
Ghana	10	2010	2466289	2466289	2262894	1575528	499171	200837	245681	87612	10	102	yes	yes
Guinea	10	1983	457837	457778	364805	275065	44403	22662	41557	20967	6	34	yes	yes
Guinea	10	1996	729071	727246	551619	397137	113872	44673	100445	37986	6	34	yes	yes
Kenya	6	1969	659310	659310	659310	394835	64079	32553	52685	25577	8	173	no	no
Kenya	6.7	1979	1033769	1031996	853843	593682	0	0	0	0	8	173	yes	yes
Kenya	5	1989	1074098	1072777	828512	578099	162592	87985	117835	58013	8	173	yes	yes
Kenya	5	1999	1407547	1407547	1191268	832083	215230	113599	120671	54811	8	173	yes	yes
Kenya	10	2009	3841935	3841935	3402695	2246737	657022	328455	284715	120285	8	173	yes	yes
Liberia	10	1974	150256	150256	127442	91811	0	0	0	0	9	50	yes	yes
Liberia	10	2008	348057	348057	294517	210111	59015	25494	35139	13253	9	50	yes	yes
Malawi	10	1987	798669	798193	657998	447247	72504	41684	62384	34856	24	223	yes	yes
Malawi	10	1998	991393	991393	826197	582694	109301	64674	78746	45320	24	223	yes	yes
Malawi	10	2008	1341977	1341046	1161773	736175	152064	89445	96963	57401	24	223	yes	yes
Mali	10	1987	785384	773407	582678	422837	111633	48553	104443	44574	9	242	no	no
Mali	10	1998	991330	986822	734156	519001	155183	68901	143926	63136	9	242	yes	yes
Mali	10	2009	1451856	1424140	1262277	776333	268699	120018	228707	101769	9	242	yes	yes
Morocco	5	1982	1012873	1012873	948008	571980	234908	94196	226850	89302	16	54	no	no
Morocco	5	1994	1294026	1293171	1293171	842330	406223	136345	376346	121039	16	54	no	no
Morocco	5	2004	1482720	1481076	1481076	1052531	514271	150544	455058	125690	16	54	no	no
Mozambique	10	1997	1551517	1550505	1248483	879255	199650	107787	184440	98072	11	143	yes	yes
Mozambique	10	2007	2047048	2047048	1616853	1103596	262286	133824	219556	110169	11	143	yes	yes
Nigeria	0.06	2006	83700	83700	82740	49282	14170	7555	7647	3975	38	37	yes	yes
Nigeria	0.06	2007	85183	85182	84122	49102	14465	7569	6958	3439	38	37	yes	yes
Nigeria	0.07	2008	107425	107425	105944	62151	19914	10092	9659	4786	38	37	yes	yes
Nigeria	0.05	2009	77896	77880	77650	45988	12867	6417	5588	2735	38	37	yes	yes
Nigeria	0.05	2010	72191	71991	58973	41830	14115	6679	6432	2918	38	37	yes	yes
Rwanda	10	1991	742918	742918	535602	372386	112661	58656	103350	52389	12	104	no	no
Rwanda	10	2002	843392	843392	645489	472153	142765	81951	112107	62684	12	104	yes	yes
Senegal	10	1988	700199	699981	527462	378289	103599	42459	90035	35464	9	28	yes	no
Senegal	10	2002	994562	994562	911891	594599	233001	82137	192271	64360	9	28	yes	yes
Sierra Leone	10	2004	494298	492922	395788	291916	94108	38245	71242	27146	14	90	yes	yes
South Africa	10	1996	3621164	3578019	3055995	2328067	753838	283482	330368	104373	9	216	yes	yes
South Africa	10	2001	3725655	3725655	3353684	2598672	880011	320148	381315	114613	9	216	yes	yes
South Africa	2	2007	1047657	1047657	842103	665305	234464	80288	89311	22387	9	216	yes	yes
South Africa	8.6	2011	4418594	4418594	3845633	3101908	919915	302412	273167	63522	9	216	yes	yes
South Sudan	7	2008	542765	542765	542333	295567	91326	41841	81749	36838	10	72	yes	yes
Sudan	16.6	2008	5066530	5066530	3902071	2790992	986776	441159	863532	385907	15	129	yes	yes
Tanzania	10	1988	2310424	2304474	1911308	1322841	263555	155341	218336	123646	23	113	yes	no
Tanzania	10	2002	3732735	3732735	3123724	2190557	494053	245738	290333	114181	23	113	yes	yes
Tanzania	10	2012	4498022	4498022	3918823	2603099	665506	327262	260554	91926	23	113	yes	yes
Uganda	10	1991	1548460	1547604	1242885	855537	183396	97908	135688	65958	38	159	yes	yes
Uganda	10	2002	2497449	2497449	2042838	1355857	304094	183083	170612	95258	38	159	yes	yes
Zambia	10	1990	787461	787461	664239	460486	142016	75070	90770	42762	8	72	yes	yes
Zambia	10	2000	996117	996117	825110	570022	192384	93412	85575	35003	8	72	yes	yes
Zambia	10	2010	1321973	1321973	1028628	704471	227855	117903	67399	29460	8	72	yes	no
TOTAL							19,039,076	8,555,617	12,947,440	5,546,205				

The first two columns give country and census-year. Column (3) shows the fraction of the census sampled by IPUMS. Column (4) is the number of observation in the original IPUMS data without restrictions. Columns (5)-(11) give the number of individuals with observations for successively tighter sample restrictions: (5) requires that age be observable, (6) in addition requires data on education, (7) in addition requires that the individual be at least 14 years old, (8) requires that the individual be co-habiting with at least one individual of an older generation, (9) has the same requirement as (8) but for individuals aged 14-18, (10) is the same as (8) but restricts the sample to children of illiterate parents, and (11) is the same as (9) again restricted to children of illiterate parents. Columns (12)-(13) show the number of geographic sub-units available for each census. (12) the number of (admin-1) provinces, and (13) the number of (admin-2) districts. Column (14) indicates whether the census has data on individual student status and (15) indicates whether the census has data on urban vs. rural residence.

Table B.2: Number of observations in migrant sample

age range	migrant status	number of observations
full migrant sample		
14-18	all	6902760
14-100	all	15315567
14-18	non-migrant	6202177
14-100	non-migrant	13577655
14-18	migrant	700583
14-100	migrant	1737912
migrant sample with years of residence		
14-18	all	4862880
14-100	all	10805595
14-18	non-migrant	4396066
14-100	non-migrant	9661701
14-18	migrant	466814
14-100	migrant	1143894

This table shows the number of observations for the full sample of individuals for whom migrant status is available as well as the sub-sample of individuals for whom both migrant status as well as time of migration data are available.

B.2 Variable construction for IM

IPUMS provides a variable for the line number of father and mother in the household, but this variable exists for only one third of all observations in the sample, and far fewer of adults with completed schooling. To maximize the number of observation for whom we observe education in a previous generation, we therefore use the variable "relationship to household head" to identify the educational attainment of the generation previous to that of any given individual. This variable takes on 32 different values. We use these different categories to assign each individual to a "generation" within the household. Based on the generation assignment, each individual is assigned a value for s_1 (her own education) and a value for s_0 (the mean education level of individuals within the household of the generation immediately above). That is, an individual of generation "1" would be assigned the mean of the education of head, spouse, siblings of the head, and cousins of the head.

Table B.3: Relationship to household head and generation assignment

relationship to head	meaning	generation	relationship to head	meaning	generation
1000	Head	0	4500	Grandparent	-2
2000	Spouse/partner	0	4600	Parent/grandparent/ascendant	-1
3000	Child	1	4700	Aunt/uncle	-1
3100	Biological child	1	4810	Nephew/niece	1
3200	Adopted child	1	4820	Cousin	0
3300	Stepchild	1	4900	Other relative, not elsewhere classified	
4000	Other relative		5000	Non-relative	
4100	Grandchild	2	5100	Friend/guest/visitor/partner	
4110	Grandchild or great grandchild	2	5120	Visitor	
4200	Parent/parent-in-law	-1	5200	Employee	
4210	Parent	-1	5210	Domestic employee	
4220	Parent-in-law	-1	5330	Foster child	1
4300	Child-in-law	1	5600	Group quarters	
4400	Sibling/sibling-in-law	0	5900	Non-relative, n.e.c.	
4410	Sibling	0	6000	Other relative or non-relative	
4430	Sibling-in-law	0	9999	Unknown	

Table B.4: Intergenerational links

	previous generation observed			previous generation's education observed		
	frequency	percent	cumulative	frequency	percent	cumulative
2 parents	5,621,042	49.92	49.92	5,260,792	49.59	49.59
1 parent	4,514,207	40.09	90.02	4,263,777	40.19	89.78
2 parents, others	150,227	1.33	91.35	143,800	1.36	91.13
1 parent, others	112,033	1.00	92.35	108,248	1.02	92.15
1 other	346,289	3.08	95.42	332,538	3.13	95.29
2+ others	515,531	4.58	100.00	500,141	4.71	100.00
	11,259,329			10,609,296		

Frequency table for intergenerational links. "2 parents" means that the two individuals observed in the previous generation are an individual's parents. "Others" means that the individuals are either not parents (aunts, uncles, parents in law etc.) or that they could be but cannot be clearly identified as an individual's parents, for example if the individual is a niece of the head and the individuals in the older generation include siblings of the head.

Table B.5: Correlation of previous generation average education with mothers', fathers', and parental average education

		(1)	(2)	(3)
estimation	variable	mothers	fathers	parental average
unconditional	years of schooling	0.933	0.932	0.993
unconditional	educational attainment, fine	0.931	0.928	0.972
unconditional	educational attainment, coarse	0.872	0.919	0.985
census FEs	years of schooling	0.925	0.889	0.989
census FEs	educational attainment, fine	0.906	0.927	1.025
census FEs	educational attainment, coarse	0.860	0.885	0.981

This table shows standardized ("beta") coefficients from regressing our measures of previous generation eduction for individual i on education of only the parents of individual i. In column (1), we regress our measure on the education of only the mother of individual i, in column (2) on that of the father, and in column (3) on the average of both parents. As with our measure, we allow fractional values for average years of schooling and round educational attainment (coarse and fine) to the nearest integer. The first three rows show unconditional estimates (hence the figures in the table are simply unconditional correlations), whereas rows four to six show estimates conditional on census (country-year) fixed effects.

C Country-level IM

C.1 Country-level estimates

Table C.1: Country-level estimates of IM

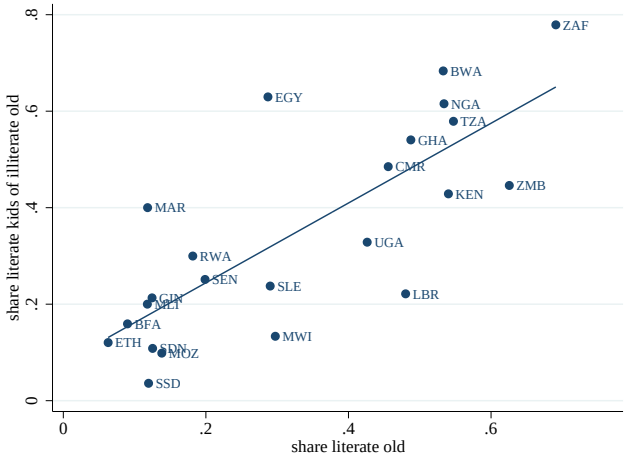
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
South Africa	0.779	0.791	0.739	0.591	0.690	0.314	0.884	0.837	0.774	0.739	0.819	0.267
Botswana	0.683	0.694	0.634	0.597	0.701	0.306	0.778	0.735	0.664	0.702	0.812	0.247
Egypt	0.630	0.622	0.600	0.501	0.558	0.196	0.630	0.622	0.600	0.506	0.616	0.089
Nigeria	0.615	0.670	0.664	0.330	0.506	0.174	0.679	0.707	0.698	0.455	0.626	0.118
Tanzania	0.579	0.627	0.618	0.544	0.708	0.362	0.705	0.693	0.672	0.540	0.741	0.233
Ghana	0.540	0.518	0.488	0.285	0.430	0.081	0.620	0.554	0.516	0.489	0.586	0.053
Cameroon	0.485	0.488	0.467	0.281	0.417	0.070	0.567	0.533	0.503	0.381	0.507	-0.013
Zambia	0.446	0.470	0.469	0.185	0.397	0.073	0.560	0.532	0.522	0.440	0.586	0.072
Kenya	0.429	0.508	0.508	0.113	0.326	0.030	0.601	0.606	0.592	0.292	0.463	-0.015
Morocco	0.400	0.368	0.339	0.140	0.239	-0.093	0.400	0.368	0.339	0.343	0.406	-0.110
Uganda	0.328	0.369	0.363	0.319	0.455	0.104	0.550	0.507	0.487	0.439	0.585	0.066
Rwanda	0.300	0.342	0.335	0.284	0.423	0.064	0.388	0.395	0.386	0.277	0.471	-0.047
Senegal	0.251	0.255	0.246	0.252	0.354	0.019	0.290	0.274	0.261	0.139	0.329	-0.160
Sierra Leone	0.238	0.244	0.236	-0.139	0.039	-0.252	0.418	0.339	0.310	0.243	0.296	-0.199
Liberia	0.221	0.309	0.317	-0.032	0.149	-0.178	0.369	0.409	0.401	0.208	0.361	-0.154
Guinea	0.213	0.209	0.197	0.092	0.149	-0.189	0.284	0.247	0.228	0.174	0.247	-0.260
Mali	0.200	0.193	0.179	-0.037	0.085	-0.247	0.254	0.221	0.203	0.040	0.160	-0.343
Burkina Faso	0.159	0.166	0.155	-0.038	0.056	-0.284	0.159	0.166	0.155	-0.035	0.107	-0.396
Malawi	0.133	0.209	0.209	-0.076	0.112	-0.231	0.455	0.424	0.411	0.316	0.404	-0.126
Ethiopia	0.120	0.150	0.145	-0.138	0.034	-0.287	0.182	0.199	0.190	0.107	0.230	-0.278
Sudan	0.108	0.180	0.195	-0.145	0.021	-0.301	0.108	0.180	0.195	-0.053	0.131	-0.365
Mozambique	0.099	0.153	0.152	-0.183	0.027	-0.295	0.099	0.153	0.152	-0.049	0.140	-0.362
South Sudan	0.036	0.072	0.076	-0.215	-0.082	-0.417	0.036	0.072	0.076	-0.123	0.027	-0.482
age range	14-18	14-30	14-100	14-18	14-30	14-100	14-18	14-30	14-100	14-18	14-30	14-100
fixed effects	none	none	none	y, b	y, b	y, b	none	none	none	y, b	y, b	y, b
student correction	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes

This table shows country-level estimates of IM (likelihood that children of illiterate parents become literate). Higher numbers → higher IM. "age range" indicates the range of ages for children in the sample. fixed effects indicates whether estimates are unconditional or conditional on census-year (y) and birth-decade for young and old (b) fixed effects. "student-correction" indicates whether individuals still in school with 4 or 5 years of schooling are classified as literate or not. Countries sorted by column (1).

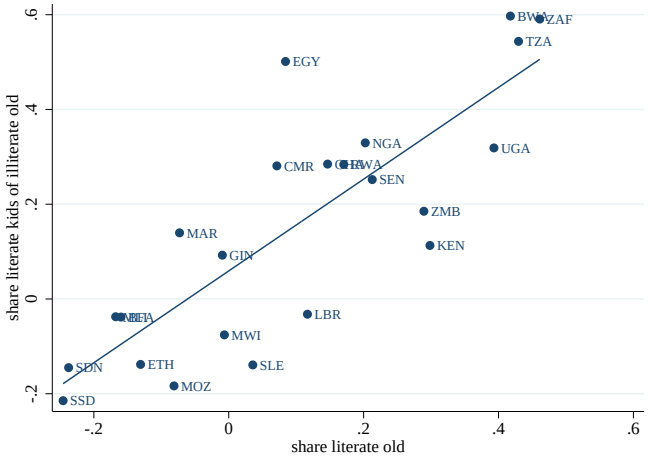
C.2 Country-level literacy and IM

Figure C.1: Literacy and IM at the country-level

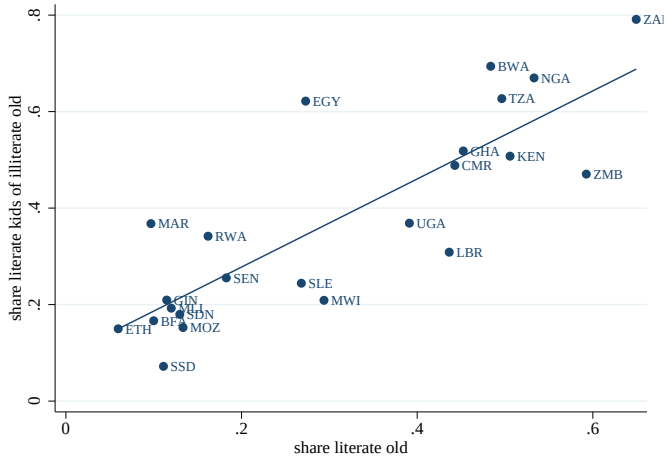
(a) unconditional, ages 14-18



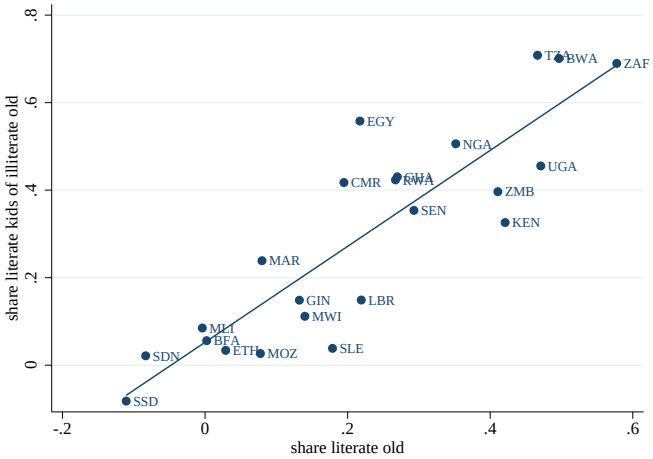
(b) year + birth-decade FEs, ages 14-18



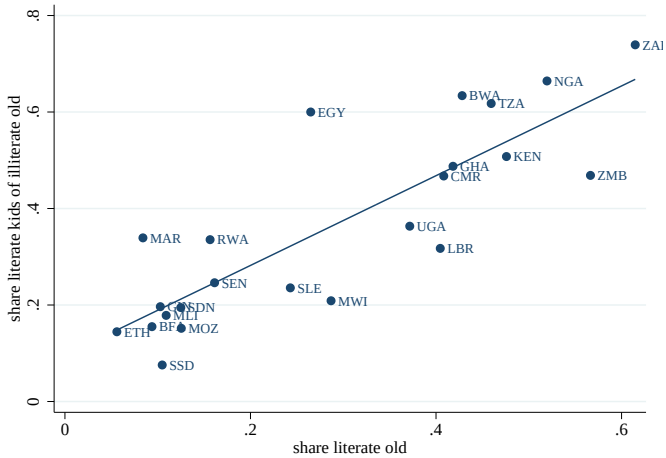
(c) unconditional, ages 14-30



(d) year + birth-decade FEs, ages 14-30



(e) unconditional, ages 14-100



(f) year + birth-decade FEs, ages 14-100

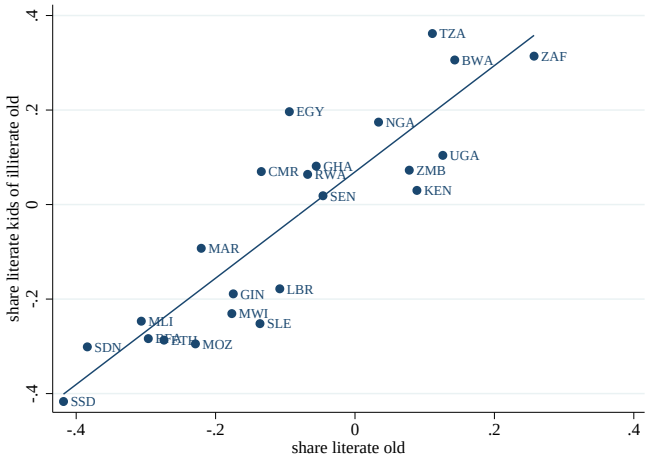
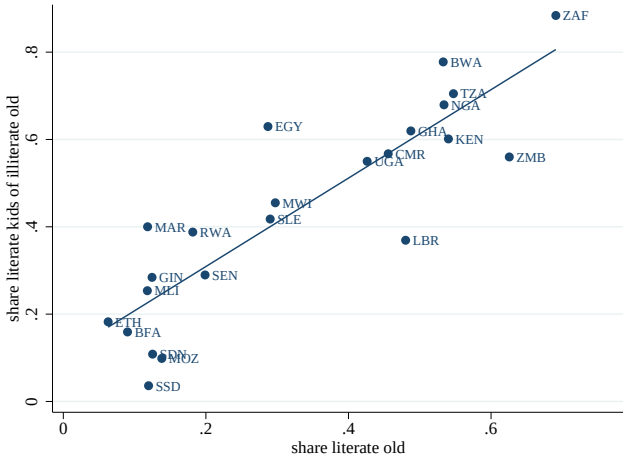
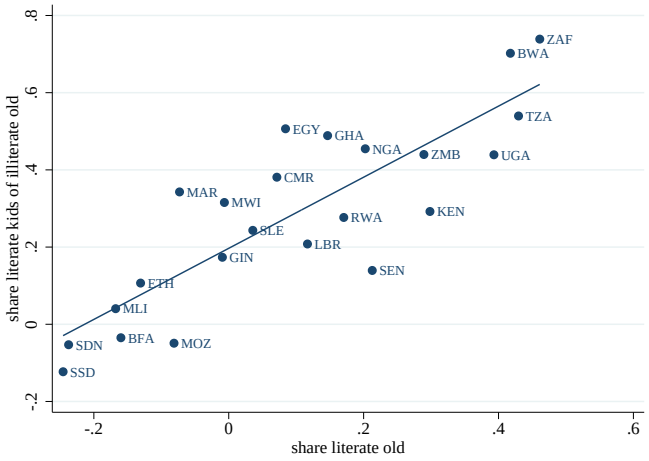


Figure C.2: Literacy and IM at the country-level, with student correction

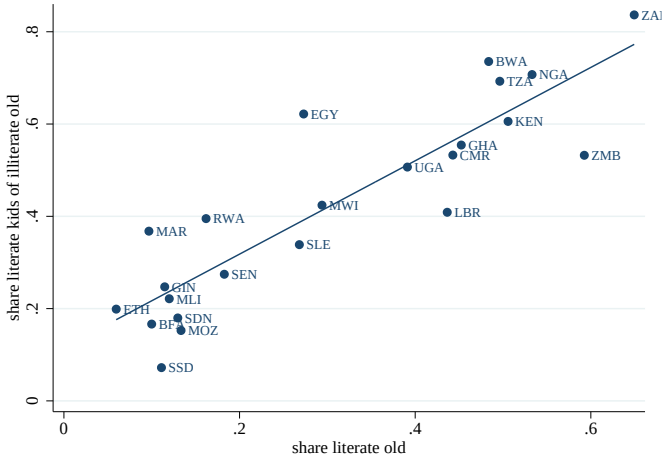
(a) unconditional, ages 14-18



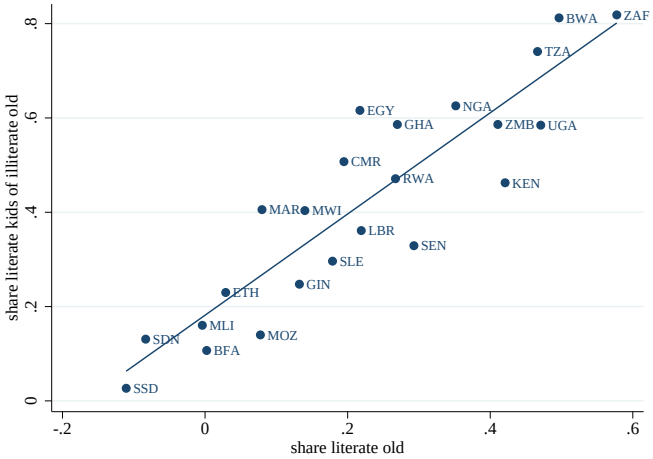
(b) year + birth-decade FEs, ages 14-18



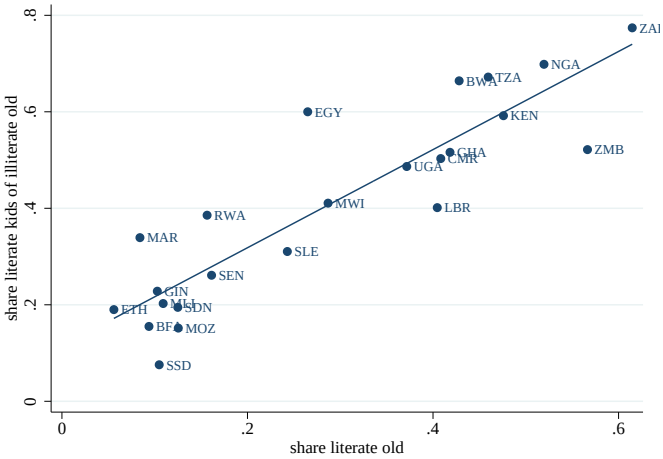
(c) unconditional, ages 14-30



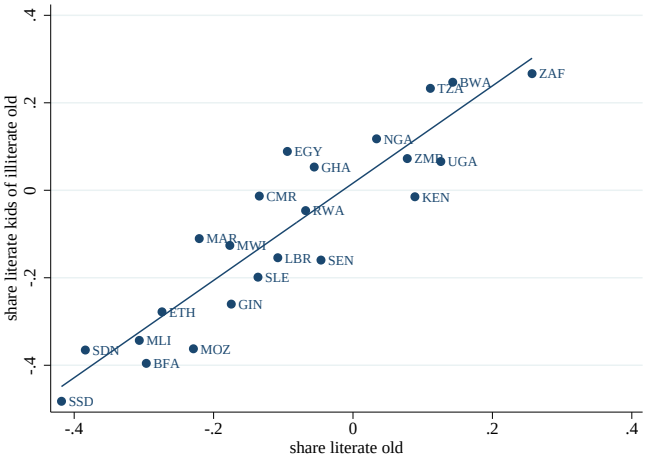
(d) year + birth-decade FEs, ages 14-30



(e) unconditional, ages 14-100



(f) year + birth-decade FEs, ages 14-100



D Movers

D.1 Within-family estimates

Table D.1: Country-by-country household fixed effects estimates of location-of-birth effects with 2-children migrant families, minimum age gap = 5

country		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Burkina Faso	point estimate	1.707***	0.720***	1.176***	0.496***	1.375***	0.518***	0.930***	0.360***
	standard error	(0.207)	(0.249)	(0.083)	(0.173)	(0.047)	(0.066)	(0.042)	(0.045)
	R-squared	0.146	0.733	0.150	0.733	0.187	0.691	0.187	0.691
	within R-squared	0.122	0.036	0.126	0.036	0.109	0.022	0.109	0.022
	N	749	749	749	749	9433	9433	9433	9433
	n birth-regions	45	45	45	45	45	45	45	45
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Botswana	point estimate	0.185	-0.052	-0.980	-0.555	0.242*	0.025	0.839***	0.275
	standard error	(0.317)	(0.214)	(0.647)	(0.694)	(0.139)	(0.126)	(0.189)	(0.304)
	R-squared	0.094	0.622	0.102	0.624	0.119	0.639	0.126	0.639
	within R-squared	0.038	0.052	0.047	0.057	0.010	0.002	0.018	0.003
	N	67	67	67	67	1760	1760	1760	1760
	n birth-regions	14	14	14	14	21	21	21	21
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Cameroon	point estimate	0.678***	-0.199*	0.732***	-0.073	1.049***	0.106***	0.962***	0.126***
	standard error	(0.073)	(0.119)	(0.069)	(0.145)	(0.052)	(0.038)	(0.020)	(0.040)
	R-squared	0.155	0.725	0.167	0.723	0.312	0.717	0.337	0.717
	within R-squared	0.114	0.007	0.127	0.003	0.232	0.038	0.260	0.038
	N	981	981	981	981	16317	16317	16317	16317
	n birth-regions	192	192	192	192	230	230	230	230
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Egypt	point estimate	0.300**	0.078	0.590***	0.120	0.514***	0.252***	0.736***	0.279***
	standard error	(0.131)	(0.073)	(0.162)	(0.107)	(0.073)	(0.023)	(0.055)	(0.035)
	R-squared	0.015	0.668	0.020	0.668	0.059	0.645	0.063	0.645
	within R-squared	0.007	0.008	0.012	0.008	0.020	0.025	0.024	0.024
	N	4106	4106	4106	4106	54437	54437	54437	54437
	n birth-regions	26	26	26	26	27	27	27	27
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Ethiopia	point estimate	-10.688***	-7.113	0.821*	0.629	0.734	-1.344*	0.475***	0.387***
	standard error	(3.021)	(7.539)	(0.451)	(1.090)	(0.949)	(0.764)	(0.102)	(0.143)
	R-squared	0.345	0.698	0.221	0.689	0.181	0.736	0.200	0.739
	within R-squared	0.280	0.138	0.144	0.112	0.064	0.149	0.086	0.159
	N	14	14	14	14	588	588	588	588
	n birth-regions	10	10	10	10	80	80	80	80
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5

The dependent variable is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. In columns (1), (2), and (5), (6), the RHS variable of interest is parental literacy estimated among non-migrants net of census-year and birth decade (for young and old) fixed effects. In columns (3), (4), and (7), (8), the RHS variable of interest is the share of literate children of illiterate old in the birth-region net of census-year and young and old birth-decade fixed effects. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table D.2: Country-by-country household fixed effects estimates of location-of-birth effects with 2-children migrant families, minimum age gap = 5

country		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ghana	point estimate	0.423***	0.104	0.727***	0.215***	0.690***	0.217***	0.775***	0.249***
	standard error	(0.111)	(0.064)	(0.095)	(0.077)	(0.037)	(0.045)	(0.089)	(0.059)
	R-squared	0.093	0.698	0.099	0.699	0.099	0.618	0.092	0.618
	within R-squared	0.034	0.011	0.041	0.013	0.070	0.038	0.062	0.038
	N	1072	1072	1072	1072	16321	16321	16321	16321
	n birth-regions	10	10	10	10	10	10	10	10
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Guinea	point estimate	0.400*	0.055	0.265**	-0.002	1.193***	0.744***	0.597***	0.366***
	standard error	(0.208)	(0.160)	(0.119)	(0.095)	(0.078)	(0.102)	(0.036)	(0.050)
	R-squared	0.072	0.658	0.075	0.658	0.164	0.606	0.167	0.606
	within R-squared	0.023	0.024	0.026	0.023	0.087	0.055	0.091	0.054
	N	415	415	415	415	5661	5661	5661	5661
	n birth-regions	34	34	34	34	34	34	34	34
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Kenya	point estimate	0.433***	-0.054	0.548***	-0.058	0.556***	0.066	0.642***	0.064
	standard error	(0.073)	(0.126)	(0.087)	(0.148)	(0.042)	(0.041)	(0.044)	(0.051)
	R-squared	0.101	0.624	0.104	0.624	0.124	0.643	0.127	0.643
	within R-squared	0.039	0.000	0.043	0.000	0.055	0.009	0.058	0.009
	N	789	789	789	789	11304	11304	11304	11304
	n birth-regions	133	133	133	133	178	178	178	178
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Liberia	point estimate	0.302	-0.100	0.690**	0.299	0.408***	0.196*	0.678***	0.303*
	standard error	(0.194)	(0.326)	(0.288)	(0.274)	(0.093)	(0.116)	(0.150)	(0.162)
	R-squared	0.122	0.672	0.141	0.675	0.070	0.629	0.084	0.630
	within R-squared	0.044	0.055	0.065	0.063	0.056	0.073	0.070	0.075
	N	75	75	75	75	1002	1002	1002	1002
	n birth-regions	13	13	13	13	15	15	15	15
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Mali	point estimate	1.259***	0.479***	0.907***	0.346***	0.948***	0.289***	0.638***	0.200***
	standard error	(0.070)	(0.123)	(0.050)	(0.093)	(0.052)	(0.042)	(0.032)	(0.029)
	R-squared	0.187	0.748	0.192	0.747	0.142	0.649	0.145	0.649
	within R-squared	0.136	0.047	0.142	0.046	0.088	0.033	0.090	0.033
	N	933	933	933	933	15250	15250	15250	15250
	n birth-regions	45	45	45	45	47	47	47	47
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5

The dependent variable is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. In columns (1), (2), and (5), (6), the RHS variable of interest is parental literacy estimated among non-migrants net of census-year and birth decade (for young and old) fixed effects. In columns (3), (4), and (7), (8), the RHS variable of interest is the share of literate children of illiterate old in the birth-region net of census-year and young and old birth-decade fixed effects. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table D.3: Country-by-country household fixed effects estimates of location-of-birth effects with 2-children migrant families, minimum age gap = 5

country		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mozambique	point estimate	0.391***	-0.062	0.682***	-0.073	0.839***	0.145***	0.809***	0.141***
	standard error	(0.062)	(0.089)	(0.084)	(0.158)	(0.043)	(0.037)	(0.027)	(0.037)
	R-squared	0.139	0.639	0.146	0.639	0.102	0.603	0.108	0.603
	within R-squared	0.030	0.022	0.037	0.022	0.069	0.032	0.075	0.032
	N	1814	1814	1814	1814	19599	19599	19599	19599
	n birth-regions	133	133	133	133	139	139	139	139
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Malawi	point estimate	0.481**	0.021	0.514**	-0.004	0.434***	-0.116	0.629***	-0.047
	standard error	(0.196)	(0.299)	(0.212)	(0.307)	(0.136)	(0.115)	(0.116)	(0.129)
	R-squared	0.091	0.642	0.090	0.642	0.067	0.598	0.075	0.598
	within R-squared	0.018	0.007	0.017	0.007	0.017	0.008	0.027	0.007
	N	384	384	384	384	2384	2384	2384	2384
	n birth-regions	31	31	31	31	31	31	31	31
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Rwanda	point estimate	0.771	0.816	1.411	1.562**	0.572***	0.068	0.996***	0.106
	standard error	(0.630)	(0.566)	(0.868)	(0.686)	(0.075)	(0.168)	(0.141)	(0.314)
	R-squared	0.189	0.624	0.192	0.626	0.042	0.544	0.042	0.544
	within R-squared	0.008	0.014	0.011	0.020	0.009	0.004	0.009	0.004
	N	270	270	270	270	1899	1899	1899	1899
	n birth-regions	12	12	12	12	12	12	12	12
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Sudan and South Sudan	point estimate	0.553***	-0.158	0.694***	-0.203	0.847***	0.159***	0.664***	0.114***
	standard error	(0.147)	(0.124)	(0.156)	(0.141)	(0.096)	(0.042)	(0.075)	(0.031)
	R-squared	0.082	0.682	0.090	0.683	0.116	0.646	0.122	0.646
	within R-squared	0.031	0.014	0.040	0.014	0.062	0.004	0.069	0.004
	N	696	696	696	696	19052	19052	19052	19052
	n birth-regions	25	25	25	25	25	25	25	25
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Senegal	point estimate	0.618***	0.031	0.743***	0.013	0.870***	0.170***	0.776***	0.147***
	standard error	(0.115)	(0.080)	(0.065)	(0.096)	(0.135)	(0.030)	(0.053)	(0.024)
	R-squared	0.110	0.692	0.133	0.692	0.133	0.649	0.144	0.648
	within R-squared	0.060	0.021	0.085	0.021	0.102	0.032	0.114	0.032
	N	1151	1151	1151	1151	22474	22474	22474	22474
	n birth-regions	34	34	34	34	34	34	34	34
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5

The dependent variable is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. In columns (1), (2), and (5), (6), the RHS variable of interest is parental literacy estimated among non-migrants net of census-year and birth decade (for young and old) fixed effects. In columns (3), (4), and (7), (8), the RHS variable of interest is the share of literate children of illiterate old in the birth-region net of census-year and young and old birth-decade fixed effects. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table D.4: Country-by-country household fixed effects estimates of location-of-birth effects with 2-children migrant families, minimum age gap = 5

country		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sierra Leone	point estimate	0.876***	0.739***	0.896***	0.628***	0.882***	0.359***	0.879***	0.324***
	standard error	(0.128)	(0.220)	(0.151)	(0.213)	(0.035)	(0.079)	(0.033)	(0.068)
	R-squared	0.181	0.691	0.187	0.689	0.196	0.629	0.203	0.629
	within R-squared	0.143	0.091	0.149	0.084	0.131	0.063	0.138	0.062
	N	230	230	230	230	4042	4042	4042	4042
	n birth-regions	70	70	70	70	90	90	90	90
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Tanzania	point estimate	0.833***	0.256	0.798***	0.137	0.814***	0.059	0.933***	0.028
	standard error	(0.144)	(0.167)	(0.171)	(0.252)	(0.141)	(0.056)	(0.094)	(0.067)
	R-squared	0.051	0.678	0.046	0.677	0.057	0.646	0.061	0.646
	within R-squared	0.022	0.003	0.016	0.001	0.022	0.015	0.026	0.015
	N	1364	1364	1364	1364	19916	19916	19916	19916
	n birth-regions	30	30	30	30	30	30	30	30
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Uganda	point estimate	0.100	0.085	0.479*	0.030	0.566***	0.107	0.893***	0.148
	standard error	(0.148)	(0.177)	(0.268)	(0.329)	(0.082)	(0.088)	(0.152)	(0.115)
	R-squared	0.063	0.644	0.069	0.644	0.083	0.608	0.086	0.608
	within R-squared	0.009	0.018	0.015	0.018	0.041	0.045	0.044	0.045
	N	532	532	532	532	5969	5969	5969	5969
	n birth-regions	48	48	48	48	56	56	56	56
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
South Africa	point estimate	0.093	0.004	1.046***	-0.146	0.093	0.028*	0.523***	0.107**
	standard error	(0.136)	(0.058)	(0.200)	(0.198)	(0.065)	(0.015)	(0.067)	(0.049)
	R-squared	0.026	0.539	0.037	0.539	0.087	0.607	0.091	0.607
	within R-squared	0.001	0.003	0.013	0.003	0.002	0.001	0.007	0.001
	N	1022	1022	1022	1022	16473	16473	16473	16473
	n birth-regions	9	9	9	9	9	9	9	9
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5
Zambia	point estimate	0.477***	0.201**	0.581***	0.225**	0.623***	0.231***	0.616***	0.232***
	standard error	(0.080)	(0.083)	(0.085)	(0.096)	(0.027)	(0.029)	(0.034)	(0.028)
	R-squared	0.052	0.645	0.056	0.645	0.085	0.580	0.084	0.580
	within R-squared	0.022	0.017	0.027	0.017	0.039	0.015	0.038	0.015
	N	1549	1549	1549	1549	21325	21325	21325	21325
	n birth-regions	71	71	71	71	72	72	72	72
	household fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
	maximum age	21	21	21	21	100	100	100	100
	minimum age gap	5	5	5	5	5	5	5	5

The dependent variable is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. In columns (1), (2), and (5), (6), the RHS variable of interest is parental literacy estimated among non-migrants net of census-year and birth decade (for young and old) fixed effects. In columns (3), (4), and (7), (8), the RHS variable of interest is the share of literate children of illiterate old in the birth-region net of census-year and young and old birth-decade fixed effects. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

E Correlates

E.1 Country fixed effects

Table E.1: District and birth-region-level correlates of the share of literate old and IM, ages 14-18, country fixed effects, without student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.314***	-0.228***	-0.064***	2437	-0.497***	-0.345***	-0.056**	1194
	(0.036)	(0.025)	(0.020)		(0.094)	(0.065)	(0.027)	
DBORD	0.028	-0.013	-0.028***	2437	0.066	0.018	-0.023	1194
	(0.029)	(0.019)	(0.011)		(0.096)	(0.065)	(0.022)	
DCOAST	-0.171***	-0.153***	-0.060***	2437	-0.371***	-0.334***	-0.120***	1194
	(0.045)	(0.029)	(0.020)		(0.128)	(0.097)	(0.031)	
MALTR	-0.300***	-0.189***	-0.025	2436	-0.391***	-0.270***	-0.034	1193
	(0.069)	(0.048)	(0.021)		(0.101)	(0.063)	(0.037)	
AGSUIT	0.012	0.062*	0.055***	2407	-0.039	-0.023	0.000	1183
	(0.057)	(0.035)	(0.019)		(0.151)	(0.091)	(0.031)	
ELEV	0.035	-0.027	-0.046**	2428	0.064	0.028	-0.011	1193
	(0.060)	(0.033)	(0.019)		(0.111)	(0.075)	(0.032)	
TERRUG	0.125**	0.101***	0.033*	2435	0.230**	0.213***	0.076**	1193
	(0.053)	(0.031)	(0.019)		(0.101)	(0.060)	(0.030)	
OILDUM	0.003	0.012	0.010	2437	0.105*	0.080*	0.016	1194
	(0.023)	(0.018)	(0.010)		(0.063)	(0.042)	(0.012)	
DIADUM	-0.015	-0.021**	-0.012	2437	-0.019	-0.018	-0.006	1194
	(0.015)	(0.009)	(0.009)		(0.017)	(0.012)	(0.009)	
history								
POPD	0.266***	0.204***	0.062***	2435	0.387***	0.317***	0.096***	1193
	(0.063)	(0.042)	(0.019)		(0.072)	(0.039)	(0.026)	
DRRD	-0.348***	-0.249***	-0.059***	1932	-0.542***	-0.383***	-0.077***	1156
	(0.054)	(0.030)	(0.016)		(0.087)	(0.061)	(0.022)	
DROAD	-0.318***	-0.194***	-0.024**	2148	-0.364***	-0.244***	-0.027	1165
	(0.038)	(0.023)	(0.011)		(0.061)	(0.038)	(0.022)	
DCMISS	-0.404***	-0.272***	-0.053**	2437	-0.512***	-0.373***	-0.074***	1194
	(0.070)	(0.047)	(0.022)		(0.113)	(0.083)	(0.028)	
DPMISS	-0.395***	-0.272***	-0.064***	2437	-0.596***	-0.423***	-0.083***	1194
	(0.052)	(0.030)	(0.018)		(0.100)	(0.065)	(0.025)	
DPCEM	0.019	-0.014	-0.024	2437	-0.052	-0.026	0.006	1194
	(0.053)	(0.027)	(0.023)		(0.072)	(0.044)	(0.027)	
DPCST	-0.019	-0.032	-0.021	2437	-0.134	-0.153***	-0.073**	1194
	(0.045)	(0.024)	(0.015)		(0.088)	(0.057)	(0.029)	
contemporary								
URBSHR	0.423***	0.202***	-0.044***	2386	0.386***	0.291***	0.064**	1170
	(0.024)	(0.018)	(0.017)		(0.036)	(0.028)	(0.031)	
AGSHR	-0.686***	-0.393***	-0.055	2264	-0.544***	-0.381***	-0.091**	1006
	(0.028)	(0.018)	(0.035)		(0.064)	(0.042)	(0.044)	
MANSHR	0.267***	0.136***	-0.007	2264	0.344***	0.212***	0.013	1006
	(0.045)	(0.030)	(0.017)		(0.079)	(0.055)	(0.024)	
SERSHR	0.635***	0.377***	0.074**	2264	0.518***	0.368***	0.090*	1006
	(0.032)	(0.019)	(0.032)		(0.064)	(0.039)	(0.046)	
ETHFRAG	0.210***	0.075***	-0.033*	1504	0.135***	0.071**	0.004	567
	(0.042)	(0.025)	(0.018)		(0.041)	(0.030)	(0.015)	
ETHPOL	0.090***	0.017	-0.028*	1504	0.069**	0.030	-0.005	567
	(0.031)	(0.019)	(0.017)		(0.029)	(0.023)	(0.013)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include country fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.2: District and birth-region-level correlates of the share of literate old and IM, ages 14-18, country fixed effects, with student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.314***	-0.184***	-0.027	2437	-0.497***	-0.286***	-0.004	1194
	(0.036)	(0.022)	(0.019)		(0.094)	(0.061)	(0.025)	
DBORD	0.028	-0.012	-0.026***	2437	0.066	0.011	-0.027	1194
	(0.029)	(0.016)	(0.009)		(0.096)	(0.057)	(0.020)	
DCOAST	-0.171***	-0.123***	-0.037**	2437	-0.371***	-0.288***	-0.086***	1194
	(0.045)	(0.025)	(0.017)		(0.128)	(0.085)	(0.024)	
MALTR	-0.300***	-0.172***	-0.021	2436	-0.391***	-0.243***	-0.023	1193
	(0.069)	(0.047)	(0.019)		(0.101)	(0.066)	(0.035)	
AGSUIT	0.012	0.062*	0.056***	2407	-0.039	-0.013	0.010	1183
	(0.057)	(0.035)	(0.018)		(0.151)	(0.087)	(0.030)	
ELEV	0.035	-0.008	-0.025	2428	0.064	0.043	0.007	1193
	(0.060)	(0.032)	(0.018)		(0.111)	(0.069)	(0.028)	
TERRUG	0.125**	0.096***	0.033*	2435	0.230**	0.202***	0.075***	1193
	(0.053)	(0.031)	(0.018)		(0.101)	(0.055)	(0.028)	
OILDUM	0.003	0.010	0.008	2437	0.105*	0.066*	0.006	1194
	(0.023)	(0.015)	(0.009)		(0.063)	(0.035)	(0.013)	
DIADUM	-0.015	-0.017**	-0.009	2437	-0.019	-0.013	-0.002	1194
	(0.015)	(0.008)	(0.008)		(0.017)	(0.011)	(0.008)	
history								
POPD	0.266***	0.178***	0.046***	2435	0.387***	0.286***	0.080***	1193
	(0.063)	(0.038)	(0.015)		(0.072)	(0.042)	(0.021)	
DRRD	-0.348***	-0.209***	-0.030*	1932	-0.542***	-0.337***	-0.049**	1156
	(0.054)	(0.031)	(0.016)		(0.087)	(0.056)	(0.019)	
DROAD	-0.318***	-0.171***	-0.015	2148	-0.364***	-0.221***	-0.019	1165
	(0.038)	(0.022)	(0.011)		(0.061)	(0.038)	(0.020)	
DCMISS	-0.404***	-0.216***	-0.010	2437	-0.512***	-0.321***	-0.037	1194
	(0.070)	(0.046)	(0.023)		(0.113)	(0.076)	(0.027)	
DPMISS	-0.395***	-0.240***	-0.045***	2437	-0.596***	-0.387***	-0.067***	1194
	(0.052)	(0.031)	(0.017)		(0.100)	(0.062)	(0.022)	
DPCEM	0.019	-0.019	-0.029	2437	-0.052	-0.047	-0.017	1194
	(0.053)	(0.027)	(0.022)		(0.072)	(0.047)	(0.023)	
DPCST	-0.019	-0.045**	-0.036**	2437	-0.134	-0.151***	-0.077***	1194
	(0.045)	(0.023)	(0.015)		(0.088)	(0.050)	(0.024)	
contemporary								
URBSHR	0.423***	0.168***	-0.069***	2386	0.386***	0.231***	0.014	1170
	(0.024)	(0.016)	(0.014)		(0.036)	(0.027)	(0.031)	
AGSHR	-0.686***	-0.341***	-0.016	2264	-0.544***	-0.323***	-0.052	1006
	(0.028)	(0.015)	(0.036)		(0.064)	(0.042)	(0.048)	
MANSHR	0.267***	0.109***	-0.023	2264	0.344***	0.174***	-0.009	1006
	(0.045)	(0.027)	(0.016)		(0.079)	(0.049)	(0.024)	
SERSHR	0.635***	0.329***	0.039	2264	0.518***	0.315***	0.057	1006
	(0.032)	(0.016)	(0.032)		(0.064)	(0.040)	(0.051)	
ETHFRAG	0.210***	0.057**	-0.045***	1504	0.135***	0.049**	-0.014	567
	(0.042)	(0.024)	(0.016)		(0.041)	(0.025)	(0.014)	
ETHPOL	0.090***	0.008	-0.035**	1504	0.069**	0.017	-0.015	567
	(0.031)	(0.019)	(0.016)		(0.029)	(0.021)	(0.013)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include country fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.3: District and birth-region-level correlates of the share of literate old and IM, ages 14-100, country fixed effects, without student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.263***	-0.187***	-0.043***	2437	-0.367***	-0.361***	-0.042*	1194
	(0.032)	(0.020)	(0.015)		(0.062)	(0.069)	(0.022)	
DBORD	0.023	-0.004	-0.017**	2437	0.039	0.015	-0.020	1194
	(0.025)	(0.016)	(0.008)		(0.067)	(0.069)	(0.018)	
DCOAST	-0.141***	-0.112***	-0.032**	2437	-0.264***	-0.333***	-0.106***	1194
	(0.038)	(0.024)	(0.014)		(0.087)	(0.102)	(0.030)	
MALTR	-0.241***	-0.169***	-0.034*	2436	-0.292***	-0.281***	-0.021	1193
	(0.056)	(0.041)	(0.019)		(0.073)	(0.065)	(0.027)	
AGSUIT	0.017	0.037	0.027**	2407	-0.014	-0.031	-0.018	1183
	(0.047)	(0.029)	(0.013)		(0.107)	(0.100)	(0.023)	
ELEV	0.033	-0.009	-0.027**	2428	0.043	0.028	-0.011	1193
	(0.050)	(0.028)	(0.014)		(0.081)	(0.077)	(0.025)	
TERRUG	0.105**	0.081***	0.021	2435	0.159**	0.208***	0.069***	1193
	(0.044)	(0.026)	(0.013)		(0.071)	(0.063)	(0.023)	
OILDUM	0.001	0.008	0.008	2437	0.076*	0.083*	0.015	1194
	(0.019)	(0.014)	(0.007)		(0.045)	(0.046)	(0.011)	
DIADUM	-0.015	-0.014**	-0.006	2437	-0.020	-0.019	-0.001	1194
	(0.013)	(0.007)	(0.006)		(0.013)	(0.012)	(0.007)	
history								
POPD	0.230***	0.166***	0.038***	2435	0.301***	0.317***	0.057***	1193
	(0.051)	(0.034)	(0.013)		(0.053)	(0.040)	(0.021)	
DRRD	-0.273***	-0.220***	-0.040***	1932	-0.398***	-0.412***	-0.064***	1156
	(0.042)	(0.028)	(0.013)		(0.059)	(0.067)	(0.019)	
DROAD	-0.265***	-0.173***	-0.017**	2148	-0.274***	-0.264***	-0.012	1165
	(0.032)	(0.020)	(0.008)		(0.045)	(0.040)	(0.020)	
DCMISS	-0.342***	-0.213***	-0.020	2437	-0.381***	-0.372***	-0.037	1194
	(0.057)	(0.038)	(0.017)		(0.078)	(0.086)	(0.025)	
DPMISS	-0.330***	-0.224***	-0.043***	2437	-0.440***	-0.442***	-0.065***	1194
	(0.042)	(0.027)	(0.012)		(0.069)	(0.071)	(0.022)	
DPCEM	0.026	-0.014	-0.029*	2437	-0.037	-0.035	-0.002	1194
	(0.046)	(0.024)	(0.016)		(0.053)	(0.047)	(0.021)	
DPCST	-0.013	-0.023	-0.016	2437	-0.096	-0.145**	-0.060***	1194
	(0.039)	(0.021)	(0.011)		(0.061)	(0.059)	(0.020)	
contemporary								
URBSHR	0.362***	0.194***	-0.019	2386	0.304***	0.298***	0.026	1170
	(0.022)	(0.013)	(0.014)		(0.025)	(0.028)	(0.027)	
AGSHR	-0.610***	-0.349***	-0.055**	2264	-0.409***	-0.405***	-0.048	1006
	(0.029)	(0.015)	(0.027)		(0.043)	(0.041)	(0.043)	
MANSHR	0.240***	0.125***	-0.002	2264	0.257***	0.238***	0.003	1006
	(0.041)	(0.025)	(0.012)		(0.056)	(0.058)	(0.020)	
SERSHR	0.561***	0.334***	0.074***	2264	0.391***	0.390***	0.047	1006
	(0.032)	(0.016)	(0.024)		(0.045)	(0.038)	(0.045)	
ETHFRAG	0.191***	0.080***	-0.015	1504	0.102***	0.083***	0.004	567
	(0.039)	(0.020)	(0.012)		(0.032)	(0.032)	(0.013)	
ETHPOL	0.080***	0.028*	-0.012	1504	0.049**	0.039	0.000	567
	(0.028)	(0.015)	(0.010)		(0.022)	(0.024)	(0.012)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include country fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.4: District and birth-region-level correlates of the share of literate old and IM, ages 14-100, country fixed effects, with student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.263***	-0.183***	-0.031*	2437	-0.367***	-0.290***	-0.016	1194
	(0.032)	(0.021)	(0.016)		(0.062)	(0.058)	(0.020)	
DBORD	0.023	-0.003	-0.017**	2437	0.039	0.010	-0.020	1194
	(0.025)	(0.016)	(0.008)		(0.067)	(0.056)	(0.015)	
DCOAST	-0.141***	-0.109***	-0.025*	2437	-0.264***	-0.269***	-0.077***	1194
	(0.038)	(0.024)	(0.014)		(0.087)	(0.083)	(0.023)	
MALTR	-0.241***	-0.178***	-0.037*	2436	-0.292***	-0.237***	-0.019	1193
	(0.056)	(0.045)	(0.020)		(0.073)	(0.059)	(0.024)	
AGSUIT	0.017	0.043	0.033**	2407	-0.014	-0.019	-0.008	1183
	(0.047)	(0.033)	(0.015)		(0.107)	(0.085)	(0.021)	
ELEV	0.033	0.001	-0.019	2428	0.043	0.033	0.000	1193
	(0.050)	(0.031)	(0.015)		(0.081)	(0.065)	(0.020)	
TERRUG	0.105**	0.088***	0.026*	2435	0.159**	0.178***	0.060***	1193
	(0.044)	(0.029)	(0.014)		(0.071)	(0.053)	(0.018)	
OILDUM	0.001	0.008	0.007	2437	0.076*	0.067*	0.009	1194
	(0.019)	(0.014)	(0.007)		(0.045)	(0.037)	(0.009)	
DIADUM	-0.015	-0.014**	-0.005	2437	-0.020	-0.014	0.001	1194
	(0.013)	(0.007)	(0.006)		(0.013)	(0.010)	(0.006)	
history								
POPD	0.230***	0.170***	0.036***	2435	0.301***	0.265***	0.047***	1193
	(0.051)	(0.036)	(0.013)		(0.053)	(0.037)	(0.016)	
DRRD	-0.273***	-0.197***	-0.026*	1932	-0.398***	-0.328***	-0.040**	1156
	(0.042)	(0.028)	(0.014)		(0.059)	(0.054)	(0.016)	
DROAD	-0.265***	-0.171***	-0.015*	2148	-0.274***	-0.216***	-0.009	1165
	(0.032)	(0.021)	(0.009)		(0.045)	(0.035)	(0.016)	
DCMISS	-0.342***	-0.210***	-0.007	2437	-0.381***	-0.306***	-0.022	1194
	(0.057)	(0.042)	(0.019)		(0.078)	(0.072)	(0.022)	
DPMISS	-0.330***	-0.230***	-0.040***	2437	-0.440***	-0.370***	-0.053***	1194
	(0.042)	(0.030)	(0.013)		(0.069)	(0.060)	(0.019)	
DPCEM	0.026	-0.016	-0.032*	2437	-0.037	-0.038	-0.011	1194
	(0.046)	(0.026)	(0.017)		(0.053)	(0.043)	(0.018)	
DPCST	-0.013	-0.031	-0.023*	2437	-0.096	-0.126***	-0.054***	1194
	(0.039)	(0.022)	(0.013)		(0.061)	(0.049)	(0.016)	
contemporary								
URBSHR	0.362***	0.191***	-0.036**	2386	0.304***	0.231***	0.004	1170
	(0.022)	(0.014)	(0.014)		(0.025)	(0.023)	(0.023)	
AGSHR	-0.610***	-0.369***	-0.040	2264	-0.409***	-0.331***	-0.031	1006
	(0.029)	(0.016)	(0.031)		(0.043)	(0.035)	(0.041)	
MANSHR	0.240***	0.129***	-0.009	2264	0.257***	0.192***	-0.004	1006
	(0.041)	(0.027)	(0.013)		(0.056)	(0.047)	(0.018)	
SERSHR	0.561***	0.354***	0.065**	2264	0.391***	0.320***	0.034	1006
	(0.032)	(0.017)	(0.028)		(0.045)	(0.033)	(0.044)	
ETHFRAG	0.191***	0.082***	-0.026**	1504	0.102***	0.060**	-0.005	567
	(0.039)	(0.024)	(0.013)		(0.032)	(0.025)	(0.011)	
ETHPOL	0.080***	0.025	-0.019	1504	0.049**	0.026	-0.005	567
	(0.028)	(0.018)	(0.012)		(0.022)	(0.020)	(0.010)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include country fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

E.2 Province fixed effects

Table E.5: District and birth-region-level correlates of the share of literate old and IM, ages 14-18, province fixed effects, without student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.373***	-0.279***	-0.095***	2379	-0.353***	-0.218***	-0.026	1151
	(0.108)	(0.045)	(0.037)		(0.124)	(0.071)	(0.027)	
DBORD	0.000	-0.015	-0.015	2379	0.003	0.012	0.010	1151
	(0.025)	(0.018)	(0.011)		(0.060)	(0.036)	(0.018)	
DCOAST	-0.181***	-0.143***	-0.051***	2379	-0.311***	-0.192***	-0.023	1151
	(0.042)	(0.034)	(0.019)		(0.043)	(0.041)	(0.032)	
MALTR	-0.284***	-0.172**	-0.029	2378	-0.426***	-0.262***	-0.035	1150
	(0.103)	(0.070)	(0.025)		(0.144)	(0.085)	(0.027)	
AGSUIT	0.118	0.108**	0.049**	2349	0.147	0.109	0.029	1140
	(0.077)	(0.050)	(0.023)		(0.124)	(0.074)	(0.045)	
ELEV	0.068	-0.012	-0.047*	2370	0.014	0.046	0.039	1150
	(0.077)	(0.043)	(0.024)		(0.132)	(0.077)	(0.033)	
TERRUG	0.069*	0.055**	0.020	2377	0.040	0.046	0.024	1150
	(0.041)	(0.026)	(0.020)		(0.064)	(0.038)	(0.020)	
OILDUM	-0.046**	-0.042***	-0.018	2379	0.016	-0.025	-0.034	1151
	(0.021)	(0.015)	(0.014)		(0.040)	(0.021)	(0.026)	
DIADUM	-0.005	-0.004	-0.001	2379	-0.004	0.000	0.002	1151
	(0.011)	(0.008)	(0.007)		(0.017)	(0.016)	(0.011)	
history								
POPD	0.265***	0.203***	0.073**	2377	0.274**	0.215***	0.071***	1150
	(0.076)	(0.052)	(0.029)		(0.138)	(0.063)	(0.027)	
DRRD	-0.329***	-0.217***	-0.050***	1874	-0.375***	-0.247***	-0.052**	1113
	(0.061)	(0.037)	(0.018)		(0.089)	(0.049)	(0.022)	
DROAD	-0.253***	-0.151***	-0.028***	2090	-0.240***	-0.167***	-0.041**	1122
	(0.035)	(0.020)	(0.010)		(0.063)	(0.037)	(0.017)	
DCMISS	-0.393***	-0.237***	-0.039	2379	-0.445***	-0.248***	-0.004	1151
	(0.065)	(0.037)	(0.024)		(0.102)	(0.050)	(0.030)	
DPMISS	-0.337***	-0.197***	-0.028*	2379	-0.389***	-0.224***	-0.013	1151
	(0.059)	(0.034)	(0.015)		(0.129)	(0.069)	(0.022)	
DPCEM	0.059	0.058	0.028	2379	-0.011	0.036	0.042	1151
	(0.049)	(0.035)	(0.036)		(0.045)	(0.040)	(0.032)	
DPCST	-0.039	-0.025	-0.005	2379	-0.117*	-0.070	-0.006	1151
	(0.041)	(0.032)	(0.014)		(0.066)	(0.056)	(0.029)	
contemporary								
URBSHR	0.358***	0.166***	-0.022	2328	0.296***	0.194***	0.038	1148
	(0.029)	(0.017)	(0.015)		(0.029)	(0.024)	(0.024)	
AGSHR	-0.663***	-0.369***	-0.107***	2206	-0.389***	-0.238***	-0.061	963
	(0.036)	(0.019)	(0.031)		(0.057)	(0.045)	(0.041)	
MANSHR	0.232***	0.110***	-0.002	2206	0.278***	0.136***	-0.006	963
	(0.067)	(0.031)	(0.018)		(0.052)	(0.034)	(0.028)	
SERSHR	0.591***	0.341***	0.117***	2206	0.369***	0.232***	0.064	963
	(0.039)	(0.019)	(0.025)		(0.062)	(0.045)	(0.041)	
ETHFRAG	0.184***	0.038	-0.052***	1483	0.134***	0.083**	0.022	536
	(0.055)	(0.031)	(0.016)		(0.047)	(0.034)	(0.020)	
ETHPOL	0.097***	0.012	-0.034**	1483	0.092***	0.060**	0.017	536
	(0.037)	(0.019)	(0.015)		(0.035)	(0.026)	(0.016)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include province fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.6: District and birth-region-level correlates of the share of literate old and IM, ages 14-18, province fixed effects, with student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.373***	-0.254***	-0.086***	2379	-0.353***	-0.199**	-0.016	1151
	(0.108)	(0.047)	(0.033)		(0.124)	(0.080)	(0.029)	
DBORD	0.000	-0.011	-0.011	2379	0.003	0.016	0.014	1151
	(0.025)	(0.016)	(0.010)		(0.060)	(0.031)	(0.020)	
DCOAST	-0.181***	-0.126***	-0.043***	2379	-0.311***	-0.168***	-0.005	1151
	(0.042)	(0.031)	(0.016)		(0.043)	(0.044)	(0.037)	
MALTR	-0.284***	-0.171**	-0.042	2378	-0.426***	-0.253**	-0.039	1150
	(0.103)	(0.080)	(0.035)		(0.144)	(0.107)	(0.038)	
AGSUIT	0.118	0.109**	0.056**	2349	0.147	0.125	0.050	1140
	(0.077)	(0.055)	(0.027)		(0.124)	(0.082)	(0.050)	
ELEV	0.068	-0.005	-0.037	2370	0.014	0.066	0.059	1150
	(0.077)	(0.044)	(0.025)		(0.132)	(0.082)	(0.042)	
TERRUG	0.069*	0.045*	0.014	2377	0.040	0.047	0.027	1150
	(0.041)	(0.026)	(0.019)		(0.064)	(0.039)	(0.023)	
OILDUM	-0.046**	-0.035***	-0.014	2379	0.016	-0.026	-0.034	1151
	(0.021)	(0.012)	(0.012)		(0.040)	(0.021)	(0.027)	
DIADUM	-0.005	-0.004	-0.001	2379	-0.004	0.000	0.002	1151
	(0.011)	(0.007)	(0.007)		(0.017)	(0.016)	(0.011)	
history								
POPD	0.265***	0.187***	0.068***	2377	0.274**	0.196***	0.059*	1150
	(0.076)	(0.048)	(0.024)		(0.138)	(0.066)	(0.035)	
DRRD	-0.329***	-0.201***	-0.048***	1874	-0.375***	-0.227***	-0.042**	1113
	(0.061)	(0.040)	(0.017)		(0.089)	(0.056)	(0.021)	
DROAD	-0.253***	-0.133***	-0.022**	2090	-0.240***	-0.149***	-0.028	1122
	(0.035)	(0.021)	(0.010)		(0.063)	(0.040)	(0.018)	
DCMISS	-0.393***	-0.202***	-0.021	2379	-0.445***	-0.224***	0.010	1151
	(0.065)	(0.036)	(0.024)		(0.102)	(0.053)	(0.034)	
DPMISS	-0.337***	-0.177***	-0.023	2379	-0.389***	-0.224***	-0.024	1151
	(0.059)	(0.039)	(0.015)		(0.129)	(0.083)	(0.024)	
DPCEM	0.059	0.042	0.015	2379	-0.011	0.025	0.031	1151
	(0.049)	(0.037)	(0.037)		(0.045)	(0.046)	(0.034)	
DPCST	-0.039	-0.031	-0.013	2379	-0.117*	-0.067	-0.006	1151
	(0.041)	(0.029)	(0.016)		(0.066)	(0.060)	(0.032)	
contemporary								
URBSHR	0.358***	0.143***	-0.033*	2328	0.296***	0.154***	-0.002	1148
	(0.029)	(0.017)	(0.017)		(0.029)	(0.027)	(0.029)	
AGSHR	-0.663***	-0.323***	-0.088***	2206	-0.389***	-0.193***	-0.027	963
	(0.036)	(0.020)	(0.029)		(0.057)	(0.053)	(0.052)	
MANSHR	0.232***	0.091***	-0.009	2206	0.278***	0.104***	-0.025	963
	(0.067)	(0.028)	(0.017)		(0.052)	(0.035)	(0.032)	
SERSHR	0.591***	0.297***	0.095***	2206	0.369***	0.189***	0.034	963
	(0.039)	(0.020)	(0.024)		(0.062)	(0.054)	(0.055)	
ETHFRAG	0.184***	0.032	-0.047***	1483	0.134***	0.061**	0.007	536
	(0.055)	(0.030)	(0.015)		(0.047)	(0.030)	(0.018)	
ETHPOL	0.097***	0.005	-0.035**	1483	0.092***	0.043*	0.006	536
	(0.037)	(0.018)	(0.014)		(0.035)	(0.024)	(0.015)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include province fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.7: District and birth-region-level correlates of the share of literate old and IM, ages 14-100, province fixed effects, without student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.320***	-0.230***	-0.065***	2379	-0.286***	-0.248***	-0.016	1151
	(0.094)	(0.042)	(0.022)		(0.091)	(0.078)	(0.029)	
DBORD	0.002	-0.010	-0.011	2379	0.000	-0.006	-0.006	1151
	(0.021)	(0.015)	(0.008)		(0.047)	(0.040)	(0.018)	
DCOAST	-0.147***	-0.122***	-0.045***	2379	-0.235***	-0.238***	-0.049	1151
	(0.033)	(0.028)	(0.016)		(0.034)	(0.043)	(0.035)	
MALTR	-0.224***	-0.150**	-0.034	2378	-0.316***	-0.303***	-0.054*	1150
	(0.083)	(0.063)	(0.024)		(0.106)	(0.102)	(0.031)	
AGSUIT	0.096	0.080*	0.030*	2349	0.108	0.129	0.042	1140
	(0.062)	(0.043)	(0.018)		(0.093)	(0.080)	(0.040)	
ELEV	0.066	-0.008	-0.043**	2370	0.005	0.014	0.010	1150
	(0.064)	(0.038)	(0.019)		(0.101)	(0.090)	(0.036)	
TERRUG	0.060*	0.042**	0.010	2377	0.022	0.032	0.015	1150
	(0.034)	(0.021)	(0.014)		(0.049)	(0.042)	(0.021)	
OILDUM	-0.035**	-0.032***	-0.013	2379	0.015	-0.010	-0.022	1151
	(0.017)	(0.010)	(0.010)		(0.034)	(0.023)	(0.019)	
DIADUM	-0.005	-0.003	-0.001	2379	-0.007	-0.005	0.000	1151
	(0.009)	(0.006)	(0.005)		(0.012)	(0.015)	(0.009)	
history								
POPD	0.225***	0.167***	0.053**	2377	0.229**	0.236***	0.056**	1150
	(0.063)	(0.044)	(0.020)		(0.105)	(0.079)	(0.026)	
DRRD	-0.258***	-0.211***	-0.053***	1874	-0.285***	-0.286***	-0.054***	1113
	(0.046)	(0.034)	(0.014)		(0.064)	(0.061)	(0.019)	
DROAD	-0.210***	-0.138***	-0.024***	2090	-0.183***	-0.186***	-0.036**	1122
	(0.029)	(0.017)	(0.006)		(0.046)	(0.043)	(0.017)	
DCMISS	-0.333***	-0.215***	-0.042**	2379	-0.343***	-0.297***	-0.019	1151
	(0.053)	(0.030)	(0.017)		(0.076)	(0.054)	(0.031)	
DPMISS	-0.287***	-0.168***	-0.019	2379	-0.298***	-0.264***	-0.024	1151
	(0.047)	(0.032)	(0.012)		(0.095)	(0.086)	(0.024)	
DPCEM	0.050	0.034	0.008	2379	-0.010	0.018	0.025	1151
	(0.042)	(0.027)	(0.024)		(0.030)	(0.042)	(0.029)	
DPCST	-0.033	-0.020	-0.002	2379	-0.087*	-0.082	-0.011	1151
	(0.032)	(0.027)	(0.012)		(0.046)	(0.058)	(0.027)	
contemporary								
URBSHR	0.305***	0.161***	-0.000	2328	0.238***	0.211***	0.021	1148
	(0.026)	(0.013)	(0.014)		(0.020)	(0.024)	(0.025)	
AGSHR	-0.593***	-0.325***	-0.095***	2206	-0.311***	-0.270***	-0.045	963
	(0.036)	(0.015)	(0.023)		(0.034)	(0.044)	(0.048)	
MANSHR	0.207***	0.102***	0.004	2206	0.225***	0.178***	0.003	963
	(0.062)	(0.028)	(0.013)		(0.040)	(0.033)	(0.028)	
SERSHR	0.528***	0.299***	0.102***	2206	0.296***	0.262***	0.049	963
	(0.039)	(0.016)	(0.019)		(0.040)	(0.045)	(0.048)	
ETHFRAG	0.173***	0.058**	-0.024**	1483	0.101***	0.105***	0.020	536
	(0.052)	(0.026)	(0.011)		(0.034)	(0.039)	(0.018)	
ETHPOL	0.089**	0.026	-0.016	1483	0.067***	0.074**	0.018	536
	(0.035)	(0.016)	(0.010)		(0.025)	(0.031)	(0.016)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include province fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

Table E.8: District and birth-region-level correlates of the share of literate old and IM, ages 14-100, province fixed effects, with student correction

	observational estimates				within-migrant-household estimates			
variable	share literate old	IM	IM controlling for share literate old	N	share literate old	IM	IM controlling for share literate old	N
geography								
DCAP	-0.320***	-0.237***	-0.069***	2379	-0.286***	-0.205***	-0.013	1151
	(0.094)	(0.046)	(0.023)		(0.091)	(0.072)	(0.029)	
DBORD	0.002	-0.009	-0.010	2379	0.000	-0.004	-0.004	1151
	(0.021)	(0.015)	(0.008)		(0.047)	(0.031)	(0.016)	
DCOAST	-0.147***	-0.123***	-0.044***	2379	-0.235***	-0.192***	-0.036	1151
	(0.033)	(0.029)	(0.017)		(0.034)	(0.037)	(0.032)	
MALTR	-0.224***	-0.161**	-0.043	2378	-0.316***	-0.256***	-0.054*	1150
	(0.083)	(0.073)	(0.031)		(0.106)	(0.095)	(0.033)	
AGSUIT	0.096	0.090*	0.038*	2349	0.108	0.118	0.047	1140
	(0.062)	(0.049)	(0.022)		(0.093)	(0.072)	(0.037)	
ELEV	0.066	-0.004	-0.039*	2370	0.005	0.022	0.019	1150
	(0.064)	(0.042)	(0.021)		(0.101)	(0.078)	(0.033)	
TERRUG	0.060*	0.043*	0.011	2377	0.022	0.030	0.015	1150
	(0.034)	(0.022)	(0.015)		(0.049)	(0.037)	(0.019)	
OILDUM	-0.035**	-0.032***	-0.013	2379	0.015	-0.008	-0.018	1151
	(0.017)	(0.010)	(0.010)		(0.034)	(0.018)	(0.016)	
DIADUM	-0.005	-0.004	-0.001	2379	-0.007	-0.004	0.000	1151
	(0.009)	(0.006)	(0.005)		(0.012)	(0.013)	(0.008)	
history								
POPD	0.225***	0.174***	0.057***	2377	0.229**	0.193***	0.045*	1150
	(0.063)	(0.045)	(0.020)		(0.105)	(0.067)	(0.026)	
DRRD	-0.258***	-0.192***	-0.048***	1874	-0.285***	-0.221***	-0.037**	1113
	(0.046)	(0.034)	(0.014)		(0.064)	(0.053)	(0.015)	
DROAD	-0.210***	-0.134***	-0.022***	2090	-0.183***	-0.145***	-0.025*	1122
	(0.029)	(0.018)	(0.007)		(0.046)	(0.037)	(0.015)	
DCMISS	-0.333***	-0.215***	-0.037**	2379	-0.343***	-0.244***	-0.015	1151
	(0.053)	(0.033)	(0.019)		(0.076)	(0.048)	(0.028)	
DPMISS	-0.287***	-0.171***	-0.018	2379	-0.298***	-0.224***	-0.027	1151
	(0.047)	(0.036)	(0.014)		(0.095)	(0.078)	(0.022)	
DPCEM	0.050	0.029	0.002	2379	-0.010	0.010	0.017	1151
	(0.042)	(0.029)	(0.027)		(0.030)	(0.038)	(0.027)	
DPCST	-0.033	-0.024	-0.006	2379	-0.087*	-0.067	-0.009	1151
	(0.032)	(0.027)	(0.013)		(0.046)	(0.051)	(0.025)	
contemporary								
URBSHR	0.305***	0.159***	-0.008	2328	0.238***	0.163***	0.003	1148
	(0.026)	(0.014)	(0.016)		(0.020)	(0.021)	(0.025)	
AGSHR	-0.593***	-0.343***	-0.099***	2206	-0.311***	-0.212***	-0.032	963
	(0.036)	(0.017)	(0.025)		(0.034)	(0.040)	(0.047)	
MANSHR	0.207***	0.106***	0.003	2206	0.225***	0.137***	-0.002	963
	(0.062)	(0.030)	(0.014)		(0.040)	(0.027)	(0.024)	
SERSHR	0.528***	0.315***	0.105***	2206	0.296***	0.207***	0.037	963
	(0.039)	(0.018)	(0.020)		(0.040)	(0.041)	(0.048)	
ETHFRAG	0.173***	0.061**	-0.029**	1483	0.101***	0.072**	0.010	536
	(0.052)	(0.029)	(0.012)		(0.034)	(0.029)	(0.014)	
ETHPOL	0.089**	0.024	-0.022**	1483	0.067***	0.051**	0.010	536
	(0.035)	(0.018)	(0.011)		(0.025)	(0.023)	(0.012)	

This is not a normal regression table. In the columns entitled "share literate old" the dependent variable is the district / birth-region share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled "IM" it is the district / birth-region-level share of children of parents with less than primary who complete at least primary (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled "IM controlling for share literate old". Each row shows the results of regressions of these variabes on the LHS on one RHS variable (indicated in the rows) at a time. The regressions in the two columns "IM controlling for share literate old" additionally control for the share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old), that is they include the LHS variable of the columns "share literate old" on the RHS. All specifications include province fixed effects (not reported). Coefficients are standardized. DCAP = distance to the capital, DBORD = distance to the national border, DCOAST = distance to the coast, MALTR = stability of malaria transmission, AGSUIT = agricultural suitability, ELEV = elevation, TERRUG = terrain ruggedness, OILDUM = oil field dummy, DIADUM = diamond mine dummy, POPD = mean population density in 1950, DRRD = distance to closest colonial railroad, DROAD = distance to closest colonial road, DCMISS = distance to closest catholic mission, DPMISS = distance to closest protestant mission, DPCEM = distance to closest pre-colonial empire, DPCST = distance to closest precolonial state, URBSHR = urban share among individuals born 1960 and earlier, MANSHR = manufacturing labour share among individuals born 1960 and earlier, SERSHR = services labour share among individuals born 1960 and earlier, ETHFRAG = ethnic fragmentation among individuals born 1960 and earlier, ETHPOL = ethnic polarization among individuals born 1960 and earlier. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.