Intergenerational Mobility in Africa*

Alberto Alesina[†]
Harvard University, IGIER, CEPR and NBER

Sebastian Hohmann[‡]
Stockholm SSE SITE

Stelios Michalopoulos[§]
Brown University, CEPR and NBER

Elias Papaioannou[¶]
London Business School and CEPR

This draft: January 2019

Abstract

We examine intergenerational mobility (IM) in educational attainment in Africa since independence, using census data from 26 countries. First, we map and characterize the geography of IM. There is substantial variation both across and within countries with differences in literacy of the old generation being the strongest correlate of IM. Inertia is stronger for rural, as compared to urban, households and present for both boys and girls. Second, we explore the correlates of mobility across more than 2,800 regions. Colonial investments in the transportation network and missionary activity are associated with upward mobility. IM is also higher in regions close to the coast and national capitals as well as in rugged areas without malaria. Upward mobility is higher and downward mobility is lower in regions that were more developed at independence, with higher urbanization and employment in services and manufacturing. Third, we identify the effects of regions on educational mobility by exploiting within-family variation from children whose families moved during primay school age. While sorting is sizable, there are considerable regional exposure effects.

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

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

*We are in particular grateful to our discussants Oriana Bandiera, Sara Lowes, and Nathaniel Hendren as well as to Francesco Caselli and Antonio Ciccone for detailed comments. We are thankful for useful discussions and feedback to James Fenske, Larry Katz, Nathan Nunn, David Weil, Elisa Cavatorta, Shaun Hargreaves, and Jim Robinson. We also got useful feedback from conference and seminar participants at the University of Zurich, UPF, Monash University, Warwick, Brown, Harvard, King's College, Copenhagen, Sheffield, Williams, the NBER Summer Institute, and the PSE Conference on Culture, Institutions, and Economic Prosperity. We also thank Remi Jedwab, Adam Storeygard, Julia Cagé, Valeria Rueda and Nathan Nunn for generously sharing their data. All errors and omissions are our responsibility.

[†]aalesina@harvard.edu
[‡]sebastian.hohmann@hhs.se
[§]smichalo@brown.edu
[¶]eliaspapaioannou@london.edu

1 Introduction

There is rising optimism about Africa's future, a continent with 1.2 billion opportunities, as the Economist (2016) recently touted. The formerly "hopeless continent" is gradually becoming the "hopeful" one (Economist (2000, 2011)). Educational attainment is rising, health is improving, and the income of many Africans is growing. Some speak about an African "growth miracle" (Young (2012)). However, anecdotal evidence indicates widespread inequalities in income, education, and social mobility. This suggests that the aggregate gains may not be broadly shared, but a comprehensive assessment is lacking.

In this paper, we take the first step toward mapping, exploring and explaining intergenerational mobility across the continent since independence. We look at educational attainment using census data covering more than 14 million individuals across 26 African countries and $2,800$ regions. Reconstructing the joint distribution of parental and offspring educational outcomes since the 1960s, when most of Africa becomes independent, allows us to shed light on many questions. Where is the land of educational opportunity in Africa? Are differences in intergenerational mobility across countries and regions small, moderate or wide? How large are gender disparities? How big is the rural-urban gap? Which elements of a region's history and geography correlate with educational mobility? Do regions matter for social mobility or do regions with higher mobility just attract families more eager to climb the social ladder?

Results Preview In the first part of the paper, we present new country and region level measures of educational opportunity in Africa. Following recent work on intergenerational mobility in income (Chetty et al. (2014); Chetty and Hendren (2018a), Chetty and Hendren (2018b)) and education (Card et al. (2018); Fletcher and Han (2018)) we construct measures of upward intergenerational mobility (IM) defined as the likelihood that children born to parents that have not completed primary schooling manage to do so. Similarly, we map downward mobility, defined as the probability that offspring of parents with completed primary education fail to do so. We use data from 14,149,328 children residing with at least one parent at the time of the census. To account for "selection on cohabitation", we follow Card et al. (2018) and focus on children between 14 and 18 years. Children in that age bracket have largely finished primary school and at the same time they still reside overwhelmingly with their parents (cohabitation rates are approximately 94%).

We document large cross-country heterogeneity in upward and downward mobility rates. The likelihood that children born to parents with no education complete primary schooling exceeds 70% in South Africa and Botswana; the corresponding statistic in Sudan, Ethiopia, Mozambique, Burkina Faso, and Malawi hovers below 20%. The analysis also uncovers substantial within-country differences in IM. For example, in Kenya, a country with a close-to-average IM of 0.50, the likelihood that children of illiterate parents will complete primary education ranges from just 5% (in Turkana county in the Northwest of the country) to more than 85% (in Westlands, an administrative division and affluent neighborhood in Nairobi). Upward IM is higher (lower) in regions and countries with

relatively higher (lower) levels of parental literacy. Variation in literacy rates among the old generation accounts for half of the observed regional variability in intergenerational mobility. Downward mobility is also linked to the stock of literacy, though the association is less strong. These findings imply considerable persistence of initial conditions. Inertia is stronger for rural Africa. While there is a gender gap in educational levels, intergenerational mobility is similar for boys and girls.

In the second part, we characterize the geography of the *land of opportunity* in Africa. Upward IM is higher and downward IM is lower in malaria-free regions. Distance to the capital and the coast correlate negatively with upward IM; this most likely reflects African states' weak capacity to broadcast power far from the main urban hubs and limited public investment in the countryside. Colonial investments in railroads and roads as well as missionary presence correlate positively with upward IM and negatively with downward IM. Though these correlations are robust to controlling for the parental stock of literacy, they do not identify causal effects. However, they are suggestive of how historical contingencies related to colonization and geographic attributes may influence not only initial economic conditions, but also the trajectories of local economies.

The observed differences in regional IM may result from two quite different forces. On the one hand, regions may have a causal impact on mobility, by providing higher quality infrastructure, more schools, and better occupational opportunities. On the other hand, regional disparities could reflect sorting, as families with higher ability and/or valuation of education move to areas with better opportunities.

In the third part, we assess the relative magnitudes of these two factors. As a starting point, we estimate within-household specifications looking at the effect of place-of-birth-IM on the probability that children born to illiterate parents will complete primary education. By comparing siblings born and partly raised in regions with different IM, we account for family characteristics. The within-family analysis reveals that while sorting is sizable, the district a child grows up in matters crucially for whether she will complete primary education. We then employ the neat approach of Chetty and Hendren (2018a) to isolate the one-way effect of regions on educational mobility. The methodology exploits differences in the age at which children of migrant households move across districts to distinguish "*selection*" from "*regional exposure effects*". Both forces are at play. Selection is far-from-negligible; families' sorting into better (worse) locations correlates strongly with children's educational attainment. This result adds to the recent findings of Young (2013), who uses survey data to document two-way rural-urban migration in developing countries based on differences in human capital. The analysis also uncovers sizable "*regional exposure effects*" both for boys and girls. A child who moves with her uneducated parents to a region with a one-standard-deviation higher IM than her birthplace at the age of 6, has a 7 percentage points higher likelihood of completing primary schooling, compared to her sibling who at the time of the move was already 11 years old.

Related Literature Our work contributes to and blends two strands of literature that have, thus far, moved in parallel. The first is the growing research that studies intergenerational mobility. 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 of children to parental outcomes; as such, most earlier works rely on relatively small samples from surveys. Card et al. (2018) 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). They show rising mobility during the first half of the 20th century, which differs across race and states.[2] Chetty et al. (2014) provide a mapping of IM in income across US counties and explore its correlates. Chetty and Hendren (2018a) use matched parents-children administrative tax records of moving families to isolate the effect of neighborhood exposure on mobility from sorting. Our work is similar in spirit to the study of Asher et al. (2018), on educational mobility across Indian regions and to the parallel work of the World Bank trying to construct measures of intergenerational mobility in education and income across many countries using survey data (Narayan et al. (2018)). Finally, our paper relates to Becker et al. (2018) who develop a theory of intergenerational mobility where parental investments in education result in persistent differences in economic status even in the absence of capital market imperfections and differences in innate ability.

The second strand is the research on African development (Young (2012), Pinkovskiy and Sala-i-Martin (2014)). The literature has moved from mostly cross-country approaches focusing on national features (e.g., Gunning and Collier (1999), Bates (2015)), to within-country analyses that connect Africa's contemporary development to its colonial and pre-colonial past. This research provides compelling evidence of historical continuity as well as instances of rupture in the evolution of the economy and polity (Michalopoulos and Papaioannou (2019) provide a review). Nevertheless, this literature has not opened the "black box" of intergenerational linkages. A natural question is whether the correlation between deeply rooted factors and current outcomes reflects the one-time effect of the former on initial (at-independence) conditions or if these identified historical legacies have also changed the transmission of opportunity across generations. By building granular data on IM across African regions and exploring in a systematic manner its correlates we can begin answering such questions.[3]

Structure In Section 2, we present the census data on educational attainment and detail the construction of the intergenerational educational mobility measures. Section

[1] Olivetti and Paserman (2015) and Olivetti et al. (2018) study IM in income in the United States from 1850 till 1940. Charles and Hurst (2004) use PSID data to estimate intergenerational persistence in wealth across US households. Alesina et al. (2018) compare actual social mobility and perceptions across several industrial countries. Early studies on intergenerational mobility in education include Bowles (1972), Blake (1985), and Spady (1967). Hertz et al. (2008) estimate country-level IM coefficients for various cohorts across 42 countries. Hilger (2017) calculates educational IM in the United States since WWII. Azam and Bhatt (2015) and Golley and Kong (2013) estimate mobility in education in India and China, respectively. Long and Ferrie (2013) describe the dynamics of occupational intergenerational mobility in the US and the UK since mid-19th century. They document higher levels of upward mobility in the early decades in the US, which however vanish in the Inter-war period.

[2] A strand of the mostly US-centered literature looks at racial differences in intergenerational mobility (e.g., Chetty et al. (2018), Davis and Mazumder (2018), Derenoncourt (2018)). These studies relate to our companion work (Alesina et al. (2019)), where we map ethnic and religious differences in social mobility across Africa.

[3] In an innovative case study, Wantchekon et al. (2015) study the intergenerational impact of colonial schools in Benin. They show that colonial schools not only raised income, education, and well-being of students and their communities, but that the benefits spread to the second and third generation.

3 describes IM across African countries and regions. Section 4 presents the results from the exploration of the geographic, historical and at-independence correlates of educational mobility. In Section 5, we first explore within-household variation on children's place of birth to isolate selection (migration) from regional features affecting intergenerational mobility. We then exploit differences in ages-at-move among migrants to estimate regional exposure effects on educational mobility. In Section 6, we summarize and discuss avenues for future research.

2 Data and methods

2.1 Why Education?

We focus on education for several reasons. First, income and wealth data are available for a tiny share of the African population and only for a handful of countries. For instance, Alvaredo et al. (2017) report that for Ghana, Kenya, Tanzania, Nigeria, and Uganda, income data encompass less than 1% of the adult population, while for most African countries tax records do not exist, as the share of the underground economy is substantial (Porta and Shleifer (2008), La Porta and Shleifer (2014)) and fiscal capacity very weak (Besley and Persson (2013)). Moreover, consumption data are noisy, cover small samples, and are not spatially disaggregated. In contrast, education data are available since the late 1960s and have a fine temporal and geographic resolution. Second, measurement error in educational attainment is a lesser concern compared to that of reported income, wealth or consumption. Education is also useful in mapping intergenerational mobility, as people tend to complete primary schooling, which is the key educational achievement for most of Africa, by the age of 14. Hence, unlike lifetime earnings or wealth, the analysis can start when adults are relatively early in the life-cycle. Third, education is strongly correlated with income/wealth across countries (e.g., Barro and Lee (2013) and regions (Gennaioli et al. (2014)); a large body of research in labor economics shows that education causally affects lifetime income (Card (1999), Krueger and Lindahl (2001)). Individual (Mincerian) returns to schooling are sizable and possibly larger in low-income countries.[4] Fourth, as we show in Appendix D with data from the Demographic and Health Surveys (DHS) and the Afrobarometer Surveys, years of schooling correlate positively with various proxies of well-being; living conditions, child mortality and fertility, attitudes toward domestic violence, and proxies of political and civic engagement.

2.2 Sample

Our analysis is based upon individual records, retrieved from 68 national censuses from 26 countries: Benin, Botswana, Burkina Faso, Cameroon, Egypt, Ethiopia, Ghana, Guinea, Kenya, Lesotho, Liberia, Malawi, Mali, Morocco, Mozambique, Nigeria, Rwanda, Sene-

[4] Young (2012) 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 et al. (2014) report lower returns in Sub-Saharan Africa of 8.5%. In line with the earlier work of Psacharopoulos (1994) they also estimate a negative relationship between Mincerian returns and years of schooling (which is steeper in 1995 as compared to 2005). Montenegro and Patrinos (2014) estimate higher Mincerian returns in SSA (12.5%) compared to the rest of the world (9.7%).

gal, Sierra Leone, South Africa, Sudan, South Sudan, Tanzania, Uganda, Zambia, and Zimbabwe. We retrieve the data from IPUMS (Integrated Public Use Microdata Series) International, hosted at the University of Minnesota Population Centre. IPUMS reports harmonized, representative samples, typically 10%.[5] As of 2015, the countries in our sample were home to slightly more than 850 million people, representing around 75 percent of Africa's population and GDP.

Overall, IPUMS records education for around 93 million individuals. We drop those younger than 14 years, so as to allow children to complete primary schooling (changing this cutoff to 12 or 16 does not change the results). This leaves us with around 66 million observations. To assign children to their parents (and estimate IM), we use information for individuals of who co-habitate with an older generation. This brings the sample down to 20.3 million observations. For households with three or more generations, an individual's education could appear both as the education of an "old" generation (vis a vis one's children) and as the education of a "young" generation (vis a vis one's parents). For simplicity, we drop such households, further reducing the sample to 14,149,328 individuals.[6] Appendix Table A.1 gives details on sample construction: census years, coverage rates, number of individuals.

The final dataset includes information on 14,149,328 "young" individuals, older than 14, who cohabitate with at least one member of an "older" generation. Estimating IM of individuals who reside with at least one older person (normally a parent) raises "cohabitation selection" issues. Following Card et al. (2018), we restrict the maximum age of "children" in the sample to either 18 or 25 years (see also Hilger (2017)). Thus we estimate IM on 12.1 million and 7.3 million individuals for the samples in the age brackets 14-25 and 14-18, respectively. IPUMS also reports information on respondents' current residence, allowing us to assign individuals to "coarse" and "fine" current administrative units. Districts are typically admin-2 and in some countries 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).[7] Our sample covers 365 provinces and $2,813$ districts across the 26 countries. Appendix table A.2 provides details of the three samples we work with: 14-18, 14-25, and 14+.

For 23 countries, hosting 10.3 million "young" aged 14-25, IPUMS also records place of birth, which allows us to assess migrant status. For a subset of 7.8 million individuals from 15 countries, we additionally have information on the timing of move, if any. Appendix tables A.3 and A.4 again provide details of the three samples: 14-18, 14-25, and 14+.

2.3 Methodology

We construct measures of *absolute* IM that reflect the likelihood that children acquire higher/lower/similar educational attainment than individuals in the same household be-

[5] One exception is Nigeria; data come from household surveys conducted in consecutive years between 2006 and 2010. We aggregate the yearly waves and count them as one census-year.

[6] In an earlier draft, we had included multigenerational households. The results are similar.

[7] For Botswana, Lesotho, and Nigeria, IPUMS reports one administrative unit, "Districts" in Botswana and Lesotho, "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, as administrative boundaries are sometimes redrawn. For our analysis, we have harmonized administrative boundaries.

longing to the immediately previous generation who cohabitate with them (parents and/or extended family members, such as aunts and uncles[8]).

There are four main educational attainment categories: (*i*) no schooling and less than completed primary; (*ii*) completed primary (and some secondary); (*iii*) completed secondary (and some tertiary); and (*iv*) completed tertiary and higher. Individuals with incomplete secondary education are assigned to completed primary, individuals with incomplete tertiary education are assigned to completed secondary. For the education of the "old" we take the average attainment of individuals one generation older in the household, rounded to the nearest integer. Our results are almost identical if we use the minimum or maximum.

For an initial look at the data, we construct 4×4 transition matrices covering all four broad educational categories. We impose two criteria to ensure that individuals have truly completed their schooling and are not misclassified: we require individuals to be at least (a) 18 years old, and (b) 9 years older than their years of schooling.[9] This is the only place in the paper where we impose this restriction. Figure 1 (a) shows the Africa-wide transition matrix using all censuses, while figures 1 (b) and (c) reproduce the transition matrices for Mozambique and Tanzania, respectively. The height of each cell (vertical axis) indicates the probability that the child has the respective educational attainment, conditional on his/her parents having the educational attainment depicted in the horizontal axis. The bars' width indicates the percentage of parents with each of the four main educational attainment categories. Across Africa (pooling across all country-censuses) roughly 75% of the "old" generation has not completed primary schooling and only 1.2% of the "old" generation has completed tertiary education. 26% of African children whose parents have not completed primary schooling, manage to do so; 12% finish high-school and 2% even get a college degree.

Since three-fourths of "old" Africans have not completed primary school, we focus on the likelihood that children born to parents without any schooling or less than completed primary (that for simplicity we label "illiterate") manage to complete primary education (we label them "literate"). This is our proxy of upward educational (social) mobility. Our measure of downward mobility is the likelihood that children born to literate parents fail to complete primary schooling themselves.[10]

To construct absolute (upward and downward) IM measures at the country, and at the district (admin-2/3) level, we first define the following indicator variables:

- $\text{lit\_par}_{ibct} = 1$ 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.

- $\text{IM\_up}_{ibct} = 1$ 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.

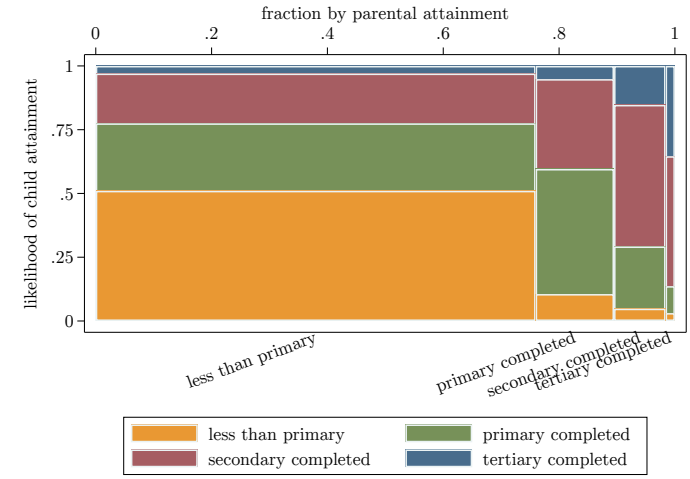
[8]See appendix A.2 for details of how we assign individuals to generations.

[9]Imposing (b), gives children a 3-year "buffer" between recorded education and the education they would have completed if they had continuously been at school up to that point. This ensures that recorded years of schooling reflect actual attainment. Since we drop younger individuals to allow Africans in our sample enough time to complete schooling, the reported statistics do not capture the sizable expansion of secondary and tertiary education during the last 20 years.

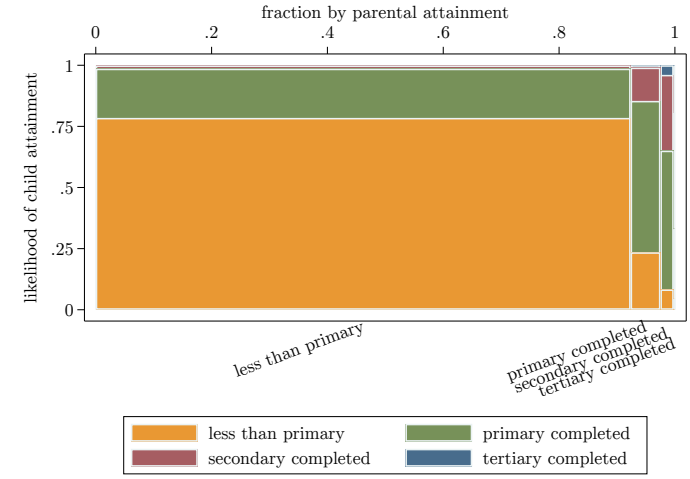
[10]Defining mobility in terms of primary education leaves our estimates less vulnerable to measurement error compared to IM estimates based on zero years of schooling. See appendix A.6 for details.

Figure 1: Educational Transition Likelihoods [Intergenerational Mobility in Education]

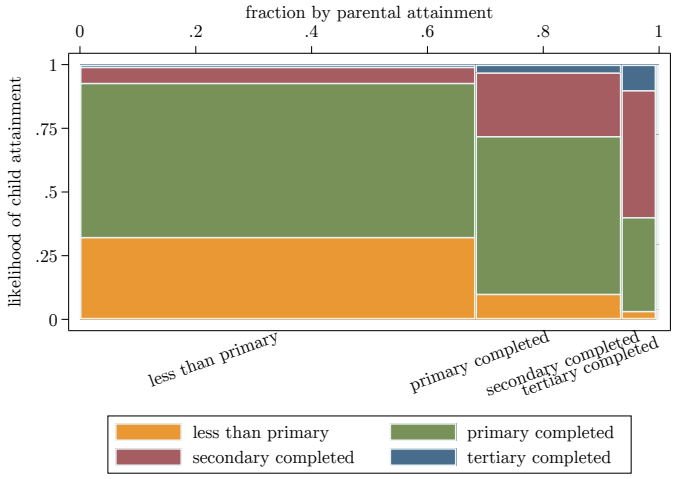
(a) Africa, 26 countries, 68 censuses



(b) Mozambique, 1997, 2007 census



(c) Tanzania, 1988, 2002, 2012 census



This figure shows the transition matrices for four broad educational attainment categories for all of Africa, Mozambique and Tanzania. Unlike in the remainder of the paper, the sample consists of individuals aged 18+ who are at least 9 years older than their years of schooling, co-residing with at least one individual of an older generation.

- $\text{IM\_down}_{ibct} = 1$ if a child i born to literate parents in birth-decade b in country c and observed in census-year t is illiterate and zero otherwise.

Then, for the country-specific analysis, we run the following regressions, pooling information from all censuses:

$$\text{lit\_par}_{ibct} = \alpha_c^o + [\gamma_b^o + \delta_b^y + \theta_t] + \epsilon_{ict} \tag{1}$$

$$\text{IM\_up/down}_{ibct} = \alpha_c^y + [\gamma_b^o + \delta_b^y + \theta_t] + \epsilon_{ict}, \tag{2}$$

where lit_par_{ibct} and IM_up/down_{ibct} are the indicators for parental literacy and child IM, respectively. To account for time trends, we condition on birth-decade fixed effects for the "young" (δ_b^y) and the "old" (γ_b^o) and census-year fixed effects (θ_t).

In estimating parental literacy (equation (1)), we are computing (conditional on cohort and time-effects) simple means among all individuals for whom we observe their parents' educational attainment. Hence, the estimated country fixed effects ($\hat{\alpha}_c^o$) reflect the shares of literate parents net of census-year and cohort effects. By contrast, when we compute measures of IM, we are computing <u>conditional</u> means. To this end, we estimate equation (2) twice: once for upward IM and once for downward IM. For upward IM, we estimate it only in the sample of children of illiterate parents, which allows us to interpret the country fixed effects ($\hat{\alpha}_y^o$) as the conditional likelihood that children born to illiterate parents become literate. For downward IM, we estimate it in the sample of literate parents and interpret $\hat{\alpha}_y^o$ as the conditional likelihood that children of literate parents become illiterate.

For the within-country analysis, we run similar specifications at the district level, r. We estimate country-by-country:

$$\text{lit\_par}_{ibcrt} = \alpha_r^o + [\gamma_b^o + \delta_b^o + \theta_t] + \epsilon_{ibcrt} \tag{3}$$

$$\text{IM\_up/down}_{ibcrt} = \alpha_r^y + [\gamma_b^o + \delta_b^o + \theta_t] + \epsilon_{ibcrt}. \tag{4}$$

2.4 Cohabitation Selection

We can only estimate IM of individuals who reside with their parents. This raises concerns of selection, as the intergenerational transmission of education may differ for cohabiting families and kids and parents who live apart. This issue is less pressing when focusing on young children that almost always cohabitate with their parents. Coresidence rates of children at the age of 8 and their parents exceed 99%. The problem is, of course, that the younger children are, 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 estimate IM for individuals aged 14-18. In this sample of $7,389,448$individuals, the coresidence rate is close to 94% (see appendix table A.6). The country in the sample with the lowest coresidence rate among 14-18 year olds is Guinea, with 82.3%, whereas Egypt and Lesotho have coresidence rates above 98%. While by looking at this sample, we miss tertiary and secondary attainment, in our setting most of the "action" is between no schooling and completed primary. We also work with a larger sample of $12,186,241$

individuals aged 14-25. This gives us a bigger sample, including college graduates, while cohabitation is still reasonably high (75%). In appendix A.5, we present simple statistics on the distribution of the *level* of education across countries and cohorts that do not rely on individuals co-residing with their parents.

3 Intergenerational Mobility across African Countries and Regions

This Section gives the main patterns of IM across African countries and regions. First, we present the cross-country statistics. Second, we provide a mapping of the African land of opportunity. Third, we report the cross-country and within-country across regions association between IM and literacy levels. Fourth, we distinguish across gender and rural-urban status.

3.1 IM across African Countries

Table 1 shows simple (unconditional) country-level estimates of intergenerational mobility.

Table 1: Country-level estimates of intergenerational mobility (IM)

	(1)	(2)	(3)	(4)	(5)	(6)
mobility / N	upward	upward	downward	downward	N with e_0 obs.	N with e_0 obs.
age range	14-18	14-25	14-18	14-25	14-18	14-25
South Africa	0.788	0.809	0.059	0.045	608,010	1,071,079
Botswana	0.704	0.713	0.057	0.049	22,558	36,415
Egypt	0.651	0.658	0.060	0.053	1,929,103	3,587,039
Nigeria	0.643	0.690	0.078	0.059	35,624	58,191
Zimbabwe	0.630	0.697	0.142	0.112	27,976	40,769
Tanzania	0.596	0.640	0.170	0.145	576,537	842,474
Ghana	0.578	0.585	0.129	0.108	298,264	483,839
Cameroon	0.522	0.541	0.103	0.093	188,275	295,440
Zambia	0.486	0.506	0.189	0.172	222,481	340,380
Kenya	0.450	0.520	0.205	0.159	469,787	719,194
Lesotho	0.437	0.487	0.253	0.208	24,197	42,910
Morocco	0.432	0.425	0.057	0.063	304,981	578,796
Benin	0.412	0.416	0.166	0.156	120,975	189,761
Uganda	0.365	0.408	0.282	0.238	274,831	401,937
Senegal	0.294	0.309	0.154	0.143	80,565	138,792
Rwanda	0.285	0.341	0.463	0.380	189,357	296,126
Sierra Leone	0.256	0.275	0.310	0.269	23,137	36,632
Guinea	0.229	0.244	0.308	0.285	51,278	81,339
Liberia	0.218	0.293	0.514	0.404	19,302	32,126
Mali	0.209	0.214	0.195	0.183	189,519	299,397
Burkina Faso	0.173	0.187	0.185	0.166	150,467	215,475
Malawi	0.158	0.233	0.445	0.341	195,856	294,976
Ethiopia	0.130	0.158	0.219	0.183	738,516	1,089,002
Sudan	0.112	0.163	0.367	0.249	400,140	651,194
Mozambique	0.101	0.143	0.432	0.341	174,810	245,422
South Sudan	0.040	0.068	0.747	0.631	36,451	58,768
mean / total	0.368	0.400	0.261	0.217	7,389,448	12,186,241

Columns (1) and (2) give upward-IM estimates. They reflect the likelihood that children, aged 14-18 and 14-25, whose parents have not completed primary schooling will manage to complete at least primary education. Columns (3) and (4) give downward-IM estimates. They reflect the likelihood that children, aged 14-18 and 14-25, whose parents have completed primary schooling or higher will not manage to complete primary education. Columns (5) and (6) give the number of observations used to estimate the country-specific IM statistics (children whose parental education is reported in the censuses). Countries are sorted from the highest to the lowest level of upward IM in the 14-18 sample (column (1)). "mean" gives the simple average of the 26 country-estimates.

Columns (1) and (2) ((3) and (4)) report upward (downward) IM means. Columns (5) and (6) give the number of children (young) for the $14-18$ and the $14-25$ sample. (The two

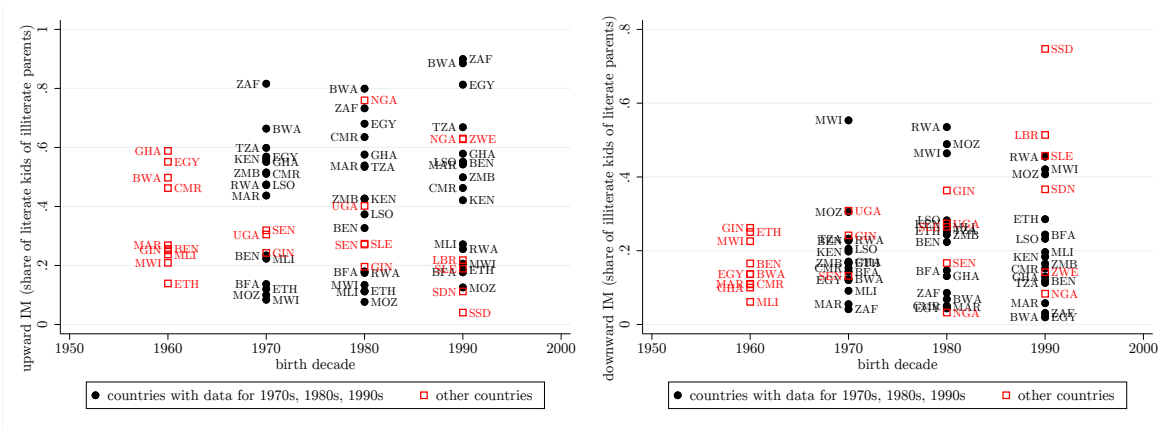
series are strongly correlated, $\rho > .97$).[11] In the average country, less than forty percent of children of illiterate parents have managed to complete primary education. Downward IM is lower, but far from negligible: approximately one out of four children born to literate parents does not complete primary education.

The pan-African mean masks sizable country variation. The likelihood that children of illiterate parents will complete at least primary education ranges from an abysmal 5% in South Sudan and 10% in Mozambique to 80% in South Africa and 70% in Botswana. The lowest upward IM is in the Sahel (Sudan, Burkina Faso and to a lesser extent Mali and Senegal) and the highest in Southern Africa (Botswana, Zambia, Zimbabwe, and South Africa) with Western and Eastern African countries in the middle. Downward mobility is negatively correlated with upward mobility ($\rho = -.76$ and $-.73$ for the two age groups). In South Sudan about 70 per cent of children born to "literate" parents fail to complete primary school, while the corresponding likelihood in Botswana, South Africa, Egypt, and Nigeria is less than 10 percent.

Do the simple cross-sectional averages obscure trends in IM? Figure 2 portrays the evolution of IM across cohorts. Panel (a) plots upward IM, while Panel (b) plots downward IM. It shows country-specific IM for children aged 14-18 and born in the 1960s (that corresponds for most countries to the first post-independence decade), 1970s and the 1980s (when many countries faced civil warfare), and the 1990s (when countries took the first steps towards democratic rule). The figure distinguishes countries with full cohort coverage and those without (Appendix Figure B.1 presents similar statistics for the $14-25$ age group).

Figure 2: IM at the country-birth-decade level, ages 14-18

(a) upward IM **(b)** downward IM



The figures report upward (panel A) and downward (panel B) Intragenerational Mobility in educational attainment (IM) across decade birth cohorts for children aged 14-18. Black solid circles indicate countries with data covering the 1970s, the 1980s, and the 1990s. Red hollow squares indicate countries with data covering just some cohorts.

Though not readily apparent because of the wide cross-country variability, there is a positive continental trend in upward-IM. The mean (median) IM goes from .36 (.27) for the 1960s-born, to .41 (.47) and .40 (.42) for children born in the 1970s and 1980s and rises to .48 (.5) for those born in the 1990s. Appendix Tables B.4 and B.5 present the regression analogues; compared to the 1960s-cohort, children born to illiterate parents in the 1990s

[11] As the census timing differs across countries, Appendix Table B.1 reports the corresponding statistics netting out cohort fixed-effects and census fixed-effects (equation (2)); the patterns are similar.

enjoy a 10 percentage points higher likelihood of completing primary education. There is sizable variation on the dynamics of IM. Upward-IM has increased considerably (by roughly 25 percentage points) in Botswana, Egypt and Benin. But it has remained roughly flat or even fallen in Ethiopia, Mozambique, Cameroon, Zambia, and Kenya. Downward IM increases slightly with an equivalent increase in variability across countries. Downward mobility increased in countries with devastating warfare in the 1980s and 1990s, such as Liberia, South Sudan, Rwanda, and Mozambique and fell in the more stable countries, Botswana, Egypt, and South Africa.

Nevertheless, the correlations between IM in the 70s, 80s, and 90s are strong, exceeding 85% for upward IM and 70% for downward IM, implying strong inertia. We further analyzed the cross-sectional and dynamic variability by regressing country-cohort IM on country fixed effects, then cohort fixed effects, and then country- and cohort fixed effects and comparing the in-sample fit. For the 17 countries with IM statistics covering the 70s, 80s, and 90s, this exercise reveals the following: When we add both sets of fixed effects, the R^2 is high, .924 and .847, for the upward IM and the downward IM specification, respectively. The strong fit reflects almost exclusively country features. The R^2 with just the country constants in the upward (downward) IM regression is .905 (.829). In contrast, the R^2 for the cohort-effects only specification is just .02 for both upward and downward IM. The stability of educational IM across cohorts in Africa echoes the pattern in India (Asher et al. (2018)).

3.2 Mapping the Land of Opportunity in Africa

Many African countries are large and there are evident disparities in geography and well-being. So where is the land of African opportunity? Figure 3 provides a mapping of social mobility across the continent. Panel (a) shows the distribution of upward IM across (mostly admin-2) districts and Panel (b) plots downward IM.

Figure 3: Pan-Africa: District-level estimates of IM, individuals aged 14-18

(a) upward; brighter colors → higher ↗ IM **(b)** downward; brighter colors → higher ↘ IM

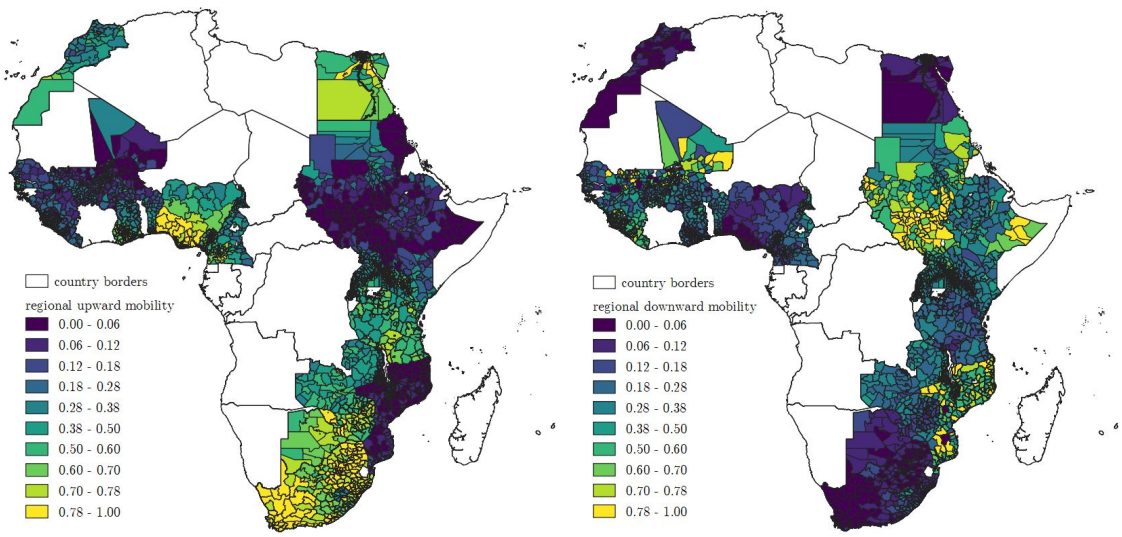
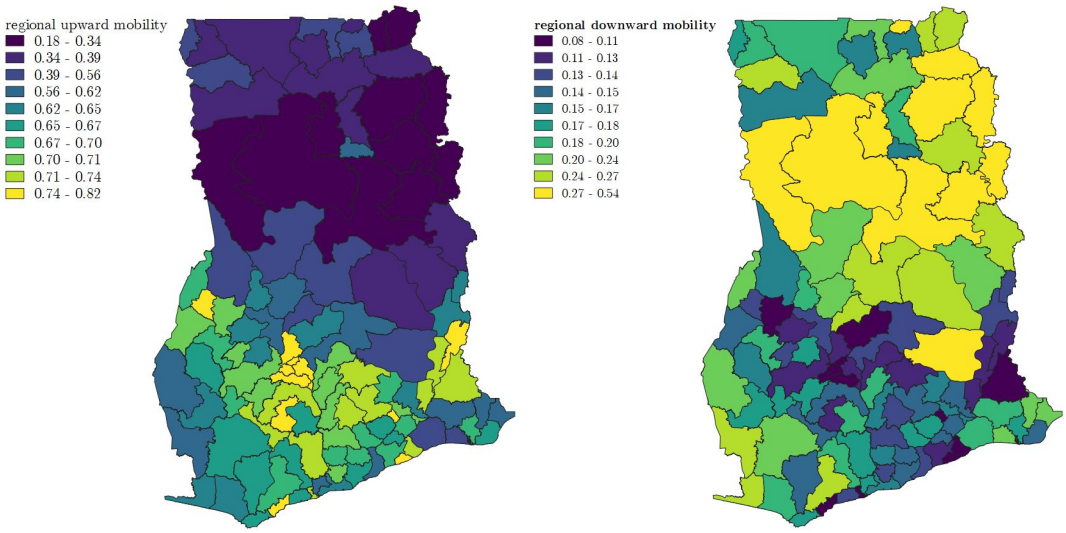


Table 2 gives summary statistics (mean, median, and range) by country. The district-level average and median for upward (downward) IM across the $2,813$ regions are 0.40 (0.30)

and 0.38 (0.25), respectively, quite close to the cross-country values. More strikingly perhaps, there is considerable variability in IM across regions in a given country.

Figure 4: Ghana: District-level estimates of IM, individuals aged 14-18

(a) upward; brighter colors → higher ↗ IM **(b)** downward; brighter colors → higher ↘ IM



As an example, figures 4 (a) and (b) portray upward and downward IM across 102 regions in Ghana. While average upward IM in Ghana is 0.57, regional IM ranges from 0.18 to 0.82 with rates below 0.4 in the Northern regions and above 0.7 in the South. The mean downward mobility is 0.20, but it varies from 0.08 to 0.50. This north-south gradient mirrors both the country's religious geography as well as colonial-era missionary activity and transportation investments, a topic we return to below.

Regional variation in IM is high in other countries (Table 2).[12] For example, in Burkina Faso the average upward-IM of 0.132 masks huge variability with regional IM ranging from 0.028 to 0.52. In Uganda the IM range is even wider $[0.015 - 0.69]$. Overall, spatial differences in IM are wider in countries with lower levels of social mobility. A simple linear regression of the coefficient of variation on mean upward IM yields a highly significant slope (s.e.) of -1.12 (0.18) with an R^2 of 0.63. But, even for countries with relatively high rates of upward IM, like Cameroon where a child born to illiterate parents has 52% probability of completing primary education, where the family resides plays a critical role: In some districts upward mobility is nearly guaranteed whereas in others it is almost impossible (the upward IM range is $0.09 - 0.89$).

Figure 5 plots the distribution of upward and downward regional IM for different cohorts. Upward-mobility for Africans born in the 1960s was quite low across the continent. Regional IM increased somewhat in the next two decades (by roughly 5%). The distribution shifts to the right for the 1990s-born children. Appendix Tables B.6 and B.7 report the regression analogues; conditional on district unobserved features, children of illiterate

[12] For some districts and census years downward mobility is 0 and in others is 1. These extreme values reflect the relatively small number of observations in some regions. The mean (median) district estimate is based on 1936 (891) matched-to-parents children ($st.dev = 3,287$). The estimates are similar if we limit attention to regions with at least 100 observations.

Table 2: Summary statistics: district-level estimates of IM

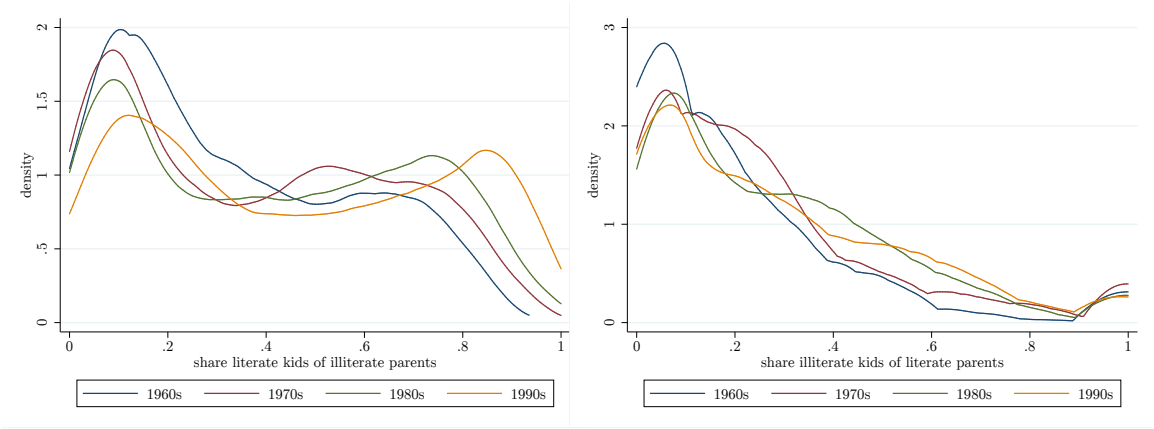
		upward					downward				
country	districts	mean	median	stdev	min	max	mean	median	stdev	min	max
South Africa	216	0.780	0.800	0.075	0.555	0.896	0.075	0.065	0.040	0.024	0.227
Botswana	23	0.719	0.714	0.079	0.554	0.909	0.069	0.070	0.031	0.000	0.143
Zimbabwe	88	0.715	0.726	0.148	0.400	1.000	0.156	0.157	0.084	0.007	0.402
Nigeria	37	0.708	0.765	0.201	0.330	0.963	0.085	0.070	0.043	0.019	0.166
Egypt	236	0.683	0.694	0.105	0.418	0.914	0.063	0.058	0.033	0.013	0.177
Tanzania	113	0.614	0.611	0.094	0.408	0.854	0.176	0.170	0.068	0.054	0.358
Ghana	110	0.588	0.650	0.158	0.181	0.820	0.183	0.166	0.074	0.077	0.541
Cameroon	230	0.551	0.588	0.203	0.088	0.896	0.210	0.168	0.147	0.000	0.875
Kenya	173	0.498	0.514	0.187	0.048	0.873	0.246	0.252	0.101	0.038	0.548
Zambia	72	0.479	0.467	0.127	0.284	0.785	0.266	0.278	0.097	0.074	0.468
Morocco	59	0.441	0.424	0.145	0.160	0.723	0.070	0.057	0.058	0.000	0.429
Lesotho	10	0.421	0.437	0.060	0.317	0.492	0.297	0.303	0.061	0.204	0.389
Uganda	161	0.383	0.382	0.128	0.015	0.696	0.352	0.353	0.117	0.124	1.000
Benin	77	0.381	0.381	0.132	0.111	0.649	0.218	0.204	0.077	0.083	0.455
Rwanda	30	0.297	0.280	0.063	0.220	0.469	0.494	0.518	0.102	0.226	0.623
Senegal	34	0.275	0.209	0.164	0.078	0.616	0.212	0.194	0.144	0.000	0.857
Sierra Leone	107	0.232	0.185	0.149	0.032	0.694	0.484	0.500	0.208	0.000	1.000
Ethiopia	97	0.208	0.119	0.235	0.000	0.865	0.349	0.311	0.202	0.000	1.000
Malawi	227	0.200	0.163	0.115	0.052	0.643	0.500	0.511	0.129	0.141	0.778
Liberia	47	0.185	0.180	0.081	0.034	0.345	0.569	0.566	0.167	0.000	1.000
Guinea	34	0.182	0.179	0.085	0.063	0.491	0.336	0.325	0.128	0.136	0.619
Sudan	129	0.153	0.097	0.144	0.001	0.551	0.533	0.500	0.200	0.224	1.000
Mali	242	0.144	0.128	0.096	0.013	0.578	0.388	0.333	0.269	0.000	1.000
Burkina Faso	45	0.129	0.123	0.080	0.029	0.526	0.252	0.231	0.118	0.000	0.714
Mozambique	144	0.091	0.064	0.086	0.015	0.707	0.587	0.581	0.202	0.000	1.000
South Sudan	72	0.043	0.024	0.056	0.000	0.319	0.831	0.848	0.158	0.400	1.000
overall	2813	0.405	0.379	0.269	0.000	1.000	0.304	0.247	0.235	0.000	1.000

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.

parents born in the 1990s face an 11 percentage points higher likelihood to complete primary education, as compared to children born in the 1960s in the same district. The variability of upward-IM has increased over time. Figure 5 (b) plots the evolution of regional downward IM. The mean has increased only slightly, but the variability has increased (st. dev. in 1960s = .20, st. dev. in 1990s = .25).

Figure 5: Distribution district × cohort level IM

(a) upward IM **(b)** downward IM

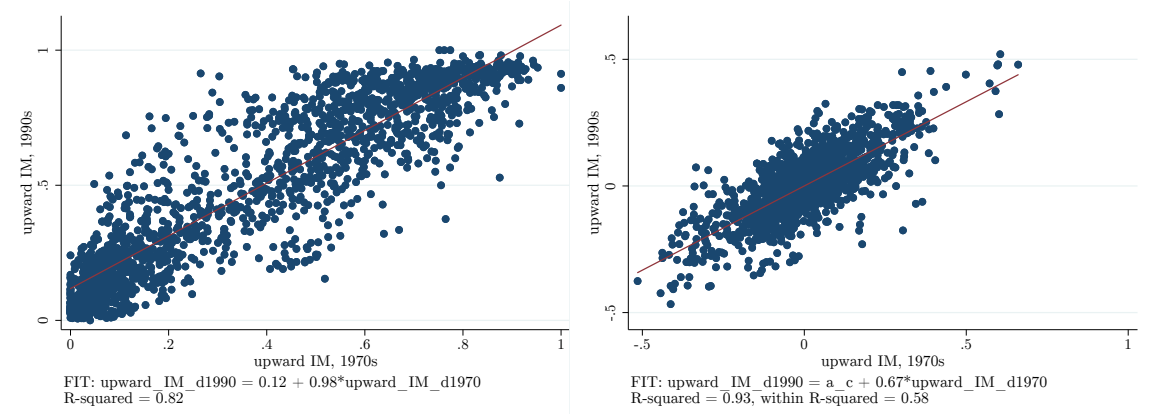


These figures plot the distribution of district-level upward (panel (a)) and downward (panel (b)) intergenerational mobility for the four birth-decades from the 1960s to the 1990s.

Similarly to the cross country patterns, IM is persistent at the sub-national level. Regressing upward district-level IM for 1990s-born cohort on the 1970s (1960s) cohort yields a slope of .98 (.87) and an R^2 of .819 (.59). Adding country fixed effects increases the R^2 to .932 (.82), while the slope decreases to .67 (.52). Figure 6 illustrates these patterns [See also Appendix figures B.2-B.3 for further evidence on persistence].

Figure 6: District-level upward IM over time

(a) OLS **(b)** country fixed effects



These figures visualize two regressions that link district-level upward IM in the 90s to district-level upward IM in the 70s. Panel (a) shows the simple linear regression, panel (b) shows the regression with country fixed effects.

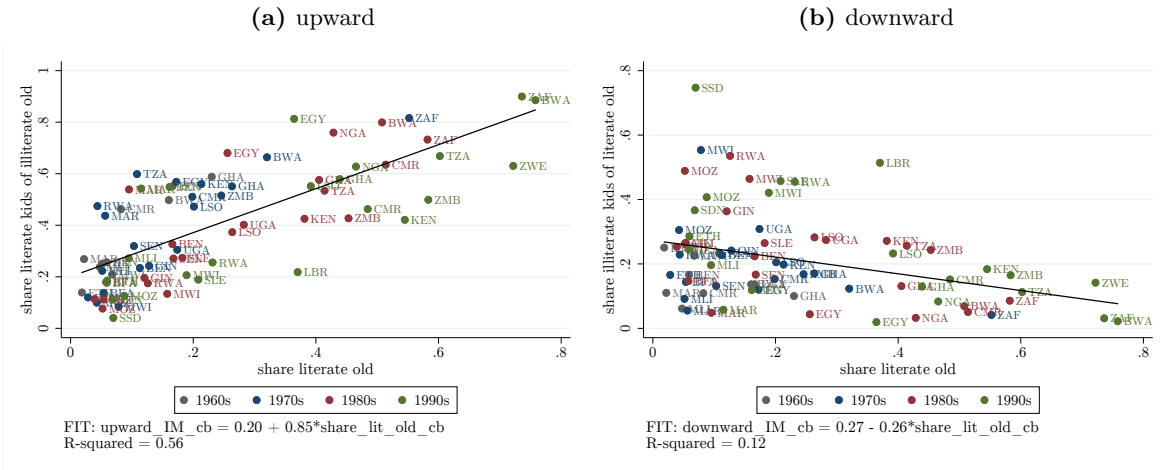
3.3 Literacy and IM

We commence the descriptive analysis correlating the newly-compiled IM measures with the literacy of the old generation. Our exploration is motivated by recent evidence, among others by Chetty and Hendren (2018a) and Chetty et al. (2016), showing that upward mobility is higher in regions with better outcomes (wealth, education, income). While these correlations do not have a causal interpretation, it is interesting to know whether the variability in social mobility varies systematically with the stock of education. [In Section 5 we tackle identification].

3.3.1 Cross-Country Patterns

While our focus rests on understanding the vast regional disparities within countries, we commence our analysis with a brief look at the cross-country picture. Figure 7 plots the relationship between IM (on the vertical axis) and the share of literacy of the old generation of the respective cohort (on the horizontal axis). Panel (a) explores the cross-country-cohort association for upward IM, while Panel (b) plots the corresponding association with downward IM. [Different colors show different cohorts.]

Figure 7: Literacy and IM at the country-birth-decade level



The figures plot upward-IM and downward-IM against across country-birth-cohorts against the share of the "old" generation that has completed primary education. The figures also report the simle OLS regression fit.

A positive association emerges between the share of completed primary of the "old" generation in the country and the likelihood that children whose parents have not completed primary schooling, manage to complete elementary school. In Ethiopia, Burkina Faso, Mozambique, North and South Sudan, where for all cohorts the share of literacy of the "old" generation is less than 20%, the likelihood that children from illiterate parents will complete primary is below or close to 20%. In contrast, the likelihood that children of illiterate parents will complete primary schooling exceeds 60% in countries-cohorts where the "old" generation is –on average– more educated, as, for example, in South Africa, and Botswana. The simple LS regression of IM on old generation's literacy pooled across all cohorts, suggests that a one percentage point increase in literacy is associated with a .85 percentage point increase in upward IM; the literacy of the "old" generation explains 56% of the cross-country-cohort variation in upward IM. The literacy of the "old" gen-

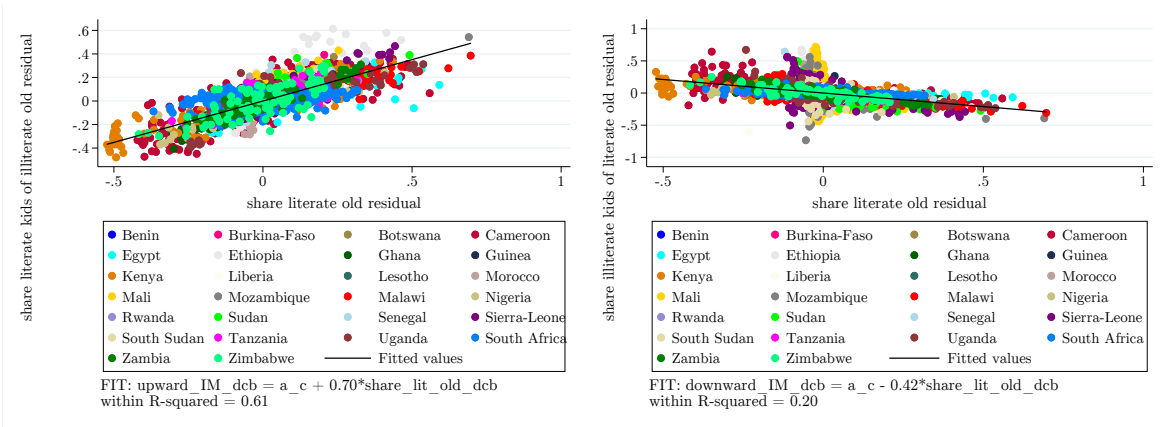
eration also correlates with downward IM, albeit more weakly.[13] A one percentage point increase in the "old" generation's literacy maps into a 0.26 fall in downward IM and the old generation's literacy explains 12% of the variation in downward IM.

3.3.2 Regional Patterns

Figures 8 (a) and (b) plot the association between upward and downward IM and mean literacy rates of the "old" generation across 2,813 districts (using different colors for different countries).

Figure 8: Literacy and IM at the district level

(a) upward **(b)** downward



To net out trends, we first run regressions (3) and (4) including all cohort and census-year fixed effects and extract coefficients $\hat{\alpha}_r^o$ and $\hat{\alpha}_r^y$. These are estimates of district-specific IM and parental literacy net of census and cohort effects. Second, we regress the "cleaned" measure of IM on the "cleaned" measure of the old generation's literacy, adding also country fixed-effects, i.e. $\hat{\alpha}_{cr}^y = \alpha_c + \beta \times \hat{\alpha}_{cr}^o + \epsilon_{cr}$. There is an evident positive association between the literacy of the "old" generation and upward IM, a pattern that echoes the cross-country one. Likewise, there is a negative -though less steep- correlation between downward IM and the literacy of the old. While there is considerable heterogeneity across countries, these correlations are not driven by some countries or cohorts.

Table 3 reports the regression estimates. Due to spatial correlation, standard errors are clustered at the province-level. The cross-sectional slope in columns (1) and (2) on the share of the old generation that has completed primary education is 0.77 and -0.486 in the upward-IM and the downward IM specification, respectively. Both estimates are highly significant. In columns (3) and (4) we add country constants. The within-country correlations retain economic and statistical significance. A 10-percentage points increase in the literacy of the "old" in the district is associated with a roughly 7 percentage point increased likelihood that the children of illiterate parents will manage to complete primary schooling and a 4.4% percentage points lower chance that kids of literate parents will not complete primary schooling. In columns (5) and (6) we replace the country constants

[13] This most likely reflects: (*i*) the smaller variability of downward-IM, as compared to upward-IM; and (*ii*) outlier observations in downward-IM, mostly coming from cohorts born in countries with sizable conflict (e.g., Sudan, Liberia, Sierra Leone).

with (280) admin-1 fixed-effects[14]); while there is nothing causal about these estimates, this accounts for provincial differences in literacy and IM. The estimates retain statistical significance though they drop in absolute value (0.56 and -0.385).

Table 3: Literacy and IM at the district-level

	(1) IM up	(2) IM down	(3) IM up	(4) IM down	(5) IM up	(6) IM down
share literate old	0.770***	-0.486***	0.703***	-0.422***	0.560***	-0.385***
	(0.029)	(0.032)	(0.040)	(0.030)	(0.041)	(0.035)
R2	0.650	0.305	0.885	0.700	0.934	0.755
within-R2			0.611	0.204	0.460	0.113
N	2809	2787	2809	2787	2726	2704
country-FEs	no	no	yes	yes	no	no
province-FEs	no	no	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). Standard errors clustered at the admin-1 (province)-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

In regions with relatively low levels of literacy among the old, the young face a lower likelihood of becoming literate. The strong positive (negative) association between the level of literacy of the old generation and upward (downward) IM among the young is suggestive of educational traps and divergence in educational attainment across countries. This finding is similar to the one of Asher et al. (2018), who also find that state's/region's mean education is the strongest correlate of upward educational mobility in India.

The estimates, while non-causal, suggest path dependence in human capital accumulation: disadvantaged children (from non-educated families) are more likely to complete primary school in regions with relatively higher initial stocks of literacy. This result is consistent with Jedwab et al. (2017), who document path-dependence stemming from colonial infrastructure investment (an issue we explore in the next section). Path dependence can reflect various mechanisms. First, inertia may stem from poverty trap dynamics that are especially salient in rural Africa, where (subsistence) agriculture is the typical mode of economic activity. Second, as regions with high levels of literacy tend to have better infrastructure (as we show in the next section), path dependence may stem from sunk costs in railroad-road construction. Third, given the limited state capacity of African states and the associated under-provision of public goods, sunk costs may also apply to school construction. Fourth, the results may reflect internal migration and sorting of families to regions with higher/lower educational opportunity (an issue that we examine in Section 5). Fifth, the estimates could at least partly reflect human capital externalities (e.g., Krueger and Lindahl (2001)) and peer effects that may be especially strong in a continent with large spatial differences in development.

[14]This is lower than the total number of provinces (365) because for Botswana, Lesotho, and Nigeria we only have province-level information.

3.4 Heterogeneity

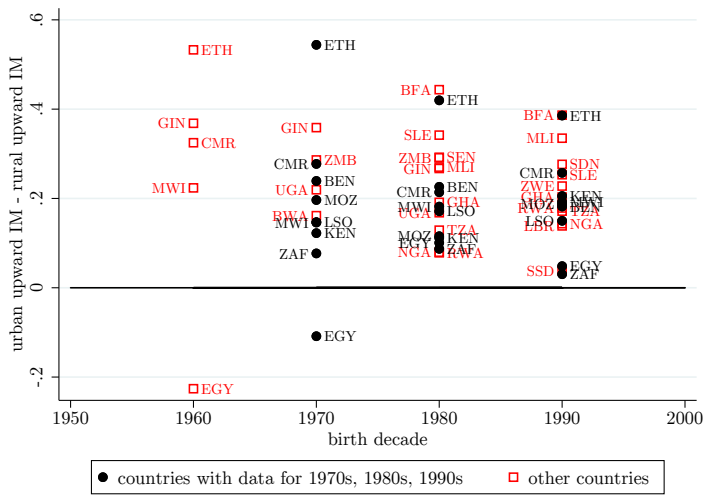
The census data allow us to construct country and regional measures of IM by gender and rural-urban status, enabling a more in-depth analysis of social mobility.

3.4.1 Rural-Urban

Appendix Table B.2 gives country-level IM for rural and urban households, using censuses' classification.[15] The country ordering is not much affected as the various IM measures correlate strongly; the correlation between rural and urban IM is 0.85 for both the upward and downward measures. Setting aside South Sudan, which is an evident outlier, upward IM in urban places ranges from 0.21 in Mozambique to 0.84 in Zimbabwe and South Africa (mean 0.33 and st.dev 0.225). The variability in rural upward IM is wider: on the one end, it hovers around 0.06 in Mozambique, Ethiopia, South and North Sudan, but on the other end it exceeds 0.6 in Nigeria, Egypt, Zimbabwe, Botswana, and South Africa (mean 0.53 and st.dev 0.197).

In Figure 9 we explore the evolution of differences in IM between rural and urban families. The horizontal axis portrays the children's birth-decade and the vertical axis plots the gap between average IM for children residing in urban versus rural areas, born in the same decade. There is a clear rural-urban divide with upward IM being lower for rural households; the average gap is 18% for all cohorts. This pattern applies to all countries, but Egypt. The rural-urban gap is the highest in countries with overall low levels of mobility and literacy. For example, there is a gap of about 40 percentage points between rural and urban places in Ethiopia and Burkina Faso; the rural-urban gap is below 10 percentage points in South Africa and Botswana.

Figure 9: Upward IM urban-rural gaps



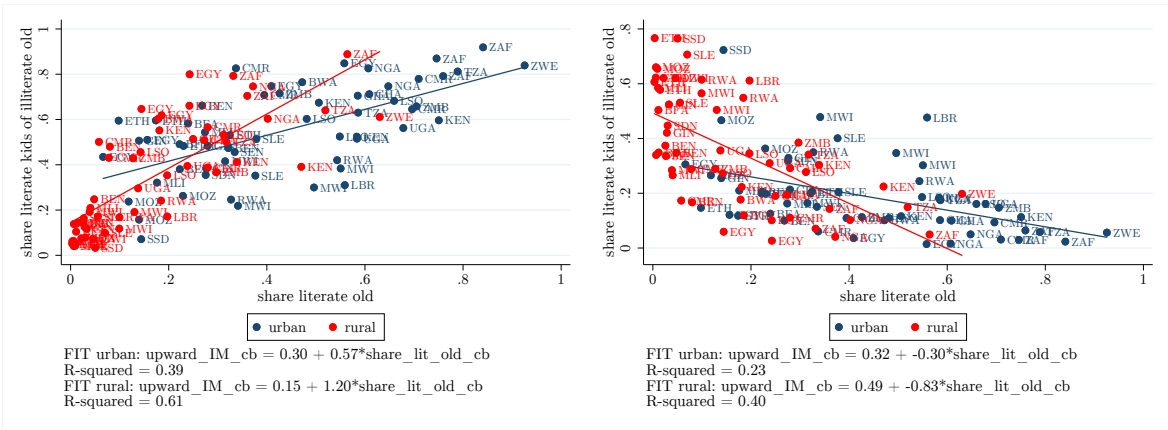
We then explored heterogeneity in the literacy-IM association between urban and rural households. Figure 10 plots the cross-country association between IM and old generation's

[15] The criteria for the rural-urban classification vary across countries. In some countries, they are based solely on population cutoffs, while in others they reflect localities' economic activity. In some instances, the statistical codebook does not provide any concrete information on the classification. Rural-urban status is not reported for Morocco. Appendix table A.1 gives details.

literacy separately for urban and rural households. Panel (a) looks at upward IM, while Panel (b) at downward IM.

Figure 10: Literacy and IM at the country-birth-decade level, urban/rural

(a) upward IM **(b)** downward IM



These figures visualize two regressions that link IM across countries-cohorts to old generation's literacy separately for rural and for urban places, that is, we estimate $\widehat{\text{IM}}_{cb}^{\text{urban or rural}} = \alpha + \beta \times \widehat{\text{lit\_par}}_{cb}^{\text{urban or rural}} + u_{cb}$. Panel (a) shows the scatter plots and regressions for upward IM and panel (b) those for downward IM.

Three observations emerge. The first regards the intercept. Children residing in rural places have a mere 13% base probability of upward mobility. The corresponding probability is 31% for children in urban areas. Likewise, in rural areas, children born to literate parents have a staggering 49% base probability of falling below their parental educational attainment. The corresponding statistic is 32% in urban areas. Second, the likelihood that kids of illiterate (literate) parents will (not) manage to complete primary education is positively (negatively) related to the mean education of the "old" generation for both urban and rural households. Third, the positive association between upward IM and literacy of the old generation is quite steep for rural households, while for urban households the association is flatter. The literacy of the old explains around 61% of the variation of rural households IM, while the R^2 for urban households is around 39%. In the downward IM plot, the slope is almost 3-times as large in rural as compared to urban households (0.3 versus 0.83); in rural areas with just a few educated old it is much more likely that children of literate parents will not complete primary schooling as compared to cities.

Table 4 explores heterogeneity in the within-country association between old generation's literacy and IM for children growing up in rural and urban places. In line with the cross-country patterns, the old's literacy - upward-IM correlation is considerately stronger in rural areas, 0.70 (in (1)) versus 0.48 (in (2)). A Chow test strongly rejects the null hypothesis of coefficient equality. In columns (5)-(6) we add province fixed-effects to partly account for broad geographic variation. The estimate in the urban sample is 0.375, while in the rural sample it is 0.55, a considerable and significant difference. The specifications in (3)-(4), (7)-(8) yield similar thought attenuated patterns. The correlation between the old's education in the district and downward IM is steeper (negatively) for rural, as compared to urban, households. The difference of the two slopes is around 10 percentage points in the province fixed-effects specifications. To the extent that those leaving villages and small towns to urban centers have higher aspirations and latent ability, the ramifica-

tions for rural Africa are dire, as the decline in the stock of education in the rural areas will lower upward and increase downward IM. We return to this issue in Section 5.

Table 4: Literacy and IM at the district-level, urban/rural

	(1) IM up	(2) IM up	(3) IM down	(4) IM down	(5) IM up	(6) IM up	(7) IM down	(8) IM down
share literate old	0.481***	0.698***	-0.259***	-0.436***	0.375***	0.546***	-0.225***	-0.318***
	(0.039)	(0.055)	(0.023)	(0.044)	(0.043)	(0.067)	(0.034)	(0.039)
R-squared	0.744	0.850	0.617	0.671	0.823	0.909	0.710	0.748
N	1930	2575	1842	2459	1930	2575	1842	2459
sub-sample	urban	rural	urban	rural	urban	rural	urban	rural
country FEs	yes	yes	yes	yes	no	no	no	no
province FEs	no	no	no	no	yes	yes	yes	yes
p: coeff-equal	0.0000		0.0000		0.0049		0.0217	

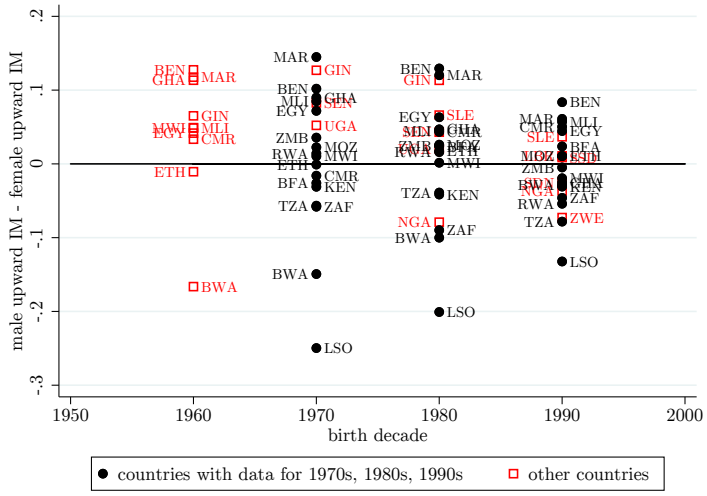
The dependent variable is the district-level share of literate kids of illiterate parents (estimated net of census year and old and young birth decade fixed effects). The independent variable is the district-level share of literate parents (also estimated net of fixed effects). Standard errors clustered at the admin-1 (province)-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. p-values for coefficient equalility in the uraban/rural sub-samples are from a Chow-test ($\sim \chi^2$ under H_0).

3.4.2 Gender Heterogeneity

To shed light on gender differences, (see Jayachandran (2015), Ashraf et al. (forthcoming), Bandiera et al. (2017), among others) we estimate IM for boys and girls. Appendix Table B.3 gives the country means. The correlation of the IM measures for boys and girls are very high (around .95) and as such the cross-country rankings are quite similar to the aggregate measures in Table 1.

Figure 11 shows the evolution of the male-female differences in IM. We do not observe major differences in the likelihood that girls whose parents have not completed primary schooling will manage to complete primary education, as compared to boys.

Figure 11: Upward IM male-female gaps, individuals aged 14-18

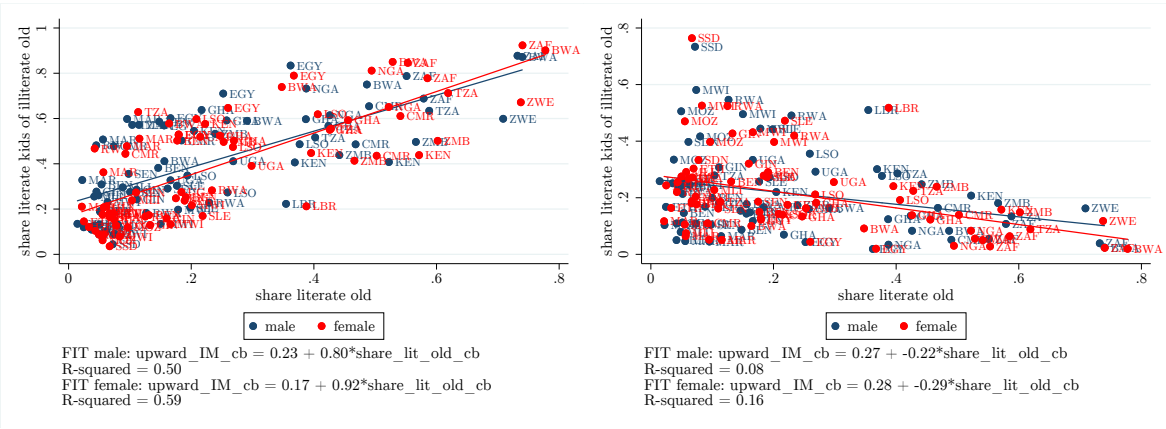


There is a small gender gap for the 1960s cohorts (especially when we exclude Botswana) that disappears in the 1980s and the 1990s cohorts. To be sure there are countries where boys are disproportionately favored compared to girls, including Benin, Mali, Egypt, Guinea, and Morocco, but girls in Lesotho, Botswana, Tanzania and South Africa, in fact, enjoy an edge over boys born to illiterate parents (see Appendix Table B.3).

We then explored heterogeneity in the association between literacy and IM across gender both at the country and the regional level. Figure 12 (a)-(b) plots the linear regression association between country-cohort level IM with country-cohort level parental literacy, estimated separately for girls and boys. Upward (downward) mobility is marginally higher (lower) for males compared to females. The positive (negative) association between the likelihood of (not) completing primary schooling for children born to illiterate (literate) parents and the share of literate old is equally strong for both boys and girls.

Figure 12: Literacy and IM at the country-birth-decade level, male/female

(a) upward, ages 14-18 **(b)** downward, ages 14-18



These figures visualize two regressions that link IM across countries-cohorts to old generation's literacy separately for male and for female places, that is, we estimate $\widehat{\text{IM}}_{cb}^{\text{male or female}} = \alpha + \beta \times \widehat{\text{lit\_par}}_{cb}^{\text{male or female}} + u_{cb}$. Panel (a) shows the scatter plots and regressions for upward IM and panel (b) those for downward IM.

Table 5: Literacy and IM at the district-level, male/female

	(1) IM up	(2) IM up	(3) IM down	(4) IM down	(5) IM up	(6) IM up	(7) IM down	(8) IM down
share literate old	0.648***	0.760***	-0.380***	-0.460***	0.506***	0.598***	-0.364***	-0.395***
	(0.039)	(0.044)	(0.028)	(0.037)	(0.043)	(0.048)	(0.036)	(0.040)
R-squared	0.864	0.869	0.660	0.629	0.921	0.925	0.718	0.706
N	2804	2805	2716	2742	2804	2805	2716	2742
sub-sample	male	female	male	female	male	female	male	female
country FEs	yes	yes	yes	yes	no	no	no	no
province FEs	no	no	no	no	yes	yes	yes	yes
p: coeff-equal	0.0000		0.0054		0.0000		0.1606	

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). Standard errors clustered at the admin-1 (province)-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$. p-values for coefficient equalility in the male/female sub-samples are from a Chow-test ($\sim \chi^2$ under H_0).

Table 5 associates the regional IM estimates with old generation's literacy separately for boys (odd-numbered columns) and for girls (even-numbered columns). Let us start with the upward-IM analysis in columns (1)-(2) and (5)-(6). The country (province) fixed-effects coefficient in the sample of boys is 0.65 (0.51), while in the girls sample it is larger, 0.76 (0.60); a Chow test of coefficient equality suggests that the difference is statistically different than zero. The environment, as captured in regional literacy, seems to matter more for girls as compared to boys. In regions with very low education level, girls from families without much schooling have a much lower chance than boys to complete primary education. The downward IM estimates in columns (3)-(4) and (7)-(8) yield similar results,

as the coefficients in the girls sample are always more negative, as compared to boys. In Section 5.2, when we re-examine gender asymmetries, we also find that regional exposure effects in the years relevant for primary education are stronger for girls.

3.5 Summary

The mapping of educational mobility across Africa reveals new regularities in the spatial distribution of opportunity across the continent. First, there are wide differences in IM across countries. Second, regional disparities in IM are even larger and appear especially wide in countries with low levels of education and mobility. Third, pan-African upward IM has somewhat increased since independence, though in many countries there have been no major changes. Fourth, upward IM is strongly linked to the average parental education in the country/region. Likewise, downward IM is inversely correlated to the literacy of the old generation, though this association is less strong. These patterns are consistent with poverty traps (or low convergence), since improvements in educational attainment are larger in regions with relatively higher human capital levels. Fifth, the correlation between the old generation's literacy and IM is especially strong for rural, as compared to urban households, a pattern that may partly be behind the recent rise in African urbanization. Sixth, the regional analysis reveals that the old's literacy IM correlations are somewhat stronger for girls, as compared to boys, suggesting that a favorable (unfavorable) environment may be particularly beneficial (detrimental) for girls.

4 The Geography of Intergenerational Mobility

In this Section, we examine the correlates of regional IM without any pretense of identifying causal effects. Our objective is to uncover a set of stylized facts that characterize the geography of educational IM. As Chetty et al. (2014), we run univariate specifications linking the proxies of intergenerational mobility with geographical, historical, and at-independence regional characteristics. As the literacy of the old generation is a strong correlate of IM, we also report specifications conditioning on it. The regression analysis, albeit simple, is useful in addressing the following question: Do the geographic or historical factors under consideration influence contemporary development through their linkages with initial conditions, that still matter due to strong inertia? Or are these factors also associated with the rate at which initial conditions are transmitted intergenerationally?

4.1 Specification

The empirical specification reads:

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

$G_{r,c}$ are geographic features of district r in country c; $H_{r,c}$ denotes historical, colonial and pre-colonial regional characteristics, and $Z_{r,c}$ are at independence economic features. Given the vast country heterogeneity we add country fixed effects (θ_c), though in the appendix we also report results without. In some permutations, we condition on the share of the old generation with completed primary schooling or higher, $L^o_{r,c}$. Appendix C.1 provides definitions and sources for all variables and also gives the summary statistics. Table

6 reports the estimates. Panel A looks at geographic features. Panel B looks at colonial and pre-colonial features, while Panel C looks at at-independence economic structure correlates of mobility. Column (1) reports the correlation between the variable specified on the left column with the share of literacy among the "old" generation; this allows benchmarking the IM estimates. Column (2) reports the correlation between the variable of interest and the likelihood that children born to illiterate parents complete primary education (upward IM); column (3) gives the correlation with upward IM, conditional on the share of the "old" generation with completed primary. Column (4) reports the number of regions. Columns (5)-(7) report analogous estimates looking at downward mobility. To make the estimates comparable, the table reports standardized "beta" coefficients that measure how many standard deviations the dependent variable changes in response to a one standard deviation change of the explanatory variable. Standard errors clustered at the province-level are reported below the estimates.

In the appendix, we report ivarious permutations that we comment on below: (i) in the $14-25$ age sample (table C.2); *(ii)* looking at the 1990s cohort that has the widest coverage (table C.5); (iii) replacing the country constants with province fixed-effects, so as to account for local unobservable features (table C.3); (iv) without the country constants (table C.4); and (v) jointly inserting geographic, historical, and at-independence factors on the RHS (tables C.6 and C.7).

4.2 Geography

Geography features prominently in explaining Africa's underdevelopment (e.g., Sachs (2006)). And given the strong inertia documented in the previous section, it is natural to examine the correlation of IM with geographic, and ecological features.

Distance to the Capital Much evidence documents the limited ability of African states to exercise control far from the capitals (e.g., Michalopoulos and Papaioannou (2014a) and Campante et al. (2019)). Even during colonization, the limited public goods were confined to the capital and a few urban hubs (Herbst (2000)). In line with this, column (1) shows that the literacy of the "old" is systematically higher in districts closer to the capitals. Column (2) reveals a significant association between proximity to capital and upward mobility. The standardized coefficient drops considerably, once we condition on the literacy of the "old" generation in column (3), from -0.29 to -0.094, though it remains precisely estimated. The picture is similar when we look at downward mobility: the likelihood that children of parents with completed primary education will not finish primary school is significantly higher in districts further from the capital. The association between IM and distance to the capital is robust to various perturbations, retaining significance when we run province fixed-effects specifications.

Distance to the Border African borders appear unruly and conflict prone.[16] The association between distance to the border and literacy of the "old" is weak (0.04) and

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

Table 6: District-level correlates of IM

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM controlling for share literate old	N	IM	IM controlling for share literate old	N
Panel A: geography							
ln(distance to capital)	-0.297***	-0.295***	-0.097***	2809	0.218***	0.092***	2787
	(0.038)	(0.039)	(0.026)		(0.029)	(0.023)	
ln(distance to border)	0.051	0.014	-0.024*	2809	-0.028	-0.003	2787
	(0.036)	(0.032)	(0.012)		(0.025)	(0.015)	
ln(distance to coast)	-0.195***	-0.242***	-0.105***	2809	0.154***	0.062***	2787
	(0.057)	(0.051)	(0.018)		(0.036)	(0.016)	
ln(1+malaria stability)	-0.242***	-0.242***	-0.071***	2798	0.168***	0.055	2776
	(0.049)	(0.052)	(0.026)		(0.044)	(0.035)	
ln(1+agricultural suitability)	-0.034	0.013	0.038**	2768	-0.018	-0.035	2746
	(0.056)	(0.049)	(0.019)		(0.036)	(0.027)	
ln(terrain ruggedness)	0.104**	0.113***	0.037*	2799	-0.087***	-0.038*	2777
	(0.048)	(0.040)	(0.019)		(0.033)	(0.020)	
oil field dummy	0.014	0.010	-0.000	2784	-0.003	0.004	2762
	(0.026)	(0.025)	(0.010)		(0.020)	(0.016)	
diamond mine dummy	-0.013	-0.015*	-0.005	2784	0.029**	0.022*	2762
	(0.012)	(0.009)	(0.007)		(0.013)	(0.012)	
Panel B: history							
ln(distance to railroad)	-0.315***	-0.329***	-0.094***	2299	0.247***	0.083***	2277
	(0.042)	(0.039)	(0.020)		(0.025)	(0.019)	
ln(distance to road)	-0.264***	-0.244***	-0.049***	2515	0.212***	0.082***	2493
	(0.028)	(0.028)	(0.016)		(0.021)	(0.020)	
ln(distance to cath. mission)	-0.362***	-0.342***	-0.087***	2809	0.235***	0.066**	2787
	(0.060)	(0.057)	(0.024)		(0.039)	(0.027)	
ln(distance to prot. mission)	-0.354***	-0.329***	-0.086***	2809	0.237***	0.081***	2787
	(0.046)	(0.038)	(0.019)		(0.029)	(0.022)	
ln(distance to precolon. empire)	0.023	-0.017	-0.034	2809	0.014	0.026	2787
	(0.040)	(0.029)	(0.026)		(0.032)	(0.031)	
ln(distance to precolon. state)	-0.037	-0.059*	-0.031*	2809	0.041*	0.022	2787
	(0.039)	(0.033)	(0.017)		(0.024)	(0.017)	
Panel C: contemporary							
ln(population density 1950)	0.226***	0.237***	0.081***	2797	-0.146***	-0.041**	2775
	(0.041)	(0.039)	(0.021)		(0.025)	(0.017)	
urban share (born < 1960)	0.378***	0.246***	-0.021	2531	-0.234***	-0.094***	2513
	(0.021)	(0.029)	(0.022)		(0.021)	(0.018)	
agri. empl. share (born < 1960)	-0.597***	-0.439***	-0.088**	2430	0.319***	0.100***	2412
	(0.029)	(0.025)	(0.038)		(0.024)	(0.024)	
manuf. empl. share (born < 1960)	0.234***	0.156***	0.004	2430	-0.105***	0.001	2412
	(0.041)	(0.033)	(0.017)		(0.027)	(0.017)	
serv. empl. share (born < 1960)	0.579***	0.434***	0.097**	2430	-0.319***	-0.110***	2412
	(0.031)	(0.026)	(0.039)		(0.022)	(0.022)	

This is not a normal regression table. In the column entitled "share literate old" the dependent variable is the district 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-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (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. 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.

the estimate does not pass standard significance thresholds. The correlation between (upward and downward) IM and distance to the border is likewise small; in most specifications the coefficient is statistically indistinguishable from zero. Neither IM nor education correlates strongly with distance to the border.

Distance to the Coast A cursory look on a satellite image of nighttime light density shows that African development is concentrated along the coastline. Proximity to the coast, among others, is linked to the presence of Europeans and associated investments during the colonial era. Distance to the coast correlates significantly with the "old" generation's literacy (column (1)). The specifications in column (2) ((5)) show that upward (downward) educational mobility is significantly higher (lower) in regions proximate to the coast. The coefficient retains significance when we condition on the literacy of the old (in (3) and (6)) though the estimate declines. The correlations retain economic and statistical significance, when we replace the country constants with province fixed effects.

Malaria Malaria has been invariably linked to Africa's underdevelopment (e.g., Gallup and Sachs (2001), Cervellati et al. (2016), Depetris-Chauvin and Weil (2018)). We associate the two proxies of intergenerational mobility (and the literacy of the "old") with an index reflecting a district's malaria ecology (from Kiszewski et al. (2004)). In line with earlier works, education is lower in regions ecologically friendly to malaria. Column (2) (column (5)) shows that upward (downward) educational IM is significantly lower (higher) in regions with an environment favorable for the transmission of malaria. Specifications (3) and (6) reveal that the negative (positive) association between malaria ecology and upward (downward) IM operates above and beyond initial differences in literacy. This – to the best of our knowledge – novel result suggests that malaria-prone regions are on a divergent trajectory. Initial educational conditions have been worse; but upward mobility is lower and downward mobility is higher in districts with malaria, even when we condition on the "level" effect.

Land Quality for Agriculture We then link IM to the quality of land (soil suitability) for agriculture, as the latter has been linked to economic outcomes both in the pre-industrial and contemporary era (Michalopoulos et al. (2019)). There is some weak positive association between soil quality and the stock of literacy; however the correlation between land suitability and IM never passes standard statistical significance thresholds.

Ruggedness We then examined the association between IM and ruggedness that correlates positively with cross-country economic performance in Africa. Different explanations have been proposed to rationalize this pattern that is unique to Africa. Nunn and Puga (2012) argue that regions with rugged terrain were shielded from Africa's slave trades that have been detrimental to Africa's long-run development (Nunn (2008)). Moreover, malaria stability is more pervasive in the lowlands compared to the highlands, protecting populations in the latter from the adverse effects of the disease. Setting the origins of this relationship aside, there is a positive and significant association between terrain ruggedness and the literacy of the "old" generation (column (1)), a result that adds to

the cross-country patterns of Nunn and Puga (2012). Columns (2)-(3) uncover that upward educational mobility is significantly higher in rugged regions, while columns (5)-(6) show that in regions with rugged terrain downward mobility is lower. However, when we replace the country constants with province fixed effects the association between IM and ruggedness breaks down.

Natural Resources A large literature on the "natural-resource curse" links conflict and other aspects of underdevelopment to the presence of oil, diamonds, and precious minerals. [See, among others, Ross (2004), Berman et al. (2017), Guidolin and La Ferrara (2007).] But, Hohmann (2018) shows that across African regions natural resource shocks are associated with higher education and structural transformation. We associated IM with indicator variables for the presence of diamond mines or oil fields. Natural resources are somewhat related to literacy, but the association with IM is weak at best. This most likely reflects opposing influences; conflict/animosity on the one hand and employment/development on the other. We also examined whether IM is related to proximity to other mineral sites (like silver or platinum mines), without detecting any correlation.

4.3 History

Drawing on the research agenda that links contemporary proxies of economic, social, and political development to colonial and precolonial features (see Michalopoulos and Papaioannou (2019) for an overview), in table 6 - Panel B we report specifications associating IM with historical features.

Colonial Road and Railroad Infrastructure Colonial investments in railroads and roads have played a crucial role in African countries' post-independence development and seem to explain path-dependence (e.g., Jedwab et al. (2017), Jedwab and Moradi (2016), Okoye et al. (2017), Huillery (2009)). We regress IM on the log distance to colonial railroads and colonial roads (data come from Jedwab and Storeygard (2017) and cover all Sub-Saharan African countries, but South Africa). There is a positive association between proximity to railroads-roads and literacy among the "old". Log distance to colonial railroads is significantly positively related to upward IM and negatively to downward IM. These patterns hold when we control for to the share of literacy of the old; the standardized coefficients are quite similar in the upward and downward IM specifications with both railroads and roads (around 0.08). Colonial railroads are not only associated with development at independence (as reflected in the education of the "old"), but also seem to boost intergenerational transmission.

Colonial Missions A considerable body of research uncovers sizable local effects of Christian, especially Protestant, missionary activity on education (Nunn (2014), Cage and Rueda (2017), Wantchekon et al. (2015), Okoye and Pongou (2014)). We thus examined the correlation between IM and proximity to colonial missions using digitized data from Nunn (2010) and Cagé and Rueda (2016). There are $1,321$ (361 Catholic, 933 Protestant, 27 British and Foreign Bible Society) and 723 (Protestant only) missions in these

datasets. The specifications in column (1) reveal a strong within-country positive correlation between proximity to Christian missions and literacy rates of the "old". There is a significantly positive (negative) correlation between proximity to missions with upward (downward) IM. When we condition on the literacy of the "old" generation, the coefficient on log distance declines, but remains statistically significant (beta around 0.09 for upward IM and 0.07 for downward IM). While data on Christian missions are incomplete and there may be systematic biases (Jedwab et al. (2018)), the analysis shows that pre-independence schooling investments of Christian missions may have lasting effects, both by shaping initial literacy which in turn increase educational mobility and also by affecting mobility directly.

Pre-colonial Political Centralization and Early Statehood We also explored the correlation between IM and pre-colonial political centralization that recent works link to contemporary development (e.g., Michalopoulos and Papaioannou (2013), Michalopoulos and Papaioannou (2014b), Gennaioli and Rainer (2007), Alsan (2015), and Depetris-Chauvin (2016)). We correlate IM with the distance to the centroid of the nearest large kingdom or empire using data from Brecke (1999), as geocoded by Besley and Reynal-Querol (2014) and log distance to pre-colonial states using Murdock's data (Murdock (1959), Murdock (1967)) though data are missing for some parts of the continent. There is no systematic link between distance to pre-colonial states and upward or downward IM with the Brecke (1999) data. The standardized coefficient is significant at the 90% level with the incomplete Murdock map. But the correlation loses significance when we control for province fixed effects in either sample.

4.4 At-Independence Correlates

In Panel C of table 6 we correlate IM with at-independence economic factors. For most variables, we use census data for individuals born before 1960. To net out migration effects (discussed in the next Section), we use information only from individuals who reside in their birth-district (the results are similar if we use all individuals). As we lack migration information for Lesotho, Nigeria, and Zimbabwe, the sample now spans 23 countries (see Appendix Table A.3).

At-Independence Development We look at how IM relates to (the log of) population density in 1950, which for most countries corresponds to the period just before independence, using data from Klein Goldewijk et al. (2010). Population density serves as a good proxy of local development in Africa that at the time was characterized by Malthusian dynamics. There is a significantly positive association between log population density in 1950 and the literacy of the "old" generation. Population density correlates positively and significantly with upward IM (column (2)) and negatively with downward IM (column (5)). The coefficients drop once we account for the share of the "old" generation with completed primary school, but the estimates are significant at the 99% level. Population density matters relatively more for upward -as compared to downward- IM ("beta" coefficients of 0.08 and -0.04).

Industrial Specialization Motivated by the literature on structural transformation in Africa (e.g., McMillan et al. (2014), Diao et al. (2017), Hohmann (2018)) and the evidence on persistence, we explored the correlation between IM with the share of employment in agriculture, manufacturing, and services at independence. The specifications in (1) show that initial human capital is considerably higher in regions with a relatively higher employment share in the "modern" sectors (services-manufacturing) as compared to the traditional sector (agriculture). The share of agriculture is significantly negatively correlated with upward mobility and positively correlated with downward mobility; these patterns also hold when we condition on the literacy of the "old" generation in the district. The standardized coefficients imply considerable magnitudes. A one standard deviation increase in the share of agricultural employment ($\sigma = .30$) is associated with a 0.44 standard deviation fall in upward mobility ($\sigma^{IM\_up} = .249$) and a 0.32 standard deviation increase in downward IM ($\sigma^{IM\_down} = .243$). The regressions with the share of services in the RHS yield a "mirror" image. Upward IM is significantly higher and downward IM is significantly lower in regions that had a higher share of employment in services. These patterns are also present in the $14-25$ age sample, when we just focus on the 1990s cohort and when we condition on admin-1 fixed-effects (see appendix C.2). Our results square well with the analysis of Asher et al. (2018) across Indian districts, who also document a strong positive association between manufacturing employment and educational mobility.

Rural-Urban Literacy is significantly higher in more urbanized regions. At the same time, upward mobility is higher and downward mobility is lower in relatively more urbanized regions. Once we condition on the share of literacy of the "old" generation, the correlation between upward IM and the share of urban households weakens and turns insignificant, while the correlation with downward IM retains its economic and statistical significance. Conditional on the education of the previous generation, in urban places the likelihood that kids of parents without schooling manage to complete primary schooling is not that different from more rural places; but, the likelihood that children of literate parents will not complete elementary schooling is considerably lower.

4.5 Summary

The correlation analysis that aims to characterize the spatial distribution of Africa's educational mobility shows that geographic and colonial-era features are related to IM. Proximity to the coast and the capital is related to higher (lower) upward (downward) mobility, even when one conditions on the initial "level" of literacy. Intergenerational mobility is also linked to terrain ruggedness (positive) and malaria (negative). In contrast, the correlation between educational mobility and natural resources is weak and statistically insignificant. Proximity to colonial railroads and Christian missions, that provided education and basic health, are also linked to higher levels of social mobility. Pre-colonial statehood correlates neither with educational mobility nor with education. At-independence development, reflected in regional population density, urbanization, and "structural transformation" proxies correlate strongly with education as well as intergenerational mobility, conditional on the education stock. This implies considerable inertia.

5 Regional Exposure Effects

Setting aside the origins of spatial differences in education and IM, to what extent do districts exert a causal effect on mobility? To answer this question, we look at migrant families and exploit within-household variation in children's exposure to regions with different degrees of intergenerational mobility. To the extent that a district's IM is a sufficient statistic of the economic and social environment that shapes educational decisions, within-household variation can help us in identifying the causal effects of regions on individual outcomes. We employ two approaches. First, we apply a straightforward household fixed-effects strategy looking at multi-children families, who have moved over time, thereby subjecting siblings to different environments. Second, we follow the approach of Chetty and Hendren (2018a) that exploit differences in the exact timing of children's moves across districts to capture regional exposure effects at different ages.[17]

5.1 Approach 1. Within-Family Estimates

5.1.1 Specification

We estimate the following regression in the sample of young individuals from households with at least two children born in different districts:

$$\begin{aligned}\text{IM\_up/down}_{ihbcrt} = \psi_h + \gamma_b^{\text{parents}} + \delta_b^{\text{child}} + \theta_{ct} + \lambda \times gender_{ihbcrt} \\ + \beta \times \widehat{\text{IM}}\text{\_up/down}_{bcr}^{\text{nm}} + \epsilon_{ihbcrt}.\end{aligned} \tag{5}$$

The dependent variable is an individual-level intergenerational mobility indicator. IM_up equals one if child i born in birth-decade b, region r and country c, to illiterate parents in household h is literate in census-year t. IM_down takes the value of one if a child of parents who have completed primary schooling is illiterate and zero otherwise. As with the purely observational estimates above, we look only at children of illiterate parents when we estimate the equation for upward IM and we only look on children of literate parents when we estimate the regression for downward IM. $\gamma_b^{\text{parents}}$ and δ_b^{child} are birth-decade fixed-effects for parents and children, respectively, *gender* indicates boys/girls, and θ_{ct} denotes country-census-year fixed effects. $\widehat{\text{IM}}\text{\_up/down}_{bcr}^{\text{nm}}$ is a country-district-birth-decade-of-the-child average IM, computed among non-movers (individuals born in the same place as the one that they reside at the time of the census), i.e.,

$$\widehat{\text{IM}}\text{\_up}_{bcr}^{\text{nm}} = \frac{\sum_i \text{IM\_up}_{ihbcrt}^{\text{nm}}}{\sum_i \mathbb{I}(\text{illiterate parents}_{ihbcrt}^{\text{nm}})} \tag{6}$$

$$\widehat{\text{IM}}\text{\_down}_{bcr}^{\text{nm}} = \frac{\sum_i \text{IM\_down}_{ihbcrt}^{\text{nm}}}{\sum_i \mathbb{I}(\text{literate parents}_{ihbcrt}^{\text{nm}})}. \tag{7}$$

[17] For all countries, except Lesotho, Nigeria, and Zimbabwe, IPUMS records an individual's birth place. For many countries, birth-place identifiers are not at the same level as residence. In some cases, birth places are at admin-1 level, whereas the residence is recorded at admin-2 level. In other cases, region of residence and birth place are at the same level. To assess migration status, we harmonized region of residence and region of birth, choosing the finest possible level of aggregation. We end up with $1,341$ "birth/current residence regions".

Crucially, ψ_h is a household constant that accounts for family features, related, among others, to ethnicity, religion, background, aspirations, etc. Hence, the coefficient of interest, β, is identified from within-family variation in children's birth place. Such variation arises because a family started out in one location, had one or more children there, and then moved to a different region, where other children were born.

Before reporting the results a caveat is in order. Estimation contains (non-negligible) measurement error, as we are not using information on the exact timing of children's move. As such, the cohort-regional place-of-birth IM variables may not reflect very well the environment that children experienced. We address this issue in the next subsection, where we exploit information in the exact timing of move. We view the simple within-family estimates as an introductory step towards the more elaborate estimates that follow.

Compared to the purely observational estimates, there are three major amendments in the sample. First, we run the above regression equation across individuals of families that have both migrant and non-migrant children. Second, since children of the same family that differ by just a couple of years will be subject to similar environments, even if born in different places, we require that they are born at least 5-years apart. Imposing this (ad hoc) restriction increases the chances that the children were not only born but also grew up in different regions. Imposing the 5-year gap means that we cannot estimate the specification in the sample of individuals $14-18$. We thus focus on the $14-25$ sample, but we also report results in the sample of individuals aged 14 and older. Third, to make the estimation of IM as clean as possible, we focus on children for whom we observe their mothers and fathers (whereas in the observational part we included extended family members in the estimation of previous generation's attainment).[18]

5.1.2 Results

Table 7 presents the results. Even-numbered specifications report the baseline household fixed effects estimates. Odd-numbered columns report otherwise identical regressions, but without the household constants. The comparison of the two sets of estimates allows us to gauge the role of selection.

Panel A looks at upward IM. The cross-sectional estimates in (1) and (3) show that the likelihood that children whose parents have not completed primary schooling will manage to finish primary school is significantly higher when born in regions with relatively high upward IM. In (2) and (4) we add household fixed effects to exploit within-family variation from migrant families with siblings born in different regions. The coefficients are positive and highly significant; the estimate in (2) implies that illiterate parents' children born in regions with a ten percentage points higher upward mobility have a 2.65 percentage points higher likelihood to complete primary schooling, as compared to their brothers and sisters. The within-family estimate is considerably smaller than the cross-sectional one, suggesting that family characteristics correlate strongly with mobility.

[18]The results are similar if we also include extended family members. We refrain from doing so to better isolate the influence of the "environment" from that of the "family".

Table 7: Household fixed effects estimates

Panel A: RHS = upward IM

	(1) upward IM	(2) upward IM	(3) upward IM	(4) upward IM
non-migrant upward IM	0.670***	0.265***	0.680***	0.341***
	(0.022)	(0.024)	(0.020)	(0.018)
R-squared	0.143	0.649	0.162	0.617
within R-squared	0.047	0.008	0.054	0.021
N	164258	164258	280055	280055
households	62960	62960	92933	92933
number of birth regions	1301	1301	1339	1339

Panel B: RHS = downward IM

	(1) downward IM	(2) downward IM	(3) downward IM	(4) downward IM
non-migrant downward IM	0.472***	0.333***	0.399***	0.269***
	(0.025)	(0.035)	(0.022)	(0.024)
R-squared	0.068	0.535	0.056	0.501
within R-squared	0.019	0.009	0.016	0.008
N	76529	76529	119030	119030
households	31207	31207	43699	43699
number of birth regions	1151	1151	1190	1190

country-year FEs	Yes	Yes	Yes	Yes
y+o cohort FEs	Yes	Yes	Yes	Yes
household FEs	No	Yes	No	Yes
age-range	14-25	14-25	14+	14+
minimum age gap	5	5	5	5

The dependent variable in panel A is an indicator equal to one if a child of parents without primary education completes at least primary and zero otherwise. In panel B it is an indicator equal to one if a child of parents with at least primary education does not complete primary education and zero otherwise. In panel A, the RHS variable of interest is average upward IM of mon-migrant individuals born in the same region and birth-decade as individual i. In panel B, it is average downward IM of mon-migrant individuals born in the same region and birth-decade as individual i. Standard errors clustered at the birth-region-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Panel B looks at downward IM. The significantly positive cross-sectional estimates in (1) and (3) suggest that the likelihood that children of parents with completed primary schooling will not manage to finish elementary schooling is significantly higher for children born in regions with low intergenerational mobility. In columns (2) and (4) we add household constants, to compare siblings born in different regions. The implied effects of regional IM fall, though the estimates are significant at the 99% confidence level. Places with higher average downward mobility do worse in keeping children of literate parents from slipping into illiteracy. Literate parents' children born in regions with a one percentage point higher downward educational mobility in their cohort face a 0.333 percent lower likelihood to complete primary schooling, as compared to their brothers and sisters.

5.2 Approach 2. Age-at-Move and Exposure Effects

Our second and preferred identification strategy to approximate regions' effects on mobility follows Chetty and Hendren (2018a) and exploits variation in the timing of move between origins and destinations to identify regional "exposure-effects". This approach focuses exclusively on migrant children; it compares the educational attainment of children who

moved to a better/worse region in terms of average mobility at different ages to identify the rate at which their education converges to those of permanent residents. The identification idea is that if regions have a causal effect on individual mobility, this effect should be stronger, the longer the exposure to the environment lasts[19]. We first describe the semi-parametric econometric specification and report the baseline estimates. Second, we present parametric estimates of regional exposure effects and explore asymmetries across gender and the direction of movement (from better to worse and regions and vice versa).

5.2.1 Econometric Specification

The starting point of the Chetty and Hendren (2018a) methodology is a generic specification that links children's outcomes to those of permanent residents in the destination. For children who moved from place of birth o to destination region d at age m, their education can be expressed with the following regression:

$$\begin{aligned}\text{IM\_up}_{ibmcod} = [\psi_h +]\ \alpha_{ob} + \alpha_m + \sum_{m=1}^{18} \beta_m \times \mathbb{I}(m_i = m) \times \Delta_{odb} \\ + \sum_{b=b_0}^{B} \kappa_b \times \mathbb{I}(b_i = b) \times \Delta_{odb}\ + \epsilon_{i,ibmcod},\end{aligned} \tag{8}$$

The dependent variable equals one if the child of illiterate parents manages to complete primary education (or higher) and zero otherwise (upward mobility). α_{ob} denote origin-region×birth-decade fixed effects. These account for unobserved factors at the level of where and when a child was born. The variable of interest, Δ_{odb}, is the difference in upward educational mobility of permanent residents (non-movers) in the destination versus origin for all children born in birth cohort b:

$$\Delta_{odb} = \widehat{\text{IM}}\text{\_up}^{\text{nm}}_{bd} - \widehat{\text{IM}}\text{\_up}^{\text{nm}}_{bo},$$

where mean region-cohort upward mobility is defined in equation (6). We estimate a different slope, β_m, for each age of move (years 1 to 18), controlling for any direct effect via age-of-move constants, α_m; these capture disruption effects and any other age-specific unobserved feature that affects the education trajectory. Following Chetty and Hendren (2018a), we augment the specification with interactions of destination-origin differences in cohort-specific IM with cohort fixed effects, to account for potential differential measurement error across cohorts (this has no effect on our estimates).

The idea behind equation (8) is that if children move from places with worse to places with better educational opportunities ($\Delta_{odb} > 0$), and exposure matters for educational outcomes, the earlier the move occurs, the greater the effect on the outcome (Chetty and Hendren (2018a)). Since we include (thousands of) origin-cohort fixed effects ($2,916$ in the $14-25$ sample and $4,175$ in the 14+ sample), variation comes from children born in the same place in the same time, who, however, move to regions with different social mobility.[20]

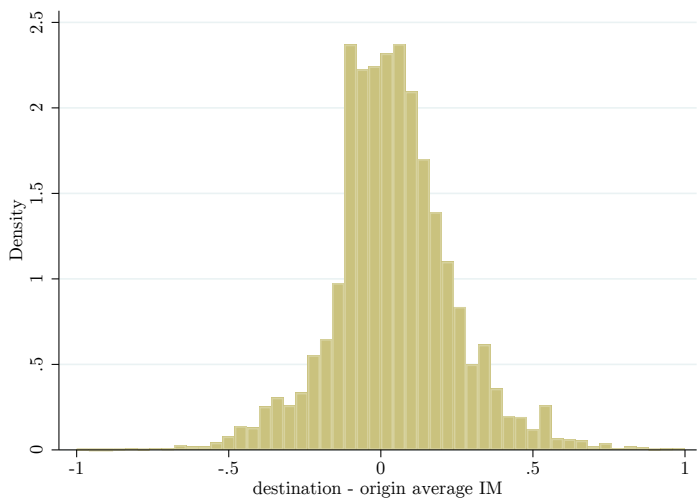
[19] See Chetty and Hendren (2018a) for references on the vast literature in sociology and economics of moving to better neighborhoods.

[20] The only difference vis a vis Chetty and Hendren (2018a) is that we are not interacting the origin-cohort effects α_{ob} with age-at-move m. Doing so would require adding more than $100,000$ fixed-effects, 1300 (regions) $\times$ 5 (cohorts) $\times$ 18 (age at move).

The age-specific slopes, β_m in equation (8), are identified even in presence of sorting; i.e., illiterate parents with higher latent propensity to educate their children are more likely to move to higher opportunity environments. The identifying assumption is that the timing of the move is not correlated with the latent ability of their (younger) children. In other words, parents who were more likely to invest in their children's education are allowed to move from worse to better environments on average compared to parents who were not going to do so, but the more ambitious parents should not move earlier rather than later. As this is not a weak assumption, we relax it estimating a household fixed-effects variant of equation (8). In this permutation (with ψ_h), the identifying assumption is that parents who move to better places do not do so to favor specifically some of their children.

Distribution of Δ^{nm}_{odb} Before reporting the results, it is useful to visualize the cohort-specific regional differences in IM (of non-movers) between origin and destination. $\Delta^{nm}_{odb} \lesseqgtr 0$, as families may move to places with better or worse IM. Figure 13 plots the histogram of Δ^{nm}_{odb} for children aged 14 to 25. The mean and median are positive, .04 and .03; on average, families move to regions with higher levels of upward mobility. However, migration flows both ways. Roughly $157,000$ children move to a region with higher IM (57%) and around $116,000$ (43%) move to regions with lower IM. These statistics complement the findings of Young (2013) who documents substantial bidirectional urban-rural migration flows across African regions with survey data. There is non-negligible variation; the standard deviation is 0.20.

Figure 13: Destination-origin differences in IM



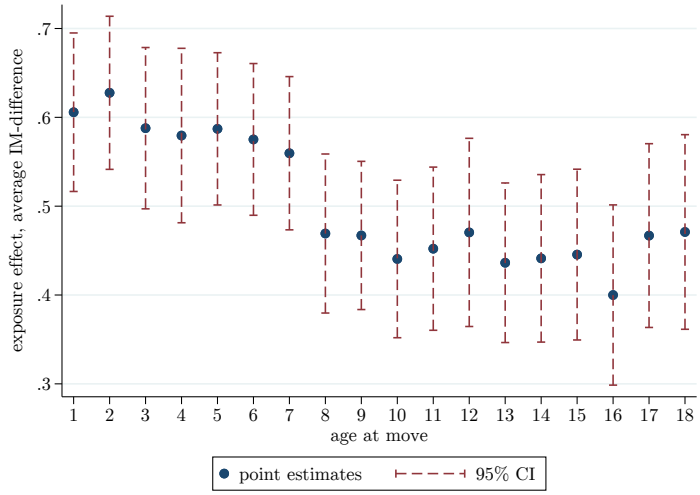
This figure plots the distribution of Δ^{nm}_{odb} – the destination minus origin differences in cohort-region average non-migrant IM – for all migrants children aged 14-25.

5.2.2 Baseline Semi-parametric Estimates

Figure 14 plots the estimated age-specific exposure effects, $\hat{\beta}_m$, against the child's age when their parents move. The figure uncovers two regularities: "*regional exposure effects*" that are particularly strong for children aged $6-11$ and "*selection effects*". The figure

also shows 95% confidence bands based on standard errors clustered at both origin and destination region levels.

Figure 14: Semi-parametric estimates of exposure effects



This figure plots observational (without household fixed effects) semi-parametric estimates of regional exposure effects $\hat{\beta}_m$ from equation (8) against the ages at which children move. 95% confidence bands based on double-clustered standard errors (origin and destination levels) are also shown.

First, the slopes are significantly positive for children moving at all ages. This applies even for children who move at the age of 14-18 ($\hat{\beta}_m \approx 0.45$).[21] As, almost by definition, moving at the age of $14-18$ cannot have a causal effect on primary educational attainment that is often completed by 12-13 years, these estimates provide direct evidence of *selection effects*. Families (with parents who have not completed primary schooling) moving to regions with higher (lower) IM have better (lower) unobservable characteristics translating into a higher likelihood that the children will finish primary school. The figure further shows that the degree of selection does not vary much with the age of children's move after the age of $14-15$.

Second, the non-parametric estimates reveal *regional exposure effects*, as moving to a better (worse) district earlier in life before ages $12-14$ generates a higher (lower) likelihood of upward mobility. The estimates are around 0.61 for children whose family moved before they turn 6 years old; the likelihood to complete primary schooling is 30 percentage points higher if their parents move to regions with 0.5 higher levels of IM (mean $IM = .6$, standard deviation $= .49$). The relationship between age of move and exposure effects is relatively flat for children moving before 6; moving to regions with higher mobility yields equally large benefits in the likelihood to complete primary schooling for children who are 1 or 4 years old. This is not surprising as primary education starts approximately at the age of 6. There is an evident declining pattern of the estimates for children moving between ages $6-12$. This suggests that a child moving at the age of 9 has a lower likelihood to complete primary education as compared to a child moving to the same region at the age of 8, 7, or 6. Following Chetty and Hendren (2018a), we define the exposure effect

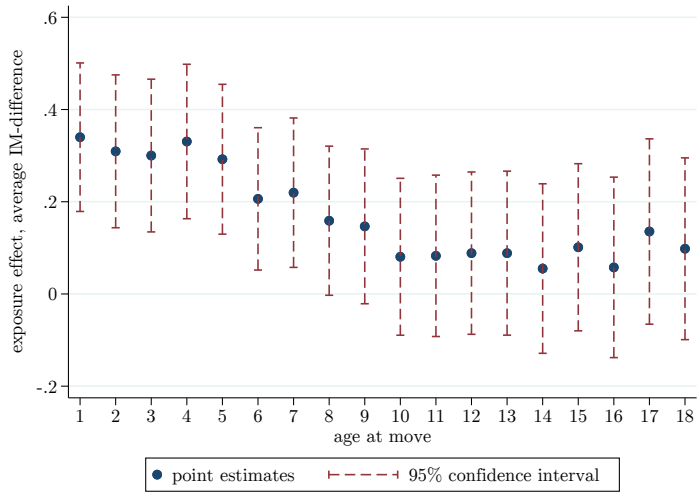
[21] We obtain similar in magnitude positive estimates for ages $19-25$.

as $\gamma_m = \hat{\beta}_{m+1} - \hat{\beta}_m$. Regressing the slopes on the age of move for ages 6 to 12 (that are relevant for primary schooling), we obtain an estimate of the average annual exposure effect of $-.019$.

5.2.3 Family Fixed-Effects Semi-parametric Estimates

We re-estimated the semi-parametric specification (equation 8) adding a vector of family constants. This is important, as family characteristics appear to be significant drivers of a child's probability of completing primary schooling. These estimates exploit variation from migrant-only children of the same family who moved at different ages. By exploiting within-family variation, we also relax the identifying assumption that required latent family features being orthogonal to the timing of move (see Chetty and Hendren (2018a)).

Figure 15: Semi-parametric estimates of exposure effects, household fixed effects



This figure plots household fixed effects semi-parametric estimates of regional exposure effects $\hat{\beta}_m$ from equation (8) against the ages at which children move. 95% confidence bands based on double-clustered standard errors (origin and destination levels) are also shown.

Figure 15 plots the age-specific exposure effects, $\hat{\beta}_m$, obtained by comparing siblings that moved at different ages. Estimation is carried out across $162,415$ children and $65,579$ multi-kid households. Two interesting patterns emerge. First, the selection effect captured in the slopes after age 11 drops significantly once we account for family unobserved features, from 0.45 to 0.07. 95% confidence intervals include 0 for all $\hat{\beta}_m$ after $10-11$. Family-specific constants account (almost) fully for selection, i.e., purge the estimation from the fact that families more (less) likely to educate their children move to regions with better (worse) educational opportunities. Second, the family-fixed-effects specifications also yield significant regional exposure effects. The slopes for children moving during ages $1-5$ are around 0.35; two siblings moving to a region with higher IM when they are 1 and 4 have, on average, the same increase in the likelihood to complete primary schooling. If the difference between destination and origin (Δ^{nm}_{odb}) is close to one standard deviation (0.5) the increase in upward-IM is around 18 percentage points for both siblings. The age-of-move slopes, $\hat{\beta}^{fe}_m$, fall for children moving when they are between ages 6 and 12,

suggesting that when a family moves to a higher IM region, the 6-year-old sibling benefits considerably more than her ten-year older sister. The estimate of the exposure effects for the critical-for-primary-schooling ages $(6-12)$ is $\gamma_m^{fe} = \hat{\beta}_{m+1}^{fe} - \hat{\beta}_m^{fe} = -.025$. This is similar to the cross-sectional estimate.

5.2.4 Baseline Parametric Estimates

Regression equation (8) is restrictive, as it includes thousands of origin-cohort fixed-effects; this issue becomes more challenging when we add family fixed-effects. Following Chetty and Hendren (2018a) we therefore estimate a parametric variant of specification (8).

$$\begin{aligned}
\text{IM_up}_{ibmcod} = & \sum_{b=b_0}^{B} \mathbb{I}(b_i = b) \times \left(\alpha_b^1 + \alpha_b^2 \times \widehat{\text{IM}}\text{_up}_{ob}^{\text{nm}}\right) + \\
& \sum_{m=1}^{18} \zeta_m \times \mathbb{I}(m_i = m) + \sum_{b=b_0}^{B} \kappa_b \times \mathbb{I}(b_i = b) \times \Delta_{odb} + \\
& \mathbb{I}(m_i \leq 5) \times (\beta_0 + (18 - m_i) \times \beta_1) \times \Delta_{odb} + \\
& \mathbb{I}(6 \leq m_i \leq 12) \times (\gamma_0 + (18 - m_i) \times \gamma_1) \times \Delta_{odb} + \\
& \mathbb{I}(m_i \geq 13) \times (\delta_0 + (18 - m_i) \times \delta_1) \times \Delta_{odb}. \qquad (9)
\end{aligned}$$

Instead of origin-cohort fixed effects, equation (9) includes birth-cohort effects interacted with a linear-in-origin-IM term (the first sum-term). The regression still includes age-at-move dummies to account for disruption effects and interactions between birth-cohorts and destination-origin differences to control for measurement error across cohorts. We no longer estimate separate age-of-move exposure effect slopes, but impose a piecewise linear structure, allowing the regional exposure effects to differ for pre-school years (ages $1-5$), the ages relevant for primary school $(6-12)$, and post-primary education years $(13-18)$.

Table 8 presents the parametric estimates in the 14-25 age sample (odd-numbered columns) and in the 14+ sample (even-numbered columns). Let us start with the cross-sectional estimates in columns (1) and (2). The exposure effect for children whose families moved after they were 13 is zero and statistically insignificant. In line with the non-parametric estimates, there is not much benefit for kids in completing primary school, when they move after that age. The estimated exposure effect for children moving between ages 1 and 5 is also small (0.015) and in general insignificant. The likelihood of completing primary school for children whose families move is positive but not much different for children who are very young at the time. In contrast, the exposure effect is significantly positive for children whose parents move when they are in the critical for primary school ages, 6 to 12. The estimate is $0.019-0.021$, similar to the semi-parametric estimates. The results are similar in the (smaller) sample of individuals that are included in the family fixed-effects specifications that for comparability we report in columns (3) and (4).

Columns (5) and (6) give household fixed-effects estimates. The regional exposure slopes are small and statistically insignificant for children whose families moved when they were older than 13 or younger than 6. The regional exposure slope is significantly positive for children moving between the age of 6 and 12. The coefficient is 0.023 and

tightly estimated. It suggests that if a family with two children moves when the old one is 11 years old and the young one is 6, from a location of zero IM to a location with an IM of one, the likelihood that the younger child would complete primary education is around 14 percentage points higher than the likelihood of her older sibling.

Table 8: Parametric exposure effects estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	IM	IM	IM	IM	IM	IM
β: 1-5	0.0139	0.0172*	-0.0000773	0.00510	0.000401	0.00693
	(0.012)	(0.010)	(0.014)	(0.011)	(0.018)	(0.013)
γ: 6-12	0.0205***	0.0191***	0.0171**	0.0144**	0.0227***	0.0190***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
δ: 13-18	-0.00515	0.000527	-0.00222	0.00400	-0.00508	0.00159
	(0.007)	(0.006)	(0.008)	(0.007)	(0.007)	(0.006)
R-squared	0.099	0.104	0.084	0.087	0.692	0.685
N	273537	335013	162708	199105	162708	199105
households	176523	214431	65694	78523	65694	78523
age at mig FE	yes	yes	yes	yes	yes	yes
birth decade FE	yes	yes	yes	yes	yes	yes
hh FE	no	no	no, hhfe sample	no, hhfe sample	yes	yes
age range	14-25	14+	14-25	14+	14-25	14+

The dependent variable in all regression is a dummy = 1 if the child has completed at least primary, and zero otherwise (i.e. a dummy for IM). The independent variables comprise a linear origin-average-IM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move dummies, birth-decade×destination dummies interacted with destination-minus-origin opportunity differences (to capture differences in measurement error across locations and cohorts), all of which not reported, as well as three linear terms for destination-minus-origin differences in relevant-birth-cohort-non-mover average IM for move-ages 1-5, 6-12, and 13-18. Standard errors clustered at origin- and destination-levels in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

5.2.5 Further Evidence. Heterogeneity

Table 9 reports parametric family fixed-effects specifications that explore heterogeneity across gender and across children that moved to regions with higher (lower) IM than their place of birth.

Gender Columns (1) and (2) examine heterogeneity across gender, estimating a variant of equation (9) where we interact the linear-in-age-at-move regional exposure effects with an indicator for girls. The slopes that capture exposure effects for boys moving before the age of 6 and after the age of 13 are unstable and insignificant. The interactions of Δ_{odb} with the female indicator that captures the additional effect for girls are also statistically indistinguishable from zero. These results are in line with the baseline estimates. The regional exposure effect for primary school age that reflects the impact for boys is significantly positive (0.013); the estimate, however, is smaller than the baseline estimate of 0.023. This is because the regional exposure effect when moving at ages $6-12$ is especially strong for girls. This is shown by the significantly positive coefficient on the interaction of Δ_{odb} with the female indicator that quantifies the extra benefit (loss) that girls get when they move to regions with higher (lower) mobility during that age. The coefficient on the interaction tern is 0.027 in the $14-25$ sample and 0.0165 in the 14+ sample, suggesting that girls benefit twice as much when moving to regions with better opportunities compared to boys.

Moving to Better-Worse Regions Columns (3) and (4) explore asymmetries between children moving to regions with higher or lower mobility than the origin. We do so by interacting the three linear-in-age-at-move regional exposure effects with a dummy variable that identifies moves to regions with lower than origin IM ($\Delta_{odb} < 0$). The estimates for ages $1-5$ and $13-18$ are small and statistically indistinguishable from zero. In contrast, the slope for children moving in the ages $6-12$ is positive and significant. The estimate is 0.03 is the $14-25$ sample and 0.019 in the 14+ sample. Moving to better places is associated with significant regional exposure effects for children of primary-school age. The interaction with moving to a region with worse mobility for the $6-12$ aged children is statistically indistinguishable from zero. This suggests that there is not much of an asymmetry between moves to better or worse conditions.

Table 9: Parametric exposure effects estimates, heterogeneity

male/female			**better/worse**		
	(1)	(2)		(3)	(4)
	IM	IM		IM	IM
β: 1-5, Δ^m_{odb}	-0.0169	0.00249	β: 1-5, Δ^+_{odb}	0.00298	0.0133
	(0.024)	(0.018)		(0.029)	(0.021)
γ: 6-12, Δ^m_{odb}	0.0137*	0.0126*	γ: 6-12, Δ^+_{odb}	0.0296**	0.0187*
	(0.008)	(0.007)		(0.012)	(0.011)
δ: 13-18, Δ^m_{odb}	-0.00739	0.00162	δ: 13-18, Δ^+_{odb}	-0.0148	-0.00186
	(0.009)	(0.008)		(0.011)	(0.009)
β: 1-5, $\Delta\Delta^f_{odb}$	0.0384	0.00810	β: 1-5, $\Delta\Delta^-_{odb}$	-0.00712	-0.0179
	(0.026)	(0.020)		(0.047)	(0.038)
γ: 6-12, $\Delta\Delta^f_{odb}$	0.0227***	0.0165**	γ: 6-12, $\Delta\Delta^-_{odb}$	-0.0179	0.000170
	(0.008)	(0.008)		(0.017)	(0.016)
δ: 13-18, $\Delta\Delta^f_{odb}$	0.00601	-0.000328	δ: 13-18, $\Delta\Delta^-_{odb}$	0.0235	0.00813
	(0.013)	(0.012)		(0.020)	(0.018)
R-squared	0.694	0.688	R-squared	0.692	0.685
N	162708	199105	N	162708	199105
age at mig FE	yes	yes	age at mig FE	yes	yes
birth decade FE	yes	yes	birth decade FE	yes	yes
hh FE	yes	yes	hh FE	yes	yes
age range	14-25	14+	age range	14-25	14+

The dependent variable in all regression is a dummy = 1 if the child has completed at least primary, and zero otherwise (i.e. a dummy for IM). The independent variables comprise a linear origin-average-IM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move dummies, birth-decade×destination dummies interacted with destination-minus-origin opportunity differences (to capture differences in measurement error across locations and cohorts), all of which not reported, as well as three linear terms for destination-minus-origin differences in relevant-birth-cohort-non-mover average IM for move-ages 1-5, 6-12, and 13-18. In columns (1) and (2), Coefficient estimates Δ^m_{odb} show the estimates for the reference group (male children). $\Delta\Delta^f_{odb}$ show estimates of differential effects for female children. In columns (3) and (4), coefficient estimates Δ^+_{odb} show the estimates for the reference group (movers to better places). $\Delta\Delta^-_{odb}$ show estimates of differential effects for movers to worse places. Standard errors clustered at origin- and destination-levels in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

6 Conclusion

In this study we conduct the first systematic exploration of intergenerational mobility in education across African countries and districts since independence.

We structure our analysis into three parts. In the first part, we construct estimates of intergenerational mobility in educational attainment across African countries and re-

gions, distinguishing by gender and rural-urban status. By mapping the African land of opportunity, we uncover sizable regional variation both across and within countries. The strongest correlate of IM is the literacy of the "old" generation. Persistence is stronger for rural, as compared to urban, households and present for both boys and girls. In the second part, we explore the geographic, historical, and economic correlates of intergenerational mobility across 2,809 regions. The goal is to characterize the wide regional variation in educational mobility and guide future research. Upward mobility is higher and downward IM is lower in regions with colonial investments in railroads and roads and areas with Christian missions. Geographical and location-specific features, including distance to the coast and the capital and an ecology favorable to malaria correlate negatively (positively) with upward (downward) IM. At-independence economic factors also relate to mobility. Upward mobility is significantly higher in initially more developed regions, with higher urbanization and employment in services and manufacturing. In the third part, we identify the causal effects of regions on educational mobility by exploiting within-family variation from children whose families moved when children where of primary school age. We document that sorting is sizable. At the same time there are significant regional exposure effects. Boys and (especially) girls whose families move from regions with lower to those with higher upward mobility have a much higher likelihood to complete primary schooling, when the move takes place before the age of 12, as compared to their older siblings.

Our analysis here -as well as in our companion paper Alesina et al. (2019) where we study ethnic and religious differences in educational mobility- call for future research. A first avenue is to examine the causal effects of historical factors on educational mobility, blending the newly compiled IM statistics – that exhibit country, region, cohort, gender, and rural-urban differences – with quasi-natural-experimental variation, exploring the economic mechanisms underlying path dependence, including colonial era and post-independence investments. A second possibility is to examine the role of nationwide educational policies and trade reforms on IM, a largely unexplored area in the context of Africa. A final avenue is to construct measures of each region's impact on IM – applying the approach of Chetty and Hendren (2018b) – and explore their determinants.

References

Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial States. *Journal of the European Economic Association 9*(2), 246–277.

Alesina, A., S. Hohmann, S. Michalopoulos, and E. Papaioannou (2019). Ethnic and Religious Intergenerational Mobility in Africa. *mimeo*.

Alesina, A., S. Stantcheva, and E. Teso (2018, February). Intergenerational Mobility and Preferences for Redistribution. *American Economic Review 108*(2), 521–554.

Alsan, M. (2015, January). The Effect of the TseTse Fly on African Development. *American Economic Review 105*(1), 382–410.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2017). Global Inequality Dynamics: New Findings from WID.World. *American Economic Review 107*(5), 404–409.

Asher, S., P. Novosad, and C. Rafkin (2018). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. *mimeo, Dartmouth College*.

Ashraf, N., N. Bau, N. Nunn, and A. Voena (forthcoming). Bride Price and Female Education. *Journal of Political Economy*.

Azam, M. and V. Bhatt (2015, December). Like Father, Like Son? Intergenerational Educational Mobility in India. *Demography 52*(6), 1929–1959.

Bandiera, O., N. Buehren, R. S. L. Burgess, M. P. Goldstein, S. Gulesci, I. Rasul, and M. Sulaimany (2017, July). Women's empowerment in action : Evidence from a randomized control trial in Africa. Technical Report 118203, The World Bank.

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

Bates, R. H. (2015, September). *When Things Fell Apart: State Failure in Late-Century Africa*. Cambridge University Press.

Becker, G. S., S. D. Kominers, K. M. Murphy, and J. L. Spenkuch (2018). A Theory of Intergenerational Mobility. *Journal of Political Economy 126*(S1), 7–25.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017, June). This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review 107*(6), 1564–1610.

Besley, T. and T. Persson (2013, February). *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton N.J.: Princeton University Press.

Besley, T. and M. Reynal-Querol (2014, May). The Legacy of Historical Conflict: Evidence from Africa. *American Political Science Review 108*(2), 319–336.

Black, S. E. and P. J. Devereux (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics*, Volume 4B, pp. 1487–1541. Elsevier.

Blake, J. (1985). Number of Siblings and Educational Mobility. *American Sociological Review 50*(1), 84–94.

Bowles, S. (1972, May). Schooling and Inequality from Generation to Generation. *Journal of Political Economy 80*(3, Part 2), S219–S251.

Brecke, P. (1999). Violent Conflicts 1400 AD to the Present in Different Regions of the World. In *1999 Meeting of the Peace Science Society, Unpublished Manusrcipt.*

Cagé, J. and V. Rueda (2016, July). The Long-Term Effects of the Printing Press in Sub-Saharan Africa. *American Economic Journal: Applied Economics 8*(3), 69–99.

Cage, J. and V. Rueda (2017, July). Sex and the Mission: The Conflicting Effects of Early Christian Investments on the HIV Epidemic in Sub-Saharan Africa. SSRN Scholarly Paper ID 3011430, Social Science Research Network, Rochester, NY.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2019). Capital Cities, Conflict, and Misgovernance. *American Economic Journal: Applied Economics*.

Card, D. (1999). The Causal Effect of Education on Earnings. In *Handbook of Labor Economics*, Volume 3, Part A, pp. 1801–1863. Elsevier.

Card, D., C. Domnisoru, and L. Taylor (2018). The Intergenerational Transmission of Human Capital: Evidencefrom the Golden Age of Upward Mobility. pp. 102.

Caselli, F., J. Ponticelli, and F. Rossi (2014). A New Dataset on Mincerian Returns. *unpublished*.

Cervellati, M., E. Esposito, U. Sunde, and S. Valmori (2016, September). Malaria Risk and Civil Violence. SSRN Scholarly Paper ID 2837708, Social Science Research Network, Rochester, NY.

Charles, K. K. and E. Hurst (2004, January). The Correlation of Wealth Across Generations. SSRN Scholarly Paper ID 468161, Social Science Research Network, Rochester, NY.

Chetty, R. and N. Hendren (2018a, August). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b, August). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2018, March). Race and Economic Opportunity in the United States: An Intergenerational Perspective. Working Paper 24441, National Bureau of Economic Research.

Chetty, R., N. Hendren, and L. F. Katz (2016, April). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review 106*(4), 855–902.

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

Davis, J. and B. Mazumder (2018, March). Racial and Ethnic Differences in the Geography of Intergenerational Mobility. SSRN Scholarly Paper ID 3138979, Social Science Research Network, Rochester, NY.

Depetris-Chauvin, E. (2016). State History and Contemporary Conflict: Evidence from Sub-Saharan Africa. Technical Report 475, Instituto de Economia. Pontificia Universidad Católica de Chile.

Depetris-Chauvin, E. and D. Weil (2018, May). Malaria and Early African Development: Evidence from the Sickle Cell Trait. SSRN Scholarly Paper ID 3174980, Social Science Research Network, Rochester, NY.

Derenoncourt, E. (2018). Can you move to opportunity? Evidence from the Great Migration. *Harvard University, mimeo*.

Diao, X., K. Harttgen, and M. McMillan (2017). The Changing Structure of Africa's Economies. *World Bank Economic Review 31*(2), 412–433.

Economist, T. (2000). Hopless Africa. *The Economist*.

Economist, T. (2011). Africa Rising. The Hopeful Continent. *The Economist*.

Economist, T. (2016). 1.2 Billion Opportunities. *The Economist*.

Fletcher, J. and J. Han (2018, December). Intergenerational Mobility in Education: Variation in Geography and Time. Working Paper 25324, National Bureau of Economic Research.

Gallup, J. L. and J. D. Sachs (2001, January). *The Economic Burden of Malaria*. American Society of Tropical Medicine and Hygiene.

Gennaioli, N., R. L. Porta, F. L. D. Silanes, and A. Shleifer (2014). Growth in Regions. *Journal of Economic Growth 19*(3), 259–309.

Gennaioli, N. and I. Rainer (2007). The Modern Impact of Precolonial Centralization in Africa. *Journal of Economic Growth 12*(3), 185–234.

Golley, J. and S. T. Kong (2013, March). Inequality in Intergenerational Mobility of Education in China. *China & World Economy 21*(2), 15–37.

Guidolin, M. and E. La Ferrara (2007, December). Diamonds Are Forever, Wars Are Not: Is Conflict Bad for Private Firms? *American Economic Review 97*(5), 1978–1993.

Gunning, J. W. and P. Collier (1999). Explaining African Economic Performance. *Journal of Economic Literature 37*(1), 64–111.

Herbst, J. (2000, March). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton, NJ: Princeton University Press.

Hertz, T., T. Jayasundera, P. Piraino, S. Selcuk, N. Smith, and A. Verashchagina (2008). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy 7*(2).

Hilger, N. (2017). The Great Escape: Intergenerational Mobility in the United States, 1930-2010. *unpulished manuscript*.

Hohmann, S. (2018). Natural Resources, Trade, and Structural Transformation in Africa. *unpublished manuscript*.

Huillery, E. (2009, April). History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa. *American Economic Journal: Applied Economics 1*(2), 176–215.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics 7*(1), 63–88.

Jedwab, R., E. Kerby, and A. Moradi (2017, August). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya. SSRN Scholarly Paper ID 3016860, Social Science Research Network, Rochester, NY.

Jedwab, R. and A. Moradi (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *The Review of Economics and Statistics 98*(2), 268–284.

Jedwab, R. and A. Storeygard (2017). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010. Technical Report 0822, Department of Economics, Tufts University.

Jedwab, R. and A. Storeygard (2018). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa. *mimeo*.

Jedwab, R., F. M. zu Selhausen, and A. Moradi (2018). The Economics of Missionary Expansion: Evidence from Africa and Implications for Development. Technical Report 2018-07, Centre for the Study of African Economies, University of Oxford.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. Ehrlich, and J. Sachs (2004). A global index representing the stability of malaria transmission. *American Journal of Tropical Medicine and Hygiene 70*(5), 486–98.

Klein Goldewijk, K., A. Beusen, and P. Janssen (2010, June). Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene 20*(4), 565–573.

Krueger, A. B. and M. Lindahl (2001, December). Education for Growth: Why and for Whom? *Journal of Economic Literature 39*(4), 1101–1136.

La Porta, R. and A. Shleifer (2014, September). Informality and Development. *Journal of Economic Perspectives 28*(3), 109–126.

Long, J. and J. Ferrie (2013, June). Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review 103*(4), 1109–1137.

Lujala, P., N. P. Gleditsch, and E. Gilmore (2005). A Diamond Curse? Civil War and a Lootable Resource. *Journal of Conflict Resolution 49*(4), 538–562.

Lujala, P., J. K. Rød, and N. Thieme (2007). Fighting over Oil: Introducing A New Dataset. *Conflict Management and Peace Science 24*(3), 239–256.

McMillan, M., D. Rodrik, and I. Verduzco-Gallo (2014). Globalization, Structural Change, and Productivity Growth, with an Update on Africa. *World Development 63*(C), 11–32.

Michalopoulos, S. and E. Papaioannou (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica 81*(1), 113–152.

Michalopoulos, S. and E. Papaioannou (2014a, February). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics 129*(1), 151–213.

Michalopoulos, S. and E. Papaioannou (2014b, December). On the Ethnic Origins of African Development: Chiefs and Precolonial Political Centralization. *Academy of Management Perspectives 29*(1), 32–71.

Michalopoulos, S. and E. Papaioannou (2016). The Long-Run Effects of the Scramble for Africa. *American Economic Review 106*(7), 1802–1848.

Michalopoulos, S. and E. Papaioannou (2019). Historical Legacies and African Development. *Journal of Economic Literature*.

Michalopoulos, S., L. Putterman, and D. N. Weil (2019). The Influence of ancestral lifeways on individual economic outcomes in sub-saharan africa. *Journal of the European Economic Association*.

Montenegro, C. E. and H. A. Patrinos (2014, September). Comparable Estimates of Returns to Schooling around the World. Technical Report WPS7020, The World Bank.

Murdock, G. P. (1959, December). *Africa: Its Peoples and Their Cultural History* (Ex-library/includes Map edition ed.). New York: McGraw-Hill Inc.,US.

Murdock, G. P. (1967). Ethnographic Atlas: A Summary. *Ethnology 6*(2), 109–236.

Narayan, A., R. V. der Weide, A. Cojocaru, S. Redaelli, C. Lakner, D. G. Mahler, R. Ramasubbaiah, and S. Thewissen (2018, May). *Fair Progress?: Economic Mobility Across Generations Around the World*. World Bank Publications.

Nunn, N. (2008). The Long-Term Effects of Africa's Slave Trades. *The Quarterly Journal of Economics 123*(1), 139–176.

Nunn, N. (2010, May). Religious Conversion in Colonial Africa. *American Economic Review 100*(2), 147–152.

Nunn, N. (2014). Gender and Missionary Influence in Colonial Africa. In *Akyeampong E, Bates R, Nunn N, Robinson JA Ed. Africa's Development in Historical Perspective.* New York: Cambridge University Press.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *The Review of Economics and Statistics 94*(1), 20–36.

Okoye, D. and R. Pongou (2014, August). Historical Missionary Activity, Schooling, and the Reversal of Fortunes: Evidence from Nigeria. SSRN Scholarly Paper ID 2484020, Social Science Research Network, Rochester, NY.

Okoye, D., R. Pongou, and T. Yokossi (2017). On the Dispensability of New Transportation Technologies: Evidence from the Heterogeneous Impact of Railroads in Nigeria. Technical Report 1703E, University of Ottawa, Department of Economics.

Olivetti, C. and M. D. Paserman (2015). In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review 105*(8), 2695–2724.

Olivetti, C., M. D. Paserman, and L. Salisbury (2018, July). Three-Generation Mobility in the United States, 1850–1940: The Role of Maternal and Paternal Grandparents. *Explorations in Economic History*.

Pinkovskiy, M. and X. Sala-i-Martin (2014). Africa Is on Time. *Journal of Economic Growth 19*(3), 311–338.

Porta, R. L. and A. Shleifer (2008). The Unofficial Economy and Economic Development. *Brookings Papers on Economic Activity 39*(2 (Fall)), 275–363.

Psacharopoulos, G. (1994, September). Returns to Investment in Education: A Global Update. *World Development 22*(9), 1325–1343.

Ramankutty, N., J. Foley, J. Norman, and K. McSweeney (2002). The global distribution of cultivable lands:current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography 11*(5), 377–392.

Ross, M. L. (2004, May). What Do We Know about Natural Resources and Civil War? *Journal of Peace Research 41*(3), 337–356.

Sachs, J. (2006, February). *The End of Poverty: Economic Possibilities for Our Time* (Reprint edition ed.). New York, NY: Penguin Books.

Solon, G. (1999). Intergenerational Mobility in the Labor Market. In *Handbook of Labor Economics*, Volume 3, Part A, pp. 1761–1800. Elsevier.

Spady, W. G. (1967, November). Educational Mobility and Access: Growth and Paradoxes. *American Journal of Sociology 73*(3), 273–286.

Survey, U. G. (1996). GTOPO30. *Sioux Falls, SD: U.S. Geological Survey Center for Earth Resources Observation and Science*.

Wantchekon, L., M. Klašnja, and N. Novta (2015, May). Education and Human Capital Externalities: Evidence from Colonial Benin. *The Quarterly Journal of Economics 130*(2), 703–757.

Young, A. (2012, August). The African Growth Miracle. *Journal of Political Economy 120*(4), 696–739.

Young, A. (2013, November). Inequality, the Urban-Rural Gap, and Migration. *The Quarterly Journal of Economics 128*(4), 1727–1785.

Social Mobility in Africa
Supplementary Online Appendix

Alberto Alesina*
Harvard University, IGIER, CEPR and NBER

Sebastian Hohmann†
Stockholm SSE SITE

Stelios Michalopoulos‡
Brown University, CEPR and NBER

Elias Papaioannou§
London Business School and CEPR

This draft: January 2019

Abstract

The Supplementary Appendix is structured into four sections. Section A gives details on sample construction. Section B reports additional descriptive evidence on the variation of intergenerational mobility (IM) in educational attainment across African countries and regions. Section C provides further correlational evidence characterizing the regional variation in IM. Section D reports a cross-validation analysis of the educational statistics. It provides graphical and descriptive evidence between education and various proxies of well-being from the Demographic and Health Surveys and the Afrobarometer Surveys.

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

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

*aalesina@harvard.edu
†sebastian.hohmann@hhs.se
‡smichalo@brown.edu
§eliaspapaioannou@london.edu

A Details on sample coverage and construction

Appendix Section A gives details on the sample. In particular, the Section gives further information on data construction; discusses the assignment of individuals across generations in the same household; presents co-residence rates by country; gives a simple cross-validation evidence of the IPUMS data with the widely-used Barro and Lee (2013) dataset; presents descriptive on educational attainment by country and cohort; and concludes with a discussion of measurement error on attainment data.

Section A.1 provides information on sample construction.
Table A.1 shows how we go from the raw IPUMS data to our sample.
Table A.2 gives details for each country-census (number of observations, number of regions and provinces, rural-urban status info) for the three core samples that cover young individuals: (i) aged 14-18, (ii) aged 14-25, and (iii) 14 and older.
Table A.3 gives for each country-census the number of observations and the number of districts for the migrant sample that we use in Section 5.1. There are three subsamples covering young individuals: (i) aged 14-18, (ii) aged 14-25, and (iii) 14 and older.
Table A.4 gives for each country-census the number of observations and the number of districts for the migrant sample for which we observe the exact timing of the move that we use in Section 5.2. There are three subsamples covering young individuals: (i) aged 14-18, (ii) aged 14-25, and (iii) 14 and older.

Section A.2 and associated Table A.5 discusses in detail the way we assign individuals to generations within households.

Section A.3 and associated Table A.6 reports co-residence rates for children aged 8 years old, children aged 14-15, and children aged 14-25 at the time of the Census, by country. The table also reports the number of observations in the 14-18 and the 14-25 age bracket.

Section A.4 reports the results of a simple cross-validation of the IPUMS data we use with the Barro and Lee (2013) statistics.
Figure A.1 (a) reports the cross-sectional correlation for mean years of schooling for individuals aged 25-99 in the two samples.
Figure A.1 (b) reports the within-country correlation for mean years of schooling for individuals aged 25-99 in the two samples.

Sub-section A.5 gives an overview of the evolution of schooling using the IPUMS data.
Table A.7 reports for each country and each birth cohort (1950s, 1960s, 1970s, and the 1980s): (i) means years of schooling; (ii) the share of individuals with less than completed primary education; (iii) the share of completed primary; (iv) the share of completed secondary; and (v) the share of completed tertiary.
Figure A.2 plots the cumulative density function (CDF) for 1950s, 1960s, 1970s, and 1980s cohorts pooling all observations (individuals aged 25-99) across the continent .

A.1 Data construction

Table A.1: Sample construction

(1) country	(2) year	(3) fraction	(4) N_{all}	(5) N_{age}	(6) N_{owned}	(7) $N^{age\geq 14}_{owned}$	(8) $N^{14\leq age\leq 25}_{owned}$	(9) $N^{14\leq age\leq 18}_{owned}$	(10) $N^{age\geq 14}_{olded}$	(11) $N^{14\leq age\leq 25}_{olded}$	(12) $N^{14\leq age\leq 18}_{olded}$	(13) $N^{age\geq 14}_{olded,\ no\ mghh}$	(14) $N^{14\leq age\leq 25}_{olded,\ no\ mghh}$	(15) $N^{14\leq age\leq 18}_{olded,\ no\ mghh}$
Benin	1979	10	331,049	329,784	244,898	171,690	62,112	24,322	44,979	24,559	13,155	21,762	16,330	9,948
Benin	1992	10	498,419	498,107	435,652	256,763	101,543	44,735	68,131	46,505	27,155	32,014	28,023	18,515
Benin	2002	10	685,467	685,467	612,658	373,452	155,832	69,048	109,876	80,022	47,492	58,448	51,876	33,734
Benin	2013	10	1,009,693	1,009,693	911,604	559,525	240,049	108,694	194,159	147,369	83,206	105,603	93,532	58,778
Botswana	1981	10	97,238	96,187	72,951	50,399	20,258	9,533	14,165	9,960	5,817	5,354	4,380	2,949
Botswana	1991	10	132,623	132,623	113,172	78,814	32,680	15,830	22,878	16,568	10,117	9,062	7,658	5,246
Botswana	2001	10	168,676	168,134	159,257	109,649	44,806	20,616	36,027	25,392	14,158	14,071	11,729	7,224
Botswana	2011	10	201,752	201,235	190,212	138,375	48,926	20,677	40,499	25,365	12,806	16,566	12,648	7,139
Burkina Faso	1985	10	884,797	883,447	484,384	410,398	159,162	75,374	0	0	0	0	0	0
Burkina Faso	1996	10	1,081,046	1,075,824	803,264	552,402	226,436	114,148	156,508	115,900	77,241	102,117	92,338	65,370
Burkina Faso	2006	10	1,417,824	1,410,123	1,244,291	770,161	321,384	151,393	178,512	155,609	103,865	133,585	123,137	85,097
Cameroon	1976	10	736,514	736,320	605,749	413,814	157,287	72,886	78,693	56,718	36,652	58,095	48,566	33,088
Cameroon	1987	10	897,211	896,649	763,652	481,727	191,552	90,805	93,198	70,283	45,721	73,851	62,728	42,368
Cameroon	2005	10	1,772,359	1,772,359	1,542,200	1,018,632	438,407	199,054	311,011	238,256	138,181	218,037	184,146	112,819
Egypt	1986	14.1	6,799,093	6,794,386	5,418,332	4,262,426	1,609,719	722,024	1,707,373	1,282,195	672,678	1,275,624	1,064,503	587,721
Egypt	1996	10	5,902,243	5,901,839	4,453,382	3,810,835	1,471,285	718,874	1,494,145	1,201,616	686,996	1,222,560	1,072,794	631,717
Egypt	2006	10	7,282,434	7,282,434	5,739,722	5,096,618	1,977,932	785,619	1,916,007	1,562,332	753,720	1,673,273	1,449,742	709,665
Ethiopia	1984	10	3,404,306	3,398,027	2,733,575	1,800,650	620,022	303,780	381,363	296,106	204,811	318,437	273,988	194,308
Ethiopia	1994	10	5,044,598	5,044,597	4,201,616	2,833,214	1,224,762	614,179	793,792	688,073	451,168	720,927	656,900	435,645
Ethiopia	2007	10	7,434,086	7,434,086	1,097,614	744,744	331,544	161,226	211,838	183,508	121,605	173,103	158,114	108,563
Ghana	1984	10	1,309,352	1,309,351	1,050,813	747,642	302,953	142,526	271,505	195,218	111,672	118,235	101,459	65,768
Ghana	2000	10	1,894,133	1,894,133	1,730,902	1,152,128	434,882	200,000	310,913	225,828	129,369	180,293	149,419	92,451
Ghana	2010	10	2,466,289	2,466,289	2,262,894	1,575,528	603,020	270,162	499,171	361,532	200,837	279,364	232,961	140,045
Guinea	1983	10	457,837	457,778	364,805	275,065	99,816	44,129	44,403	36,885	22,662	44,403	36,885	22,662
Guinea	1996	10	729,071	727,246	551,619	397,137	148,064	69,165	114,081	77,077	44,747	53,012	44,454	28,616
Kenya	1969	6	659,310	659,310	659,310	394,835	167,003	67,260	64,079	50,053	32,553	42,229	37,861	26,334
Kenya	1979	6.7	1,033,769	1,031,996	853,843	593,682	267,515	132,599	0	0	0	0	0	0
Kenya	1989	5	1,074,098	1,072,777	828,512	578,099	259,837	125,884	162,767	135,792	88,062	149,203	129,864	84,770
Kenya	1999	5	1,407,547	1,407,547	1,191,268	832,083	378,922	176,867	215,230	181,182	113,599	202,324	175,342	110,568
Kenya	2009	10	3,841,935	3,841,935	3,402,695	2,246,737	955,548	432,424	657,022	536,829	328,455	418,780	376,127	248,115
Lesotho	1996	10	187,795	187,795	165,960	121,446	50,160	24,283	50,332	36,835	19,312	29,347	24,226	13,863
Lesotho	2006	10	180,208	180,208	171,947	123,644	50,609	22,361	45,362	33,269	16,401	22,139	18,684	10,334
Liberia	1974	10	150,256	150,256	127,442	91,811	34,393	16,014	0	0	0	0	0	0
Liberia	2008	10	348,057	348,057	294,517	210,111	87,459	38,854	60,004	45,987	25,951	38,266	32,126	19,302
Malawi	1987	10	798,669	798,193	657,998	447,247	176,370	81,029	72,558	62,944	41,720	72,558	62,944	41,720
Malawi	1998	10	991,393	991,393	826,197	582,694	251,873	114,846	109,301	96,672	64,674	109,301	96,672	64,674
Malawi	2008	10	1,341,977	1,341,046	1,161,773	736,175	307,167	135,833	152,144	135,360	89,462	152,144	135,360	89,462
Mali	1987	10	785,384	773,407	582,678	422,837	162,820	76,364	111,641	81,398	48,559	79,558	68,459	42,898
Mali	1998	10	991,330	986,822	734,156	519,001	207,852	102,961	155,752	113,342	69,050	112,397	95,138	60,236
Mali	2009	10	1,451,856	1,424,140	1,262,277	776,333	326,105	158,458	270,954	202,869	120,639	153,897	135,800	86,385
Morocco	1982	5	1,012,873	1,012,873	948,008	571,980	242,307	115,031	234,908	177,902	94,196	197,418	162,099	86,824
Morocco	1994	5	1,294,026	1,293,171	1,293,171	842,330	322,163	149,529	406,223	271,392	136,345	239,441	189,156	101,918
Morocco	2004	5	1,482,720	1,481,076	1,481,076	1,052,531	363,627	161,892	514,724	313,558	150,610	316,587	227,541	116,239
Mozambique	1997	10	1,551,517	1,550,505	1,248,483	879,255	370,427	167,753	199,650	167,263	107,787	119,546	109,546	78,564
Mozambique	2007	10	2,047,048	2,047,048	1,616,853	1,103,596	439,299	193,512	262,286	214,265	133,824	150,307	135,876	96,246
Nigeria	2006	.06	83,700	83,700	82,740	49,282	18,063	8,803	14,170	12,222	7,555	12,444	11,142	6,988
Nigeria	2007	.06	85,183	85,182	84,122	49,102	18,013	8,811	14,465	12,213	7,569	12,319	10,902	6,877
Nigeria	2008	.07	107,425	107,425	105,944	62,151	23,183	11,453	19,914	16,790	10,092	17,700	15,411	9,401
Nigeria	2009	.05	77,896	77,880	77,650	45,988	16,676	8,050	12,872	10,705	6,419	11,577	9,925	6,044
Nigeria	2010	.05	72,191	71,991	58,973	41,830	15,485	7,534	14,118	11,580	6,679	12,629	10,811	6,314
Rwanda	1991	10	742,918	742,918	535,602	372,386	146,839	71,287	112,661	96,102	58,656	74,940	68,798	45,347

Table A.1: Sample construction, continued

(1) country	(2) year	(3) fraction	(4) N_{all}	(5) N_{age}	(6) N_{owned}	(7) $N_{owned}^{age\geq 14}$	(8) $N_{owned}^{14\leq age\leq 25}$	(9) $N_{owned}^{14\leq age\leq 18}$	(10) $N_{olded}^{age\geq 14}$	(11) $N_{olded}^{14\leq age\leq 25}$	(12) $N_{olded}^{14\leq age\leq 18}$	(13) $N_{olded,\ no\ mghh}^{age\geq 14}$	(14) $N_{olded,\ no\ mghh}^{14\leq age\leq 25}$	(15) $N_{olded,\ no\ mghh}^{14\leq age\leq 18}$
Rwanda	2002	10	843,392	843,392	629,146	473,714	221,106	109,367	142,049	128,617	81,595	106,942	101,797	67,823
Rwanda	2012	10	1,038,369	1,038,369	938,201	624,155	250,162	112,248	189,127	162,006	92,149	137,882	125,531	76,187
Senegal	1988	10	700,199	699,981	527,462	378,289	153,541	68,971	103,599	76,483	42,459	78,981	65,979	37,607
Senegal	2002	10	994,562	994,562	911,891	594,599	260,317	124,706	233,001	152,603	82,137	87,659	72,813	42,958
Sierra Leone	2004	10	494,298	492,922	395,788	291,916	120,773	55,346	95,188	67,231	38,567	44,073	36,632	23,137
South Africa	1996	10	3,621,164	3,578,019	3,055,995	2,328,067	840,077	376,601	757,110	537,486	284,595	403,592	326,826	189,176
South Africa	2001	10	3,725,655	3,725,655	3,353,684	2,598,672	915,973	421,066	880,011	603,037	320,148	397,275	325,140	190,487
South Africa	2007	2	1,047,657	1,047,657	842,103	665,305	233,345	105,048	234,464	158,413	80,288	100,958	80,697	45,161
South Africa	2011	8.6	4,418,594	4,418,594	3,845,633	3,101,908	1,020,126	422,182	919,915	608,842	302,412	443,274	338,416	183,186
South Sudan	2008	7	542,765	542,765	542,765	295,979	120,722	57,942	91,414	70,408	41,862	70,186	58,768	36,451
Sudan	2008	16.6	5,066,530	5,066,530	4,055,673	2,919,766	1,238,223	578,339	1,037,575	791,575	462,619	773,891	651,194	400,140
Tanzania	1988	10	2,310,424	2,304,474	1,911,308	1,322,841	556,836	278,218	264,594	228,184	155,786	264,594	228,184	155,786
Tanzania	2002	10	3,732,735	3,732,735	3,123,724	2,190,557	903,114	416,283	494,053	381,631	245,738	317,998	281,324	192,737
Tanzania	2012	10	4,498,022	4,498,022	3,918,823	2,603,099	1,036,707	491,497	665,506	523,475	327,262	372,921	332,966	228,014
Uganda	1991	10	1,548,460	1,547,604	1,242,885	855,537	378,505	179,263	183,439	149,677	97,917	166,998	142,299	94,490
Uganda	2002	10	2,497,449	2,497,449	2,042,838	1,355,857	601,101	289,123	304,094	264,174	183,083	294,850	259,638	180,341
Zambia	1990	10	787,461	787,461	664,239	460,486	216,756	108,294	142,016	120,274	75,070	142,016	120,274	75,070
Zambia	2000	10	996,117	996,117	825,110	570,022	259,096	119,089	192,384	156,274	93,412	110,078	98,159	63,959
Zambia	2010	10	1,321,973	1,321,973	1,028,628	704,471	307,786	147,933	227,855	187,502	117,903	133,646	121,947	83,452
Zimbabwe	2012	5	654,688	653,276	587,010	397,356	157,602	74,305	85,295	64,881	41,073	46,667	40,769	27,976
total			117,279,705	117,142,326	92,685,319	66,561,550	26,476,016	12,194,312	20,269,053	15,572,173	8,960,075	14,149,328	12,127,473	7,352,997

This table shows how we go from the raw IPUMS data to the sample used in our estimates. "fraction" = fraction of the full census obtained by IPUMS; "*age*" = we observe individual age, "*owned*" = we observe the individual's education; "*olded*" = we observe own and parental education, "*no mghh*" we drop individuals in multigenerational households. Observations from three early censuses, Burkina Faso 1985, Kenya 1979, and Liberia 1974 drop, because there are no household identifiers.

Table A.2: Base sample

(1) country	(2) year	(3) $N^{age\geq 14}_{olded,\ no\ mghh}$	(4) $N^{14\leq age\leq 25}_{olded,\ no\ mghh}$	(5) $N^{14\leq age\leq 18}_{olded,\ no\ mghh}$	(6) n_P	(7) n_D	(8) student	(9) urban/rural
Benin	1979	21,762	16,330	9,948	12	77	yes	no
Benin	1992	32,014	28,023	18,515	12	77	yes	yes
Benin	2002	58,448	51,876	33,734	12	77	yes	yes
Benin	2013	105,603	93,532	58,778	12	77	yes	yes
Botswana	1981	5,354	4,380	2,949	21	23	yes	no
Botswana	1991	9,062	7,658	5,246	21	23	yes	yes
Botswana	2001	14,071	11,729	7,224	21	23	yes	no
Botswana	2011	16,566	12,648	7,139	21	23	yes	no
Burkina Faso	1996	102,117	92,338	65,370	13	45	yes	no
Burkina Faso	2006	133,585	123,137	85,097	13	45	yes	yes
Cameroon	1976	58,095	48,566	33,088	7	228	yes	no
Cameroon	1987	73,851	62,728	42,368	7	228	yes	yes
Cameroon	2005	218,037	184,146	112,819	7	228	yes	yes
Egypt	1986	1,275,624	1,064,503	587,721	24	235	yes	yes
Egypt	1996	1,222,560	1,072,794	631,717	24	235	yes	yes
Egypt	2006	1,673,273	1,449,742	709,665	24	235	yes	yes
Ethiopia	1984	318,437	273,988	194,308	12	98	yes	yes
Ethiopia	1994	720,927	656,900	435,645	12	98	yes	yes
Ethiopia	2007	173,103	158,114	108,563	12	98	yes	yes
Ghana	1984	118,235	101,459	65,768	10	110	yes	no
Ghana	2000	180,293	149,419	92,451	10	110	yes	yes
Ghana	2010	279,364	232,961	140,045	10	110	yes	yes
Guinea	1983	44,403	36,885	22,662	6	34	yes	yes
Guinea	1996	53,012	44,454	28,616	6	34	yes	yes
Kenya	1969	42,229	37,861	26,334	8	154	yes	no
Kenya	1989	149,203	129,864	84,770	8	154	yes	yes
Kenya	1999	202,324	175,342	110,568	8	154	yes	yes
Kenya	2009	418,780	376,127	248,115	8	154	yes	yes
Lesotho	1996	29,347	24,226	13,863	11	10	yes	yes
Lesotho	2006	22,139	18,684	10,334	11	10	yes	yes
Liberia	2008	38,266	32,126	19,302	5	47	yes	yes
Malawi	1987	72,558	62,944	41,720	26	227	yes	yes
Malawi	1998	109,301	96,672	64,674	26	227	yes	yes
Malawi	2008	152,144	135,360	89,462	26	227	yes	yes
Mali	1987	79,558	68,459	42,898	9	242	yes	no
Mali	1998	112,397	95,138	60,236	9	242	yes	yes
Mali	2009	153,897	135,800	86,385	9	242	yes	yes
Morocco	1982	197,418	162,099	86,824	16	59	yes	no
Morocco	1994	239,441	189,156	101,918	16	59	yes	no
Morocco	2004	316,587	227,541	116,239	16	59	yes	no
Mozambique	1997	119,546	109,546	78,564	11	144	yes	yes
Mozambique	2007	150,307	135,876	96,246	11	144	yes	yes
Nigeria	2006	12,444	11,142	6,988	38	38	yes	yes
Nigeria	2007	12,319	10,902	6,877	38	38	yes	yes
Nigeria	2008	17,700	15,411	9,401	38	38	yes	yes
Nigeria	2009	11,577	9,925	6,044	38	38	yes	yes
Nigeria	2010	12,629	10,811	6,314	38	38	yes	yes
Rwanda	1991	74,940	68,798	45,347	8	30	yes	no
Rwanda	2002	106,942	101,797	67,823	8	30	yes	yes
Rwanda	2012	137,882	125,531	76,187	8	30	yes	yes
Senegal	1988	78,981	65,979	37,607	9	34	yes	no
Senegal	2002	87,659	72,813	42,958	9	34	yes	yes
Sierra Leone	2004	44,073	36,632	23,137	14	107	yes	yes
South Africa	1996	403,592	326,826	189,176	5	216	yes	yes
South Africa	2001	397,275	325,140	190,487	5	216	yes	yes
South Africa	2007	100,958	80,697	45,161	5	216	yes	yes
South Africa	2011	443,274	338,416	183,186	5	216	yes	yes
South Sudan	2008	70,186	58,768	36,451	10	72	yes	yes
Sudan	2008	773,891	651,194	400,140	15	129	yes	yes
Tanzania	1988	264,594	228,184	155,786	23	113	yes	no
Tanzania	2002	317,998	281,324	192,737	23	113	yes	yes
Tanzania	2012	372,921	332,966	228,014	23	113	yes	yes
Uganda	1991	166,998	142,299	94,490	36	161	yes	yes
Uganda	2002	294,850	259,638	180,341	36	161	yes	yes
Zambia	1990	142,016	120,274	75,070	8	72	yes	yes
Zambia	2000	110,078	98,159	63,959	8	72	yes	yes
Zambia	2010	133,646	121,947	83,452	8	72	yes	no
Zimbabwe	2012	46,667	40,769	27,976	10	88	yes	yes
total		14,149,328	12,127,473	7,352,997				

This table shows the number of observations per census in our final sample for which we observe individual ages and education as well as parental age and parental education and where we have excluded individuals in multigenerational households. "n_P" = number of (admin-1) provinces, "n_D" = number of admin-2/3 districts. "student" = we observe student status, "urban/rural" we observe urban/rural residence.

Table A.3: Observations in full migrant sample

country	year	(1) $N_{nonmig}^{age\geq 14}$	(2) $N_{mig}^{14\leq age\leq 25}$	(3) $N_{mig}^{14\leq age\leq 18}$	(4) n_D
Benin	1979	3,990	2,900	1,718	77
Benin	1992	6,485	5,608	3,578	77
Benin	2002	13,087	11,612	7,396	77
Benin	2013	20,594	17,839	10,574	77
Botswana	2001	4,113	3,521	2,002	21
Botswana	2011	5,831	4,722	2,517	21
Burkina Faso	1996	11,045	9,785	6,385	45
Burkina Faso	2006	11,369	9,959	5,926	45
Cameroon	1976	9,563	8,118	5,462	228
Cameroon	1987	16,062	13,669	9,031	228
Cameroon	2005	50,248	41,997	24,325	228
Egypt	1986	90,095	69,658	36,260	29
Egypt	1996	46,052	36,148	17,309	29
Egypt	2006	131,327	113,044	55,981	29
Ethiopia	1984	14,783	11,240	7,588	87
Ghana	1984	9,580	7,526	4,696	10
Ghana	2000	27,350	21,817	12,620	10
Ghana	2010	39,517	31,961	17,807	10
Guinea	1983	5,192	4,337	2,432	34
Guinea	1996	7,879	6,297	3,505	34
Kenya	1969	8,396	7,684	5,383	158
Kenya	1989	15,132	13,015	8,225	158
Kenya	1999	19,517	16,539	9,995	158
Kenya	2009	68,795	60,924	37,072	158
Liberia	2008	7,255	6,150	3,640	15
Malawi	1987	10,096	8,851	5,879	31
Malawi	2008	58,389	51,386	33,011	31
Mali	1987	6,737	5,559	3,333	47
Mali	1998	9,206	7,342	4,313	47
Mali	2009	17,674	14,976	8,385	47
Morocco	2004	43,116	29,041	13,900	58
Mozambique	1997	18,912	16,926	11,545	139
Mozambique	2007	23,954	20,986	13,675	139
Rwanda	1991	48,600	44,849	29,647	30
Rwanda	2002	5,977	5,610	3,713	30
Rwanda	2012	13,415	12,287	7,927	30
Senegal	1988	18,035	14,699	7,978	34
Senegal	2002	18,795	15,273	8,360	34
Sierra Leone	2004	12,897	10,713	6,502	107
South Africa	1996	18,756	14,938	8,131	9
South Africa	2001	48,987	39,557	21,227	9
South Africa	2007	14,153	11,185	5,655	9
South Africa	2011	56,851	42,882	20,910	9
Sudan	2008	44,197	31,856	15,154	25
Tanzania	1988	26,881	22,949	15,179	30
Tanzania	2002	32,804	27,743	17,173	30
Tanzania	2012	38,435	32,964	20,812	30
Uganda	1991	26,046	21,909	14,438	56
Uganda	2002	27,946	23,890	15,775	56
Zambia	1990	36,685	30,983	18,469	72
Zambia	2000	33,727	29,223	17,326	72
Zambia	2010	34,880	30,858	19,530	72
total		1,389,408	1,155,505	669,374	

This table shows the number of observations for the full sample of individuals for whom parental education is observed and who were born in a different region than their current residence. "n_D" shows the number of districts used for the migration analysis.

Table A.4: Observations in migrant sample with time of migration information

country	year	(1) $N_{nonmig}^{age\geq 14}$	(2) $N_{mig}^{14\leq age\leq 25}$	(3) $N_{mig}^{14\leq age\leq 18}$	(4) n_D
Benin	1979	3,990	2,900	1,718	77
Benin	1992	6,485	5,608	3,578	77
Benin	2002	13,087	11,612	7,396	77
Benin	2013	20,594	17,839	10,574	77
Cameroon	1976	9,563	8,118	5,462	228
Cameroon	1987	16,062	13,669	9,031	228
Cameroon	2005	50,248	41,997	24,325	228
Egypt	1986	90,095	69,658	36,260	29
Egypt	1996	46,052	36,148	17,309	29
Egypt	2006	131,327	113,044	55,981	29
Ethiopia	1984	14,783	11,240	7,588	87
Ghana	2010	39,517	31,961	17,807	10
Guinea	1996	7,879	6,297	3,505	34
Kenya	1999	19,517	16,539	9,995	158
Kenya	2009	68,795	60,924	37,072	158
Malawi	2008	58,389	51,386	33,011	31
Mali	1987	6,737	5,559	3,333	47
Mali	1998	9,206	7,342	4,313	47
Mali	2009	17,674	14,976	8,385	47
Morocco	2004	43,116	29,041	13,900	58
Rwanda	1991	48,600	44,849	29,647	30
Rwanda	2002	5,977	5,610	3,713	30
South Africa	1996	18,756	14,938	8,131	9
South Africa	2001	48,987	39,557	21,227	9
South Africa	2007	14,153	11,185	5,655	9
Sudan	2008	44,197	31,856	15,154	25
Uganda	1991	26,046	21,909	14,438	56
Uganda	2002	27,946	23,890	15,775	56
Zambia	1990	36,685	30,983	18,469	72
Zambia	2000	33,727	29,223	17,326	72
Zambia	2010	34,880	30,858	19,530	72
total		1,013,070	840,716	479,608	

This table shows the number of observations for the full sample of individuals for whom parental education is observed, who were born in a different region than their current residence, and for whom time of migration data are available. "n_D" shows the number of districts used for the migration analysis.

A.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, and far fewer of adults with completed schooling. To maximize coverage, we therefore use the variable "relationship to household head" to identify the educational attainment of the previous generation. This variable takes on 32 different values. We use this classification to assign young individuals to the previous generation. Based on the generation assignment, each individual is assigned the mean education level of individuals within the household of the generation immediately above. For example, 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 A.5: Relationship to household head and generation assignment

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

A.3 Co-residence rates

Table A.6: Co-residence rates

	(1)	(2)	(3)	(4)	(5)
country	age 8	age 14-18	age 14-25	N_{14-18}	N_{14-25}
Benin	99.04	90.50	64.17	120,975	189,761
Botswana	95.77	86.84	72.77	22,558	36,415
Burkina Faso	97.98	84.00	57.83	150,467	215,475
Cameroon	98.67	85.96	64.69	188,275	295,440
Egypt	99.29	98.50	85.33	1,929,103	3,587,039
Ethiopia	99.22	91.36	68.73	738,516	1,089,002
Ghana	98.98	94.35	77.86	298,264	483,839
Guinea	98.97	82.28	60.82	51,278	81,339
Kenya	99.40	93.90	70.74	469,787	719,194
Lesotho	99.66	98.69	86.61	24,197	42,910
Liberia	99.65	95.72	75.00	19,302	32,126
Malawi	99.99	92.85	61.67	195,856	294,976
Mali	99.02	86.20	65.97	189,519	299,397
Morocco	99.89	98.60	91.13	304,981	578,796
Mozambique	99.49	87.78	54.20	174,810	245,422
Nigeria	99.00	93.70	77.15	35,624	58,191
Rwanda	99.43	96.37	76.46	189,357	296,126
Senegal	99.02	92.58	80.68	80,565	138,792
Sierra Leone	95.61	84.69	65.47	23,137	36,632
South Africa	97.52	93.91	83.21	608,010	1,071,079
South Sudan	97.49	92.52	75.82	36,451	58,768
Sudan	99.52	92.93	73.45	400,140	651,194
Tanzania	99.84	95.18	68.78	576,537	842,474
Uganda	98.50	89.08	63.05	274,831	401,937
Zambia	98.89	90.54	68.17	222,481	340,380
Zimbabwe	98.30	89.97	60.60	27,976	40,769
overall	99.10	93.73	74.78	7,352,997	12,127,473

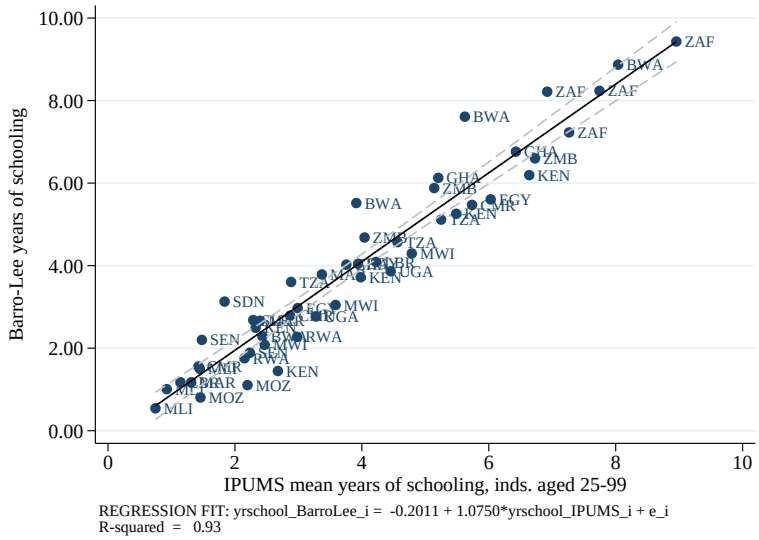
This table shows co-residence rates for individuals aged 8, individuals aged 14-18 and invidiuals aged 14-25. To compute co-residence rates, we start with a sample of individuals for whom their own eduactional attainment as well as relationship to household head is observed. The latter does not exclude single-person households, since these individuals will be labelled "head". The co-residence rate is then simply the total number of individuals that co-reside with at least one member of an immediately older generation in the household divided by the total number of individuals in that age group.

A.4 Barro-Lee crosscheck

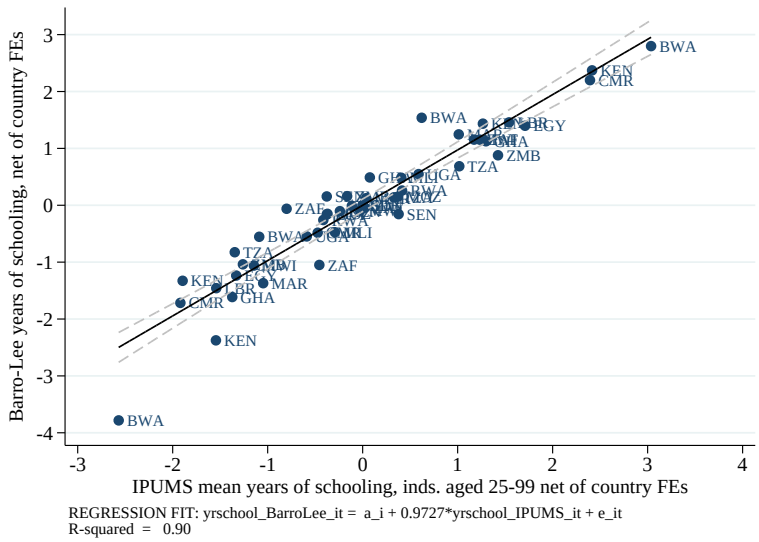
To gauge the quality of our education data, we correlate mean years of schooling for individuals aged 25-99 to the data from Barro and Lee (2013) who also report figures for years of schooling for this age range As Barro and Lee (2013) only report their estimates of years of schooling at 5-year intervals, we correlate our estimates with the closest years they report. Barro and Lee (2013) provide two separate estimates for years for schooling – one based on an age range of 15-99, the other 25-99. Strictly for this comparison only, we compute measures for the 25-99 age range. Since we have several countries with more than one census, we can also explore the panel-correlation with Barro and Lee. There are six countries for which we only have one census: Sierra-Leone, Egypt, Rwanda, South Sudan, Sudan, and Zimbabwe.

Figure A.1: Barro-Lee crosscheck

(a) Years of schooling in our sample compared to Barro and Lee (2011), levels, full sample of countries



(b) Years of schooling in our sample compared to Barro and Lee (2011), controlling for country fixed effects, full sample of countries



A.5 Schooling by cohort

In this appendix section, we summarize education levels by country and for an average invidiual of the 26 countries for four birth-decades since 1950 using data on individuals 25 and above.

Table A.7: Average education by country-cohort

country	birth-cohort	(1) mean years of schooling	(2) share less than primary	(3) share primary	(4) share secondary	(5) share tertiary
Benin	1950	1.951	0.812	0.128	0.031	0.014
Benin	1960	2.757	0.740	0.186	0.045	0.013
Benin	1970	2.902	0.738	0.199	0.042	0.015
Benin	1980	3.930	0.679	0.197	0.097	0.027
Botswana	1950	4.670	0.531	0.343	0.055	0.039
Botswana	1960	6.714	0.333	0.495	0.099	0.066
Botswana	1970	9.011	0.149	0.564	0.189	0.088
Botswana	1980	10.477	0.079	0.556	0.264	0.092
Burkina Faso	1950	0.664	0.824	0.050	0.008	0.008
Burkina Faso	1960	0.842	0.837	0.063	0.017	0.012
Burkina Faso	1970	1.306	0.813	0.099	0.036	0.014
Burkina Faso	1980	1.430	0.797	0.127	0.043	0.007
Cameroon	1950	4.591	0.499	0.412	0.044	0.011
Cameroon	1960	5.765	0.401	0.475	0.076	0.021
Cameroon	1970	6.888	0.326	0.501	0.088	0.061
Cameroon	1980	6.213	0.398	0.415	0.139	0.020
Egypt	1950	4.091	0.676	0.054	0.169	0.100
Egypt	1960	5.531	0.563	0.058	0.265	0.114
Egypt	1970	7.471	0.403	0.084	0.367	0.145
Egypt	1980	8.559	0.324	0.076	0.421	0.178
Ethiopia	1950	1.106	0.707	0.032	0.018	0.002
Ethiopia	1960	2.039	0.512	0.058	0.031	0.002
Ethiopia	1970	2.186	0.146	0.020	0.010	0.001
Ethiopia	1980	2.397	0.143	0.024	0.010	0.001
Ghana	1950	5.985	0.485	0.385	0.115	0.015
Ghana	1960	6.361	0.461	0.397	0.122	0.020
Ghana	1970	6.249	0.434	0.389	0.158	0.020
Ghana	1980	7.086	0.357	0.389	0.231	0.023
Guinea	1950	1.780	0.843	0.061	0.058	0.021
Guinea	1960	1.736	0.819	0.105	0.030	0.012
Guinea	1970	1.476	0.830	0.100	0.026	0.004
Kenya	1950	4.961	0.515	0.311	0.150	0.013
Kenya	1960	6.721	0.333	0.388	0.248	0.021
Kenya	1970	7.669	0.241	0.457	0.271	0.023
Kenya	1980	7.897	0.224	0.453	0.290	0.022
Lesotho	1950	5.022	0.643	0.283	0.058	0.016
Lesotho	1960	6.243	0.468	0.415	0.100	0.016
Lesotho	1970	6.970	0.372	0.470	0.139	0.018
Lesotho	1980	7.352	0.341	0.472	0.173	0.014
Liberia	1950	4.009	0.663	0.133	0.166	0.038
Liberia	1960	4.687	0.593	0.185	0.190	0.032
Liberia	1970	4.665	0.583	0.228	0.171	0.017
Liberia	1980	4.804	0.558	0.282	0.153	0.007
Malawi	1950	3.435	0.787	0.163	0.040	0.004
Malawi	1960	4.275	0.727	0.196	0.063	0.006
Malawi	1970	5.234	0.653	0.207	0.113	0.008
Malawi	1980	6.228	0.560	0.258	0.140	0.007
Mali	1950	1.297	0.855	0.080	0.022	0.011
Mali	1960	1.438	0.846	0.094	0.017	0.010
Mali	1970	1.573	0.856	0.088	0.020	0.014
Mali	1980	2.111	0.825	0.110	0.036	0.022
Morocco	1950	2.774	0.796	0.109	0.058	0.021
Morocco	1960	3.846	0.730	0.122	0.107	0.042

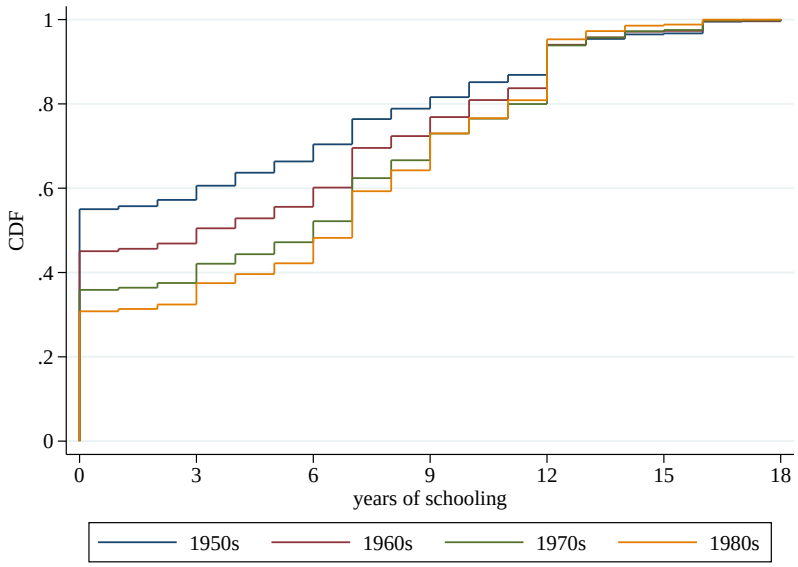
Table A.7: Average education by country-cohort, continued

country	birth-cohort	(1) mean years of schooling	(2) share less than primary	(3) share primary	(4) share secondary	(5) share tertiary
Morocco	1970	4.826	0.633	0.215	0.107	0.045
Mozambique	1950	1.539	0.912	0.058	0.012	0.004
Mozambique	1960	2.158	0.859	0.100	0.020	0.005
Mozambique	1970	2.385	0.833	0.122	0.023	0.004
Mozambique	1980	2.852	0.780	0.163	0.033	0.003
Nigeria	1950	3.970	0.573	0.218	0.138	0.050
Nigeria	1960	5.017	0.485	0.225	0.216	0.050
Nigeria	1970	5.663	0.431	0.231	0.266	0.047
Nigeria	1980	6.559	0.369	0.215	0.330	0.052
Rwanda	1950	2.435	0.772	0.174	0.007	0.003
Rwanda	1960	3.404	0.653	0.277	0.017	0.007
Rwanda	1970	4.727	0.526	0.383	0.047	0.017
Rwanda	1980	4.745	0.642	0.235	0.084	0.027
Senegal	1950	2.281	0.766	0.165	0.040	0.016
Senegal	1960	2.444	0.752	0.184	0.044	0.012
Senegal	1970	2.756	0.724	0.210	0.056	0.010
Sierra Leone	1950	2.176	0.768	0.165	0.023	0.020
Sierra Leone	1960	2.441	0.739	0.207	0.018	0.015
Sierra Leone	1970	2.714	0.718	0.239	0.016	0.013
South Africa	1950	6.765	0.347	0.417	0.161	0.042
South Africa	1960	8.128	0.232	0.446	0.244	0.047
South Africa	1970	9.463	0.135	0.424	0.358	0.054
South Africa	1980	10.359	0.066	0.423	0.414	0.062
South Sudan	1950	0.753	0.931	0.045	0.013	0.011
South Sudan	1960	1.129	0.889	0.078	0.021	0.012
South Sudan	1970	1.206	0.878	0.093	0.020	0.009
South Sudan	1980	1.473	0.850	0.117	0.024	0.008
Sudan	1950	1.295	0.895	0.069	0.010	0.023
Sudan	1960	1.736	0.843	0.113	0.010	0.031
Sudan	1970	2.039	0.818	0.125	0.010	0.044
Sudan	1980	2.273	0.801	0.134	0.010	0.050
Tanzania	1950	3.896	0.601	0.329	0.062	0.008
Tanzania	1960	5.299	0.349	0.572	0.067	0.011
Tanzania	1970	5.900	0.266	0.641	0.075	0.017
Tanzania	1980	6.148	0.272	0.589	0.097	0.042
Uganda	1950	4.026	0.631	0.326	0.032	0.010
Uganda	1960	4.651	0.570	0.371	0.048	0.009
Uganda	1970	5.352	0.517	0.395	0.077	0.011
Zambia	1950	5.394	0.460	0.382	0.118	0.024
Zambia	1960	5.968	0.386	0.437	0.132	0.024
Zambia	1970	6.576	0.332	0.446	0.171	0.016
Zambia	1980	7.200	0.285	0.424	0.220	0.008
Zimbabwe	1950	5.600	0.515	0.413	0.040	0.021
Zimbabwe	1960	8.091	0.272	0.590	0.087	0.044
Zimbabwe	1970	9.613	0.100	0.768	0.097	0.031
Zimbabwe	1980	9.803	0.084	0.786	0.096	0.030

This table shows average education by country and birth-decade for ages 25+. Column (1) shows mean years of schooling, column (2) the share of individuals with less than primary education, column (3) the share with primary education, column (4) the share with secondary education, and column (5) the share with tertiary education.

For a rough idea of the overall evolution of schooling, figure A.2 plots the CDF of years of schooling for four birth-decades since the 1950s for our full dataset, again restricting individuals to ages 25+. Note that these data represent unweighted averages across all available censuses for individuals born in each birth-decade.

Figure A.2: CDF of years of schooling by birth-decade



A.6 Measurement error

In this subsection we give an example with simulated data showing that measurement error in education may lead to a mechanical association between parental literacy and IM when assessing them in terms of years of schooling. We then show that this correlation disappears when both are defined on the basis of completed primary and measurement error is in years of schooling.

Case 1: Mobility and literacy defined directly, measurement error affects literacy.

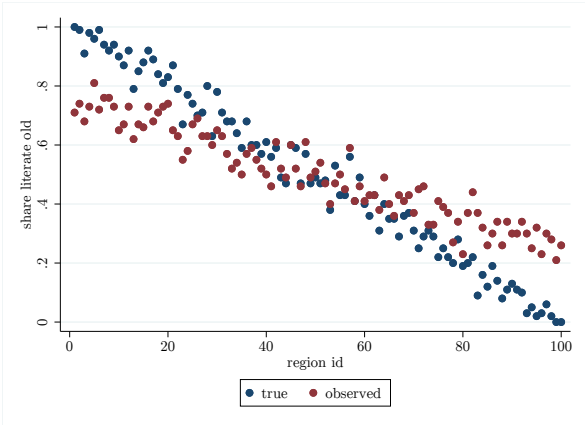
Consider a country with $R = 100$ regions. Each region $r \in \{1, \ldots, 100\}$ is populated by $n_r = 100$ old individuals who each have one child. Old literacy is determined by the value of a random variable $u \sim \mathcal{U}[0, 1]$ drawn independently across all individuals in all regions. Individuals in region r are literate if $u \geq \frac{r}{100}$.

Old literacy is observed with error e^o drawn i.i.d from

$$e^o = \begin{cases} -1 & \text{w.p. } \frac{1}{4} \\ 0 & \text{w.p. } \frac{1}{2} \\ 1 & \text{w.p. } \frac{1}{4}. \end{cases}$$

Observed and true literacy are related as $\text{lit}^{\text{o,obs}} = \min\left[1, \max\left[0, \text{lit}^{\text{o,true}} + e^o\right]\right]$. Figure A.3 shows the distribution of average true and observed parental literacy by region.

Figure A.3: Distribution of true and observed parental literacy across regions

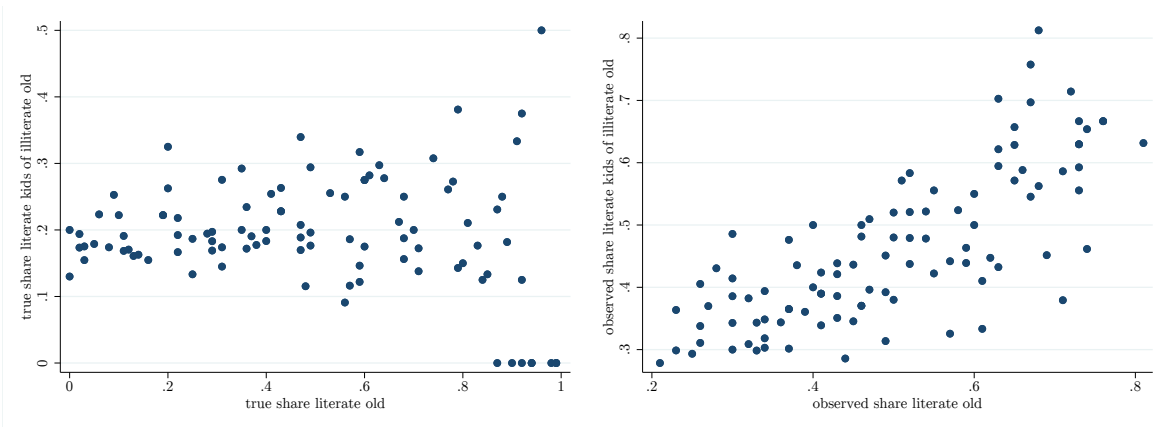


Child literacy is determined by a random variable $v \sim \mathcal{U}[0,1]$ drawn independently across all individuals in all regions. Children of truly literate parents have a 0.8 chance of being literate and children of truly illiterate parents have a 0.2 chance of being literate. As with the old, literacy of the young is observed with an i.i.d error e^y, which has the same distribution as e^o and true and observed literacy of the young are related as $\text{lit}^{\text{y,obs}} = \min\left[1, \max\left[0, \text{lit}^{\text{y,true}} + e^y\right]\right]$.

Given this data generating process, figure A.4 shows the true and observed relationship between parental literacy and IM across regions.

Figure A.4: True and observed literacy and IM across regions

(a) True **(b)** Observed



The true relationship shows heteroskedasticity (when there are fewer illiterate old, the estimates of mobility are noisier because they are based on fewer observations) but no slope. By contrast, measurement error introduces a clear positive relationship between the two variables in the observed data.

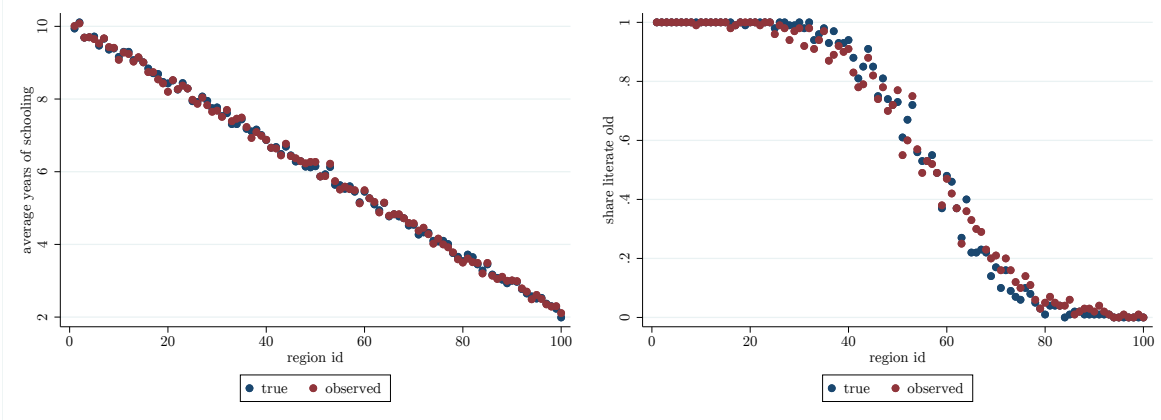
Case 2: Mobility and literacy defined in terms of primary schooling, measurement error affects schooling.

Once again there are 100 regions, with 100 old-young pairs per region. The old in region r receive a draw from a random variable $u \sim \mathcal{N}(\mu_r, 1)$, where $\mu_r = 2 + (100 - r + 1)\frac{8}{100}$. This ensures that $\mu_1 \approx 10$ and $\mu_{100} = 2$. True schooling is defined as $\text{ysc}^{\text{o,true}} = \min\left[12, \max\left[0, u\right]\right]$. We then draw the same e^o as in case 1 and compute observed schooling of the old as $\text{ysc}^{\text{o,obs}} = \min\left[12, \max\left[0, \text{ysc}^{\text{o,true}} + e^o\right]\right]$. Parental literacy is defined in

terms of years of schooling: 6 or more years make people literate. Figure A.5 shows the distribution of average true and observed parental schooling and literacy by region.

Figure A.5: True and observed parental schooling and literacy across regions

(a) Schooling **(b)** Literacy



Child schooling is related to parental schooling through a transition matrix M (representative element $m_{i,j}$, which we define in three steps:

1) set $m_{i,i} = 0.5$ (the likelihood that children have the same education as their parents)

2) $\forall i \neq j$, set $m_{i,j} = \frac{1}{|i-j|+1}$. This means that if parents have 5 years of schooling, the (un-normalized, see step 3) likelihood that their children have 4 or 6 years is 0.5. As j moves further from i, the likelihood declines.

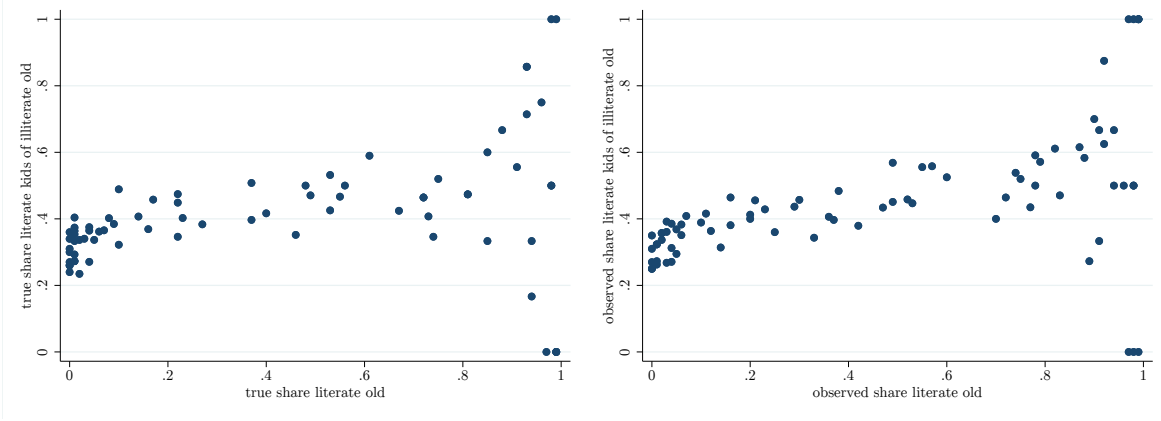
3) normalize by rows so that the transition probabilities for every parent sum to 1.

Using M, we compute true schooling of children. We then add an error e^y that has the same distribution as in case 1 and compute $\text{ysc}^{\text{y,obs}} = \min\left[12, \max\left[0, \text{ysc}^{\text{y,true}} + e^y\right]\right]$. Finally, we use the same 6-year cutoff to determine literacy as we did for parents.

Given this data generating process, figure A.6 shows the true and observed relationship between parental literacy and IM across regions.

Figure A.6: True and observed literacy and IM across regions

(a) True **(b)** Observed



As expected, there is a small positive slope between literacy and IM at the lower end of the literacy distribution (where the error has some bite) but overall, measurement error has no effect on our main relationship.

B IM across countries and regions

Appendix Section B gives additional summary and descriptive statistics on intergenerational mobility (IM) in educational attainment across African countries and regions, as well as further evidence on the association between educational IM and the literacy of the "old" generation.

Section B.1 gives further evidence on IM at the country level.
Table B.1 reports country-level estimates of upward IM and downward IM, conditional on census-year and cohort fixed-effects. This Table "mirrors" Table 1 in the paper, but accounts for birth-cohort factors and trends and census-year factors.
Table B.2 reports country-level estimates of upward IM and downward IM for rural and urban households, using the census classification for individuals aged 14-18 and aged 14-25. This Table "mirrors" Table 1 in the paper, but distinguishes individuals by rural-urban status.
Table B.3 reports country-level estimates of upward IM and downward IM for boys and girls aged 14-18 and aged 14-25. This Table "mirrors" Table 1 in the paper, but distinguishes children by gender.
Table B.4 reports OLS estimates associating country-cohort upward IM and downward IM with cohort dummy variables for the 1970s, 1980s, and 1990s (with the 1960s serving as the omitted category). This allows examining trends in IM.
Table B.5 reports OLS estimates associating country-cohort upward IM and downward IM with cohort indicator variables looking only at countries with full cohort coverage. This allows examining trends in IM.
Figure B.1 plots the country-specific upward and downward IM estimates for children aged 14-25, born in the 1960s, the 1970s, the 1980s, and the 1990s. The figure distinguishes countries with full cohort coverage and those without. This figure "mirrors" Figure 2 in the paper that looked at children aged 14-18. age group).

Section B.2 gives further evidence on IM at the region level.
Table B.6 reports OLS estimates associating region-cohort upward IM and downward IM with cohort dummy variables for the 1970s, 1980s, and 1990s (with the 1960s serving as the omitted category). This allows examining trends in regional IM.
Table B.7 reports OLS estimates associating region-cohort upward IM and downward IM with cohort indicator variables looking only at countries with full cohort coverage. This allows examining trends in regional IM.
Figure B.2 plots district-level upward IM for the 1990s-born cohort (in the vertical axis) against upward-IM for the 1960s-born cohort (in the horizontal axis). Panel A reports the unconditional scatter-plot; Panel B nets country factors (conditioning on country fixed-effects)

Figure B.2 plots district-level downward IM for the 1990s-born cohort (in the vertical axis) against downward-IM for the 1960s-born cohort (in the horizontal axis). Panel A reports the unconditional scatter-plot; Panel B nets country factors (conditioning on country fixed-effects)

B.1 Country-level IM

Table B.1: Country-level estimates of IM conditional on census-year and cohort effects

	(1)	(2)	(3)	(4)	(5)	(6)
mobility / N	upward	upward	downward	downward	N with e_0 obs.	N with e_0 obs.
age range	14-18	14-25	14-18	14-25	14-18	14-25
Botswana	0.113	0.139	-0.006	-0.009	22,558	36,415
South Africa	0.096	0.128	0.012	0.007	608,010	1,071,079
Tanzania	0.068	0.113	0.110	0.092	576,537	842,474
Zimbabwe	0.059	0.136	0.132	0.098	27,976	40,769
Egypt	0.000	0.000	0.000	0.000	1,929,103	3,587,039
Uganda	-0.151	-0.097	0.150	0.124	274,831	401,937
Nigeria	-0.152	-0.066	0.064	0.029	35,624	58,191
Benin	-0.161	-0.153	0.092	0.105	120,975	189,761
Ghana	-0.183	-0.110	0.083	0.037	298,264	483,839
Cameroon	-0.189	-0.109	0.109	0.069	188,275	295,440
Senegal	-0.218	-0.211	0.029	0.045	80,565	138,792
Lesotho	-0.250	-0.196	0.208	0.169	24,197	42,910
Rwanda	-0.255	-0.189	0.416	0.339	189,357	296,126
Zambia	-0.290	-0.189	0.134	0.081	222,481	340,380
Morocco	-0.320	-0.279	0.154	0.152	304,981	578,796
Kenya	-0.380	-0.274	0.145	0.080	469,787	719,194
Guinea	-0.408	-0.399	0.239	0.235	51,278	81,339
Mali	-0.530	-0.477	0.156	0.126	189,519	299,397
Burkina Faso	-0.539	-0.509	0.135	0.118	150,467	215,475
Liberia	-0.554	-0.453	0.498	0.369	19,302	32,126
Malawi	-0.567	-0.455	0.411	0.288	195,856	294,976
Sierra Leone	-0.622	-0.537	0.399	0.349	23,137	36,632
Ethiopia	-0.633	-0.550	0.255	0.205	738,516	1,089,002
Sudan	-0.662	-0.585	0.355	0.222	400,140	651,194
Mozambique	-0.684	-0.569	0.361	0.266	174,810	245,422
South Sudan	-0.731	-0.677	0.732	0.596	36,451	58,768
mean / total	-0.329	-0.268	0.226	0.177	7,389,448	12,186,241

Columns (1) and (2) give upward-IM estimates. They reflect the likelihood that children, aged 14-18 and 14-25, whose parents have not completed primary schooling will manage to complete at least primary education. Columns (3) and (4) give downward-IM estimates. They reflect the likelihood that children, aged 14-18 and 14-25, whose parents have completed primary schooling or higher will not manage to complete primary education. All estimates conditional on census-year and birth-cohort (of old and young) fixed effects. Columns (5) and (6) give the number of observations used to estimate the country-specific IM statistics (children whose parental education is reported in the censuses). Countries are sorted from the highest to the lowest level of upward IM in the 14-18 sample (column (1)). "mean" gives the simple average of the 26 country-estimates.

Table B.2: Country-level estimates of IM, urban/rural heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sample	urban				rural			
mobility	upward		downward		upward		downward	
age range	14-18	14-25	14-18	14-25	14-18	14-25	14-18	14-25
Zimbabwe	0.839	0.888	0.057	0.042	0.611	0.676	0.198	0.168
South Africa	0.839	0.859	0.044	0.033	0.762	0.779	0.096	0.079
Botswana	0.765	0.767	0.101	0.092	0.603	0.596	0.177	0.166
Nigeria	0.756	0.793	0.047	0.035	0.620	0.667	0.096	0.074
Cameroon	0.722	0.740	0.064	0.059	0.464	0.478	0.210	0.193
Ghana	0.709	0.722	0.101	0.082	0.511	0.513	0.193	0.173
Zambia	0.698	0.700	0.133	0.133	0.403	0.415	0.354	0.349
Tanzania	0.692	0.725	0.108	0.089	0.567	0.600	0.211	0.187
Egypt	0.614	0.638	0.049	0.040	0.674	0.671	0.073	0.071
Kenya	0.587	0.646	0.126	0.093	0.443	0.514	0.243	0.194
Lesotho	0.581	0.652	0.165	0.124	0.417	0.464	0.310	0.263
Ethiopia	0.573	0.631	0.154	0.126	0.066	0.085	0.622	0.598
Benin	0.557	0.572	0.149	0.139	0.350	0.351	0.224	0.222
Uganda	0.546	0.593	0.166	0.133	0.357	0.400	0.321	0.276
Burkina Faso	0.524	0.546	0.171	0.150	0.115	0.120	0.458	0.460
Sierra Leone	0.475	0.505	0.251	0.217	0.158	0.169	0.572	0.523
Guinea	0.458	0.498	0.281	0.260	0.128	0.124	0.473	0.471
Senegal	0.456	0.476	0.151	0.140	0.164	0.163	0.336	0.336
Mali	0.439	0.453	0.178	0.167	0.144	0.137	0.315	0.301
Sudan	0.355	0.489	0.319	0.200	0.079	0.108	0.447	0.355
Rwanda	0.327	0.416	0.278	0.208	0.205	0.280	0.570	0.474
Malawi	0.322	0.431	0.335	0.238	0.144	0.215	0.531	0.427
Liberia	0.310	0.404	0.477	0.362	0.172	0.237	0.612	0.514
Mozambique	0.209	0.290	0.397	0.305	0.058	0.079	0.632	0.569
South Sudan	0.071	0.128	0.723	0.577	0.033	0.054	0.765	0.671
mean	0.519	0.565	0.221	0.178	0.318	0.344	0.377	0.338

This table shows estimates of IM (likelihood that children of illiterate parents become literate, higher numbers → higher upward IM; likelihood that children of parents with at least primary, who complete less than primary, higher numbers → higher downward IM). Columns (1)-(4) show estimates for individuals residing in urban areas, columns (5)-(8) those for individuals in rural areas. "age range" indicates the range of ages for children in the sample. Countries sorted by column (1). "mean" is the mean of the country-level estimates.

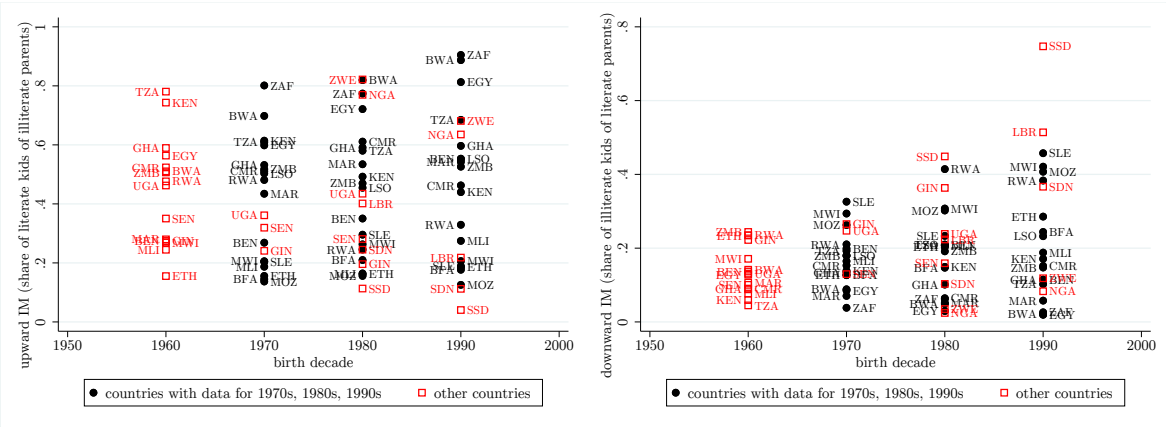
Table B.3: Country-level estimates of IM, male/female heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sample	male				female			
mobility	upward		downward		upward		downward	
age range	14-18	14-25	14-18	14-25	14-18	14-25	14-18	14-25
South Africa	0.758	0.792	0.085	0.063	0.832	0.844	0.049	0.038
Egypt	0.669	0.661	0.063	0.057	0.606	0.593	0.065	0.061
Botswana	0.658	0.673	0.084	0.071	0.763	0.774	0.054	0.047
Zimbabwe	0.634	0.722	0.170	0.126	0.706	0.762	0.117	0.092
Nigeria	0.632	0.680	0.079	0.059	0.670	0.712	0.079	0.061
Ghana	0.600	0.610	0.135	0.110	0.556	0.538	0.146	0.131
Tanzania	0.578	0.635	0.196	0.162	0.634	0.663	0.142	0.124
Cameroon	0.542	0.554	0.112	0.101	0.499	0.507	0.101	0.096
Zambia	0.499	0.538	0.203	0.173	0.486	0.493	0.180	0.167
Morocco	0.493	0.470	0.068	0.071	0.356	0.336	0.079	0.089
Kenya	0.450	0.536	0.231	0.172	0.479	0.537	0.180	0.140
Benin	0.444	0.433	0.137	0.129	0.340	0.315	0.225	0.217
Uganda	0.379	0.440	0.303	0.247	0.349	0.364	0.262	0.229
Lesotho	0.343	0.392	0.356	0.297	0.538	0.615	0.210	0.168
Senegal	0.309	0.310	0.172	0.163	0.246	0.254	0.220	0.203
Sierra Leone	0.295	0.325	0.297	0.250	0.230	0.226	0.353	0.309
Rwanda	0.288	0.355	0.495	0.402	0.301	0.356	0.443	0.363
Guinea	0.268	0.267	0.271	0.255	0.159	0.167	0.381	0.360
Mali	0.247	0.241	0.172	0.168	0.181	0.182	0.230	0.217
Liberia	0.230	0.335	0.512	0.380	0.221	0.278	0.522	0.419
Burkina Faso	0.184	0.190	0.163	0.156	0.164	0.182	0.234	0.205
Malawi	0.157	0.254	0.471	0.347	0.161	0.207	0.418	0.333
Ethiopia	0.133	0.163	0.220	0.179	0.132	0.154	0.224	0.192
Mozambique	0.121	0.186	0.475	0.358	0.106	0.143	0.447	0.363
Sudan	0.111	0.165	0.398	0.273	0.138	0.201	0.335	0.226
South Sudan	0.046	0.086	0.737	0.616	0.037	0.060	0.768	0.654
mean	0.375	0.411	0.272	0.222	0.368	0.390	0.268	0.228

This table shows estimates of IM (likelihood that children of illiterate parents become literate, higher numbers → higher upward IM; likelihood that children of parents with at least primary, who complete less than primary, higher numbers → higher downward IM). Columns (1)-(4) show estimates for male individuals, columns (5)-(8) those for female individuals. "age range" indicates the range of ages for children in the sample. Countries sorted by column (1). "mean" is the mean of the country-level estimates.

Figure B.1: IM at the country-birth-decade level, ages 14-25

(a) upward **(b)** downward



The figures report upward (panel A) and downward (panel B) Intragenerational Mobility in educational attainment (IM) across decade birth cohorts for children aged 14-25. Black solid circles indicate countries with data covering the 1970s, the 1980s, and the 1990s. Red hollow squares indicate countries with data covering just some cohorts.

Table B.4: IM at the country × cohort level, trends

	(1) IM up	(2) IM up	(3) IM up	(4) IM down	(5) IM down	(6) IM down
born 1970s	0.0505	0.0554	0.168	0.0374	0.0130	0.0339
	(1.17)	(1.57)	(1.15)	(1.59)	(0.38)	(0.47)
born 1980s	0.0512	0.0523	0.133	0.0697**	0.0458	0.107
	(1.21)	(1.22)	(0.92)	(2.33)	(1.32)	(1.50)
born 1990s	0.0847	0.108**	0.194	0.0852**	0.0248	0.0471
	(1.54)	(2.40)	(1.35)	(2.13)	(0.75)	(0.66)
R2	0.014	0.899		0.040	0.789	
within R2	0.014	0.176		0.040	0.067	
N	75	71	75	75	71	75
estimator	OLS	country FE	QREG	OLS	country FE	QREG

This table shows regressions of average IM in country c for individuals born in birth-cohort b on cohort dummies. 1960s is the omitted category. t-statistics based on standard errors clustered at the country-level (except for quantile regression) in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table B.5: IM at the country × cohort level, trends, balanced sample

	(1) IM up	(2) IM up	(3) IM up	(4) IM down	(5) IM down	(6) IM down
born 1970s	0.0589	0.0647	0.206	0.0433*	0.0125	0.0317
	(1.21)	(1.66)	(1.46)	(1.76)	(0.32)	(0.60)
born 1980s	0.0456	0.0514	0.157	0.0830**	0.0523	0.107**
	(0.98)	(1.11)	(1.11)	(2.26)	(1.36)	(2.02)
born 1990s	0.119**	0.125**	0.230	0.0462	0.0154	0.0284
	(2.14)	(2.66)	(1.63)	(1.44)	(0.44)	(0.53)
R2	0.032	0.900		0.043	0.787	
within R2	0.032	0.242		0.043	0.094	
N	60	60	60	60	60	60
estimator	OLS	country FE	QREG	OLS	country FE	QREG

This table shows regressions of average IM in country c for individuals born in birth-cohort b on cohort dummies. 1960s is the omitted category. t-statistics based on standard errors clustered at the country-level (except for quantile regression) in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

B.2 Region-level IM

Table B.6: IM at the district × cohort level, trends

	(1) IM up	(2) IM up	(3) IM up	(4) IM down	(5) IM down	(6) IM down
born 1970s	0.0440	0.0124	0.0105	0.0604	0.0485	0.0387
	(0.84)	(0.44)	(0.36)	(1.46)	(0.93)	(0.76)
born 1980s	0.0854*	0.0539	0.0489	0.0847*	0.0544	0.0350
	(1.94)	(1.16)	(0.98)	(1.81)	(0.89)	(0.58)
born 1990s	0.128**	0.117***	0.111**	0.0892*	0.0290	0.0105
	(2.23)	(3.07)	(2.64)	(1.79)	(0.51)	(0.19)
R2	0.024	0.701	0.910	0.011	0.424	0.665
within R2	0.024	0.067	0.182	0.011	0.007	0.009
N	8031	8031	7551	7289	7289	6738
FEs	none	country	district	none	country	district

This table shows regressions of average IM in district d in country c for individuals born in birth-cohort b on cohort dummies. 1960s is the omitted category. t-statistics based on standard errors clustered at the country-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

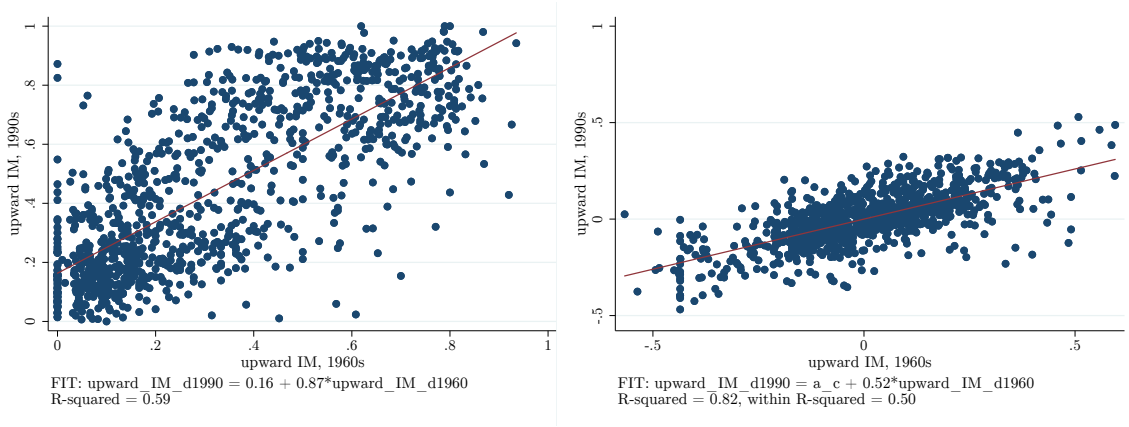
Table B.7: IM at the district × cohort level, trends, balanced sample

	(1) IM up	(2) IM up	(3) IM up	(4) IM down	(5) IM down	(6) IM down
born 1970s	0.0264	0.00499	0.00502	0.0814*	0.0610	0.0461
	(0.82)	(0.18)	(0.18)	(1.87)	(1.10)	(0.85)
born 1980s	0.0674	0.0461	0.0463	0.0926	0.0692	0.0415
	(1.52)	(0.83)	(0.83)	(1.52)	(0.97)	(0.62)
born 1990s	0.140***	0.118**	0.118**	0.0480	0.0216	-0.00621
	(3.39)	(2.68)	(2.67)	(0.96)	(0.35)	(-0.11)
R2	0.036	0.634	0.884	0.017	0.316	0.610
within R2	0.036	0.075	0.203	0.017	0.016	0.023
N	5032	5032	5032	4432	4432	4432
FEs	none	country	district	none	country	district

This table shows regressions of average IM in district d in country c for individuals born in birth-cohort b on cohort dummies. 1960s is the omitted category. t-statistics based on standard errors clustered at the country-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Figure B.2: Persistence of district-level upward IM over time (60s and 90s birth-decades)

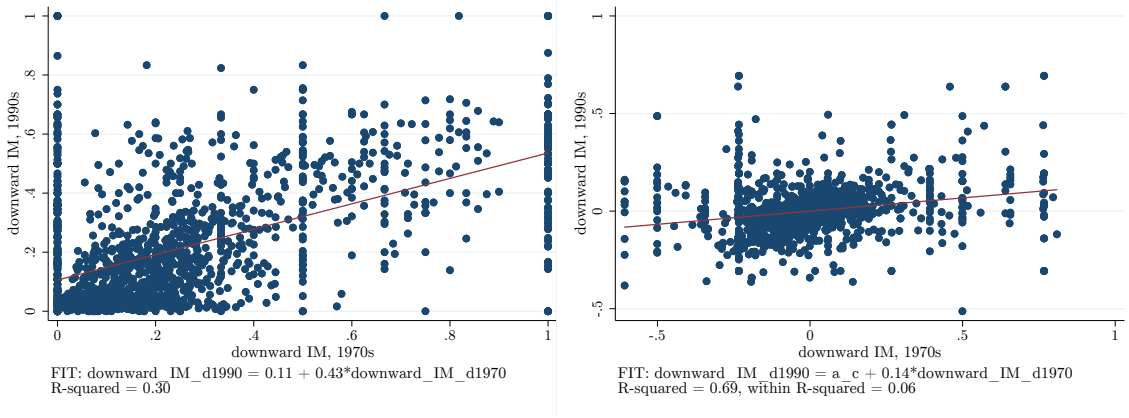
(a) OLS **(b)** country fixed effects



These figures visualize two regressions that link district-level upward IM in the 90s to district-level upward IM in the 60s. Panel (a) shows the simple linear regression, panel (b) shows the regression with country fixed effects.

Figure B.3: Persistence of district-level downward IM over time (70s and 90s birth-decades)

(a) OLS **(b)** country fixed effects



These figures visualize two regressions that link district-level downward IM in the 90s to district-level downward IM in the 70s. Panel (a) shows the simple linear regression, panel (b) shows the regression with country fixed effects.

C Correlates

Appendix Section C gives definitions and sources for all variables employed in the correlation analysis (Section 4 of the paper) and additional regression results.

Section C.1 gives variable definitions and sources for the geographic, locational, and ecological, historical (colonial and precolonial) and at-independence economic factors employed in Section 4 of the paper.
Table C.1 reports summary statistics.

Section C.2 presents sensitivity checks of the association between literacy, upward IM, and downward IM and the variable indicated in the left column. Panel A looks at geographic features. Panel B looks at colonial and pre-colonial features. Panel C looks at at-independence economic features. The tables give standardized "beta" coefficients. All tables "mirror" Table 6 that reports the baseline correlates at the sample of individuals aged 14-18, conditional on country fixed-effects.
Table C.2 reports the regional level correlation analysis looking at individuals ages 14-25.
Table C.3 reports the regional level correlation analysis looking at individuals ages 14-18, conditioning on province (admin-1) fixed-effects.
Table C.4 reports the regional level correlation analysis looking at individuals ages 14-18, conditioning neither on country fixed effects nor on province fixed-effects.
Table C.5 reports the regional level correlation analysis looking at individuals aged 14-18 and born in the 1990s.

Section C.3 reports multivariate regression estimates associating regional upward IM with the various geographic, historical, and economic characteristics of the regions. While the correlational analysis does not aim to identify causal effects, it allows understanding the strength of the univariate correlations, as the various correlates are related to each other.
Table C.6 gives OLS regression country fixed-effects estimates associating regional upward IM with geographic, historical, and at independence factors. The table gives standardized beta coefficients and clustered at the province-level standard errors.
Table C.7 gives OLS regression country fixed-effects estimates associating regional downward IM with geographic, historical, and at independence factors. The table gives standardized beta coefficients and clustered at the province-level standard errors.

C.1 Variable definitions

ln(distance to the capital) The natural logarithm of the geodesic distance from the district centroid to the national capital. Computed using GIS software.
ln(distance to the border) The natural logarithm of the geodesic distance from the district centroid to clostest point on the national border. Computed using GIS software.
ln(distance to the coast) The natural logarithm of the geodesic distance from the district centroid to clostest point on the coastline. Computed using GIS software.
ln(1+malaria stability) The natural logarithm of 1 + mean stability of malaria transmission in the district. The latter variable is computed, using GIS software, as the within-

district zonal statistic of a raster provided by Kiszewski et al. (2004), which we resample to a resolution of 30 arc-seconds prior to computing the statistic.

ln(1+agricultural suitability) The natural logarithm of 1 + mean agricultural suitability in the district. The latter variable is computed, using GIS software, as the within-district zonal statistic of a raster provided by Ramankutty et al. (2002), which we resample to a resolution of 30 arc-seconds prior to computing the statistic.

ln(terrain ruggedness) The natural logarithm of terrain ruggedness. The latter is computed using cell-level data on elevation at 30 arc-second resolution from Survey (1996). Given the grid cell data, picture a 3×3 block of 9 cells and let $e_{r,c}$ be the elevation of the cell in row r, column c of the grid. Following Nunn and Puga (2012), we compute ruggedness as $\sqrt{\sum_{i=r-1^{r+1}} \sum_{i=c-1}^{c+1} (e_{i,j} - e_{r,c})^2}$, that is, the square root of the sum of all the squared differences in elevation between the middle cell and the surrounding 8 cells.

oil field dummy A dummy = 1 if the district is intersected by an oil field, and zero otherwise. Data on oil fields come from Lujala et al. (2007)

diamond mine dummy A dummy = 1 if the district is intersected by a diamond mine, and zero otherwise. Data on oil fields come from Lujala et al. (2005)

ln(population density 1950 The natural logarithm of mean population density in the district in 1950. The latter variable is computed, using GIS software, as the within-district zonal statistic of a raster provided by Klein Goldewijk et al. (2010), which we resample to a resolution of 30 arc-seconds prior to computing the statistic.

ln(distance to railroad) The natural logarithm of the geodesic distance from the district centroid to clostest point on a colonial railroad. Computed using GIS software. Data on colonial railroads come from Jedwab and Moradi (2016).

ln(distance to road) The natural logarithm of the geodesic distance from the district centroid to clostest point on a colonial road. Computed using GIS software. Data on colonial roads come from Jedwab and Storeygard (2018).

ln(distance to Catholic mission) The natural logarithm of the geodesic distance from the district centroid to the closest Catholic Mission. Computed using GIS software. Data on missions come from Nunn (2010).

ln(distance to Protestant mission) The natural logarithm of the geodesic distance from the district centroid to the closest Protestant Mission. Computed using GIS software. Data on missions come from Nunn (2010) and Cagé and Rueda (2016).

ln(distance to precolonon. empire) The natural logarithm of the geodesic distance from the district centroid to the closest pre-colonial empire. Computed using GIS software. Data on the extent of pre-colonial empires come from Besley and Reynal-Querol (2014).

ln(distance to precolonon. state) The natural logarithm of the geodesic distance from the district centroid to the closest pre-colonial state. Data on the extent of pre-colonial states are obtained by combining the maps of pre-colonial ethnic homeland in Murdock (1959) with the levels of jurisdictional hierarchy beyond the local community level of these societies, a variable found in Murdock (1967). Societies with more than 3 levels are classified as states.

urban share (born $<$ 1960**)** The share of the (non-migrant) district population born prior to 1960 classified as urban. Computed using the IPUMS census data.

agri. empl share (born $<$ 1960**)** The share of the (non-migrant) district population born prior to 1960 and working in agriculture. Computed using the IPUMS census data.
manuf. empl share (born $<$ 1960**)** The share of the (non-migrant) district population born prior to 1960 and working in manufacturing. Computed using the IPUMS census data.
serv. empl share (born $<$ 1960**)** The share of the (non-migrant) district population born prior to 1960 and working in services. Computed using the IPUMS census data.

Table C.1: Summary statistics for correlates

Variable	Obs	Mean	Std. Dev.	Min	Max
Geography					
ln(distance to the capital)	2,813	5.357	1.152	-0.573	7.528
ln(distance to the border)	2,813	3.932	1.153	-3.304	6.269
ln(distance to the coast)	2,813	5.548	1.422	-2.010	7.450
ln(1+malaria stability)	2,809	2.041	1.160	0	3.652
ln(1+agricultural suitability)	2,779	0.296	0.184	0	0.692
ln(terrain ruggedness)	2,810	3.696	1.199	0.438	6.224
oil field dummy	2,796	0.049	0.216	0	1
diamond mine dummy	2,796	0.038	0.190	0	1
History					
ln(distance to railroad)	2,310	4.041	1.664	-3.730	6.984
ln(distance to road)	2,526	2.540	1.690	-6.250	6.521
ln(distance to Catholic mission)	2,813	5.219	1.414	-0.468	7.798
ln(distance to Protestant mission)	2,813	3.949	1.436	-1.471	7.042
ln(distance to precolonon. empire)	2,813	5.901	0.762	1.532	7.426
ln(distance to precolonon. state)	2,813	4.896	0.833	-0.471	6.926
Contemporary					
ln(population density 1950)	2,808	2.890	1.870	-10.597	9.751
urban share (born $<$ 1960)	2,531	0.245	0.294	0	1
agri. empl share (born $<$ 1960)	2,430	0.678	0.303	0	1
manuf. empl share (born $<$ 1960)	2,430	0.044	0.060	0	0.555
serv. empl share (born $<$ 1960)	2,430	0.270	0.263	0	1

C.2 Additional results: One variable at a time

Table C.2: District-level correlates of IM, country fixed effects, all birth cohorts, ages 14-25

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM controlling for share literate old	N	IM	IM controlling for share literate old	N
Panel A: geography							
ln(distance to capital)	-0.302***	-0.293***	-0.097***	2809	0.229***	0.094***	2787
	(0.038)	(0.039)	(0.026)		(0.032)	(0.025)	
ln(distance to border)	0.050	0.015	-0.021*	2809	-0.034	-0.008	2787
	(0.036)	(0.032)	(0.012)		(0.027)	(0.015)	
ln(distance to coast)	-0.200***	-0.230***	-0.092***	2809	0.167***	0.069***	2787
	(0.057)	(0.051)	(0.018)		(0.040)	(0.017)	
ln(1+malaria stability)	-0.242***	-0.252***	-0.085***	2798	0.177***	0.060*	2776
	(0.049)	(0.053)	(0.027)		(0.043)	(0.033)	
ln(1+agricultural suitability)	-0.034	0.010	0.034*	2768	-0.004	-0.021	2746
	(0.056)	(0.049)	(0.019)		(0.038)	(0.027)	
ln(terrain ruggedness)	0.103**	0.113***	0.041**	2799	-0.094***	-0.043*	2777
	(0.048)	(0.039)	(0.019)		(0.036)	(0.023)	
oil field dummy	0.013	0.009	-0.001	2784	-0.008	-0.001	2762
	(0.027)	(0.024)	(0.010)		(0.024)	(0.019)	
diamond mine dummy	-0.012	-0.013	-0.004	2784	0.029**	0.022*	2762
	(0.013)	(0.008)	(0.007)		(0.013)	(0.011)	
Panel B: history							
ln(distance to railroad)	-0.320***	-0.334***	-0.086***	2299	0.246***	0.068***	2277
	(0.041)	(0.041)	(0.021)		(0.027)	(0.018)	
ln(distance to road)	-0.273***	-0.255***	-0.052***	2515	0.220***	0.080***	2493
	(0.028)	(0.029)	(0.017)		(0.021)	(0.021)	
ln(distance to cath. mission)	-0.374***	-0.325***	-0.067***	2809	0.246***	0.063**	2787
	(0.060)	(0.056)	(0.025)		(0.044)	(0.029)	
ln(distance to prot. mission)	-0.364***	-0.326***	-0.082***	2809	0.252***	0.083***	2787
	(0.046)	(0.039)	(0.019)		(0.032)	(0.023)	
ln(distance to precolon. empire)	0.025	-0.023	-0.041*	2809	0.012	0.025	2787
	(0.041)	(0.029)	(0.024)		(0.035)	(0.034)	
ln(distance to precolon. state)	-0.038	-0.059*	-0.033**	2809	0.042	0.023	2787
	(0.040)	(0.032)	(0.016)		(0.027)	(0.020)	
Panel C: contemporary							
ln(population density 1950)	0.236***	0.233***	0.073***	2797	-0.157***	-0.043**	2775
	(0.041)	(0.039)	(0.019)		(0.027)	(0.019)	
urban share (born < 1960)	0.392***	0.252***	-0.014	2531	-0.241***	-0.086***	2513
	(0.022)	(0.028)	(0.021)		(0.022)	(0.020)	
agri. empl. share (born < 1960)	-0.619***	-0.443***	-0.090**	2430	0.336***	0.094***	2412
	(0.030)	(0.025)	(0.039)		(0.026)	(0.027)	
manuf. empl. share (born < 1960)	0.243***	0.156***	0.002	2430	-0.113***	0.001	2412
	(0.042)	(0.034)	(0.015)		(0.028)	(0.017)	
serv. empl. share (born < 1960)	0.601***	0.440***	0.100**	2430	-0.335***	-0.105***	2412
	(0.032)	(0.026)	(0.039)		(0.024)	(0.025)	

This is not a normal regression table. In the column entitled "share literate old" the dependent variable is the district 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-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (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. 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 C.3: District-level correlates of IM, province fixed effects, all birth cohorts, ages 14-18

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM controlling for share literate old	N	IM	IM controlling for share literate old	N
Panel A: geography							
ln(distance to capital)	-0.303***	-0.274***	-0.105***	2749	0.197***	0.070**	2727
	(0.067)	(0.039)	(0.038)		(0.046)	(0.030)	
ln(distance to border)	0.009	-0.000	-0.005	2749	0.010	0.014	2727
	(0.024)	(0.018)	(0.012)		(0.020)	(0.016)	
ln(distance to coast)	-0.167***	-0.159***	-0.063**	2749	0.147***	0.076***	2727
	(0.036)	(0.031)	(0.025)		(0.032)	(0.027)	
ln(1+malaria stability)	-0.253***	-0.165***	-0.018	2738	0.097***	-0.014	2716
	(0.060)	(0.062)	(0.031)		(0.028)	(0.024)	
ln(1+agricultural suitability)	0.060	0.102**	0.067***	2708	-0.042	-0.016	2686
	(0.046)	(0.040)	(0.024)		(0.036)	(0.035)	
ln(terrain ruggedness)	0.024	0.024	0.010	2739	-0.029	-0.019	2717
	(0.025)	(0.025)	(0.022)		(0.025)	(0.023)	
oil field dummy	-0.012	-0.040***	-0.033**	2724	0.006	0.000	2702
	(0.018)	(0.012)	(0.014)		(0.019)	(0.020)	
diamond mine dummy	-0.017*	-0.014**	-0.003	2724	0.022*	0.014	2702
	(0.010)	(0.006)	(0.009)		(0.012)	(0.012)	
Panel B: history							
ln(distance to railroad)	-0.241***	-0.232***	-0.079***	2239	0.200***	0.072***	2217
	(0.029)	(0.024)	(0.015)		(0.024)	(0.026)	
ln(distance to road)	-0.211***	-0.178***	-0.060***	2455	0.160***	0.064***	2433
	(0.023)	(0.019)	(0.009)		(0.018)	(0.018)	
ln(distance to cath. mission)	-0.335***	-0.252***	-0.061**	2749	0.219***	0.078**	2727
	(0.042)	(0.032)	(0.024)		(0.031)	(0.032)	
ln(distance to prot. mission)	-0.288***	-0.210***	-0.048***	2749	0.149***	0.027	2727
	(0.038)	(0.029)	(0.017)		(0.020)	(0.020)	
ln(distance to precolon. empire)	0.065	-0.015	-0.053	2749	0.007	0.035	2727
	(0.056)	(0.038)	(0.057)		(0.038)	(0.048)	
ln(distance to precolon. state)	-0.035	-0.037	-0.017	2749	0.009	-0.006	2727
	(0.032)	(0.023)	(0.016)		(0.021)	(0.019)	
Panel C: contemporary							
ln(population density 1950)	0.208***	0.214***	0.100***	2737	-0.133***	-0.044**	2715
	(0.051)	(0.041)	(0.029)		(0.029)	(0.022)	
urban share (born < 1960)	0.325***	0.173***	-0.003	2531	-0.211***	-0.119***	2513
	(0.021)	(0.027)	(0.020)		(0.025)	(0.021)	
agri. empl. share (born < 1960)	-0.571***	-0.365***	-0.151***	2409	0.311***	0.160***	2391
	(0.034)	(0.024)	(0.034)		(0.028)	(0.028)	
manuf. empl. share (born < 1960)	0.187***	0.113***	0.022	2409	-0.116***	-0.045**	2391
	(0.051)	(0.027)	(0.016)		(0.026)	(0.018)	
serv. empl. share (born < 1960)	0.540***	0.353***	0.149***	2409	-0.293***	-0.144***	2391
	(0.036)	(0.025)	(0.031)		(0.028)	(0.024)	

This is not a normal regression table. In the column entitled "share literate old" the dependent variable is the district 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-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (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). Nigeria and Botswana are omitted. Coefficients are standardized. 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 C.4: District-level correlates of IM, without fixed effects, all birth cohorts, ages 14-18

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM controlling for share literate old	N	IM	IM controlling for share literate old	N
Panel A: geography							
ln(distance to capital)	-0.219***	-0.262***	-0.090**	2809	0.234***	0.120**	2787
	(0.059)	(0.066)	(0.046)		(0.062)	(0.051)	
ln(distance to border)	0.080	0.018	-0.046*	2809	-0.079*	-0.033	2787
	(0.055)	(0.055)	(0.028)		(0.044)	(0.033)	
ln(distance to coast)	-0.153***	-0.333***	-0.215***	2809	0.274***	0.195***	2787
	(0.054)	(0.054)	(0.034)		(0.052)	(0.042)	
ln(1+malaria stability)	-0.363***	-0.372***	-0.092**	2798	0.375***	0.203***	2776
	(0.058)	(0.054)	(0.036)		(0.044)	(0.040)	
ln(1+agricultural suitability)	-0.044	-0.151**	-0.116***	2768	0.156**	0.134***	2746
	(0.065)	(0.072)	(0.037)		(0.064)	(0.049)	
ln(terrain ruggedness)	0.234***	0.163**	-0.026	2799	-0.104*	0.026	2777
	(0.048)	(0.070)	(0.041)		(0.063)	(0.051)	
oil field dummy	0.003	0.099**	0.097***	2784	-0.085*	-0.084**	2762
	(0.038)	(0.046)	(0.026)		(0.048)	(0.041)	
diamond mine dummy	0.035	-0.004	-0.032***	2784	-0.002	0.016	2762
	(0.033)	(0.032)	(0.012)		(0.036)	(0.022)	
Panel B: history							
ln(distance to railroad)	-0.379***	-0.371***	-0.058	2299	0.258***	0.048	2277
	(0.047)	(0.056)	(0.041)		(0.043)	(0.036)	
ln(distance to road)	-0.341***	-0.372***	-0.093***	2515	0.327***	0.137***	2493
	(0.036)	(0.038)	(0.026)		(0.039)	(0.030)	
ln(distance to cath. mission)	-0.154**	0.063	0.192***	2809	-0.212***	-0.304***	2787
	(0.066)	(0.063)	(0.031)		(0.054)	(0.032)	
ln(distance to prot. mission)	-0.488***	-0.461***	-0.090**	2809	0.292***	0.032	2787
	(0.035)	(0.042)	(0.037)		(0.042)	(0.044)	
ln(distance to precolon. empire)	0.020	0.059	0.043	2809	-0.068	-0.055	2787
	(0.063)	(0.064)	(0.038)		(0.055)	(0.043)	
ln(distance to precolon. state)	-0.167***	-0.189***	-0.055	2809	0.164***	0.074**	2787
	(0.058)	(0.063)	(0.035)		(0.054)	(0.038)	
Panel C: contemporary							
ln(population density 1950)	0.163***	0.317***	0.191***	2797	-0.282***	-0.198***	2775
	(0.047)	(0.047)	(0.028)		(0.047)	(0.039)	
urban share (born < 1960)	0.451***	0.386***	0.023	2531	-0.394***	-0.161***	2513
	(0.038)	(0.041)	(0.038)		(0.034)	(0.037)	
agri. empl. share (born < 1960)	-0.479***	-0.475***	-0.123**	2430	0.457***	0.238***	2412
	(0.051)	(0.046)	(0.048)		(0.036)	(0.036)	
manuf. empl. share (born < 1960)	0.290***	0.369***	0.152***	2430	-0.375***	-0.225***	2412
	(0.063)	(0.060)	(0.034)		(0.046)	(0.028)	
serv. empl. share (born < 1960)	0.457***	0.444***	0.104**	2430	-0.426***	-0.208***	2412
	(0.053)	(0.051)	(0.046)		(0.042)	(0.038)	

This is not a normal regression table. In the column entitled "share literate old" the dependent variable is the district 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-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (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 without fixed effects. Coefficients are standardized. 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 C.5: District-level correlates of IM, country fixed effects, 1990s birth cohort, ages 14-25

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM controlling for share literate old	N	IM	IM controlling for share literate old	N
Panel A: geography							
ln(distance to capital)	-0.322***	-0.221***	-0.067**	2557	0.190***	0.080***	2506
	(0.045)	(0.038)	(0.027)		(0.032)	(0.026)	
ln(distance to border)	0.051	0.007	-0.020	2557	-0.014	0.006	2506
	(0.042)	(0.027)	(0.012)		(0.026)	(0.016)	
ln(distance to coast)	-0.201***	-0.187***	-0.087***	2557	0.156***	0.083***	2506
	(0.065)	(0.041)	(0.017)		(0.035)	(0.021)	
ln(1+malaria stability)	-0.259***	-0.209***	-0.081***	2546	0.194***	0.100***	2495
	(0.058)	(0.045)	(0.025)		(0.043)	(0.035)	
ln(1+agricultural suitability)	-0.029	0.037	0.052***	2517	-0.008	-0.020	2466
	(0.068)	(0.043)	(0.019)		(0.042)	(0.031)	
ln(terrain ruggedness)	0.130**	0.091***	0.025	2547	-0.054*	-0.004	2496
	(0.055)	(0.034)	(0.020)		(0.030)	(0.020)	
oil field dummy	0.012	0.013	0.007	2533	-0.004	0.000	2483
	(0.030)	(0.019)	(0.010)		(0.023)	(0.018)	
diamond mine dummy	-0.010	-0.012*	-0.006	2533	-0.001	-0.005	2483
	(0.012)	(0.007)	(0.006)		(0.011)	(0.010)	
Panel B: history							
ln(distance to railroad)	-0.362***	-0.323***	-0.093***	2047	0.228***	0.074***	1998
	(0.053)	(0.039)	(0.019)		(0.027)	(0.021)	
ln(distance to road)	-0.280***	-0.210***	-0.050***	2263	0.196***	0.082***	2214
	(0.032)	(0.026)	(0.013)		(0.023)	(0.021)	
ln(distance to cath. mission)	-0.345***	-0.244***	-0.071***	2557	0.182***	0.052*	2506
	(0.073)	(0.049)	(0.024)		(0.040)	(0.027)	
ln(distance to prot. mission)	-0.366***	-0.264***	-0.088***	2557	0.212***	0.084***	2506
	(0.055)	(0.034)	(0.019)		(0.029)	(0.021)	
ln(distance to precolon. empire)	0.041	-0.016	-0.038*	2557	0.000	0.016	2506
	(0.040)	(0.024)	(0.021)		(0.037)	(0.036)	
ln(distance to precolon. state)	-0.042	-0.052*	-0.031*	2557	0.025	0.009	2506
	(0.043)	(0.028)	(0.016)		(0.026)	(0.020)	
Panel C: contemporary							
ln(population density 1950)	0.221***	0.172***	0.063***	2545	-0.098***	-0.013	2494
	(0.047)	(0.033)	(0.018)		(0.025)	(0.021)	
urban share (born < 1960)	0.358***	0.190***	0.021	2282	-0.214***	-0.110***	2236
	(0.026)	(0.024)	(0.021)		(0.023)	(0.021)	
agri. empl. share (born < 1960)	-0.597***	-0.312***	-0.069**	2207	0.292***	0.127***	2159
	(0.032)	(0.024)	(0.033)		(0.026)	(0.029)	
manuf. empl. share (born < 1960)	0.231***	0.082***	-0.026*	2207	-0.096***	-0.010	2159
	(0.041)	(0.029)	(0.015)		(0.026)	(0.015)	
serv. empl. share (born < 1960)	0.576***	0.318***	0.091***	2207	-0.289***	-0.131***	2159
	(0.035)	(0.022)	(0.029)		(0.024)	(0.027)	

This is not a normal regression table. In the column entitled "share literate old" the dependent variable is the district share of parents with at least primary schooling (the simple average for parents of children born in the 1990s). In the columns entitled "IM" it is the district-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (the simple average for children born in the 1990s), 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 (the simple average for parents of children born in the 1990s), – 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. 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.

C.3 Multivariate correlates regressions

Table C.6: Multivariate regression of district-level upward IM on convariates, by category and kitchen-sink, ages 14-18, regressions conditional on country-fixed effects

	(1) geography	(2) geography	(3) history	(4) history	(5) contemporary	(6) contemporary	(7) kitchen-sink	(8) kitchen-sink
parental literacy		0.629***		0.699***		0.626***		0.607***
		(0.037)		(0.040)		(0.063)		(0.045)
oil field dummy	0.0191	-0.000759					0.0129	0.00591
	(0.014)	(0.010)					(0.010)	(0.008)
diamond mine dummy	0.00163	0.00505					0.00717	0.0108*
	(0.008)	(0.007)					(0.008)	(0.006)
DCAP	-0.224***	-0.0805***					-0.0589**	-0.0124
	(0.029)	(0.026)					(0.030)	(0.027)
ln(distance to border)	0.0816***	0.0243**					0.118***	0.0338***
	(0.023)	(0.012)					(0.023)	(0.013)
ln(distance to coast)	-0.208***	-0.0985***					-0.150***	-0.0823***
	(0.044)	(0.022)					(0.046)	(0.027)
ln(1+agricultural suitability)	0.00249	0.0305*					-0.0354	0.0182
	(0.028)	(0.017)					(0.022)	(0.017)
ln(1+malaria stability)	-0.146***	-0.0472*					-0.0775**	-0.0686***
	(0.041)	(0.027)					(0.035)	(0.026)
ln(terrain ruggedness)	0.0914***	0.0373**					0.111***	0.0524***
	(0.023)	(0.019)					(0.017)	(0.019)
ln(population density 1950)			0.0660**	0.0336*			0.0444	0.0493**
			(0.029)	(0.018)			(0.031)	(0.020)
ln(distance to railroad)			-0.161***	-0.0656***			-0.0871***	-0.0643***
			(0.036)	(0.020)			(0.022)	(0.016)
ln(distance to road)			-0.100***	-0.0150			-0.0209	-0.00225
			(0.024)	(0.016)			(0.016)	(0.012)
ln(distance to cath. mission)			-0.0611**	-0.0123			-0.0214	0.00624
			(0.031)	(0.017)			(0.026)	(0.018)
ln(distance to prot. mission)			-0.167***	-0.0280			-0.0785***	-0.0262
			(0.029)	(0.023)			(0.028)	(0.023)
ln(distance to precolon. empire)			-0.0246	0.00130			0.0550**	0.00250
			(0.029)	(0.020)			(0.025)	(0.019)
ln(distance to precolon. state)			-0.0338	-0.0210			-0.0268	-0.0104
			(0.024)	(0.016)			(0.022)	(0.015)
urban share (born < 1960)					0.0393	-0.0649***	0.0747***	-0.00602
					(0.024)	(0.020)	(0.023)	(0.024)
migrant share (born < 1960)					0.0161	-0.0129	-0.0244	-0.0302**
					(0.024)	(0.017)	(0.018)	(0.015)
agri. empl. share (born < 1960)					-0.240***	-0.0415	-0.280***	-0.131*
					(0.081)	(0.112)	(0.098)	(0.075)
manuf. empl. share (born < 1960)					-0.0116	-0.0177	-0.0743***	-0.0544***
					(0.027)	(0.025)	(0.021)	(0.016)
serv. empl. share (born < 1960)					0.156**	0.0938	-0.110	-0.113
					(0.069)	(0.112)	(0.093)	(0.072)
R2	0.803	0.897	0.805	0.901	0.823	0.898	0.860	0.914
within-R2	0.315	0.643	0.377	0.685	0.326	0.613	0.524	0.709
N	2747	2747	2288	2288	2350	2350	1867	1867

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). parental literacy = district-level share of literate parents (also estimated net of fixed effects). Coefficients are standardized. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table C.7: Multivariate regression of district-level downward IM on convariates, by category and kitchen-sink, ages 14-18, regressions conditional on country-fixed effects

	(1) geography	(2) geography	(3) history	(4) history	(5) contemporary	(6) contemporary	(7) kitchen-sink	(8) kitchen-sink
parental literacy		-0.386***		-0.445***		-0.336***		-0.320***
		(0.032)		(0.041)		(0.047)		(0.041)
oil field dummy	-0.00529	0.00686					-0.00316	0.000321
	(0.015)	(0.015)					(0.021)	(0.021)
diamond mine dummy	0.0125	0.0102					0.0341**	0.0321**
	(0.011)	(0.011)					(0.015)	(0.016)
DCAP	0.167***	0.0781***					0.0824***	0.0578*
	(0.022)	(0.022)					(0.026)	(0.030)
ln(distance to border)	-0.0750***	-0.0396**					-0.0804***	-0.0360*
	(0.021)	(0.017)					(0.023)	(0.019)
ln(distance to coast)	0.130***	0.0624***					0.0467	0.0107
	(0.030)	(0.020)					(0.036)	(0.029)
ln(1+agricultural suitability)	-0.0113	-0.0282					-0.0174	-0.0456**
	(0.029)	(0.027)					(0.022)	(0.023)
ln(1+malaria stability)	0.0967***	0.0360					0.0378	0.0333
	(0.037)	(0.036)					(0.040)	(0.038)
ln(terrain ruggedness)	-0.0632***	-0.0303					-0.0847***	-0.0543***
	(0.021)	(0.019)					(0.021)	(0.020)
ln(population density 1950)			-0.0209	-0.000147			-0.00510	-0.00763
			(0.025)	(0.023)			(0.031)	(0.027)
ln(distance to railroad)			0.118***	0.0570***			0.0524***	0.0404**
			(0.023)	(0.016)			(0.018)	(0.019)
ln(distance to road)			0.117***	0.0629***			0.0479**	0.0381*
			(0.024)	(0.022)			(0.023)	(0.023)
ln(distance to cath. mission)			0.00968	-0.0216			-0.0243	-0.0389*
			(0.025)	(0.019)			(0.023)	(0.021)
ln(distance to prot. mission)			0.148***	0.0599**			0.0994***	0.0719***
			(0.029)	(0.027)			(0.024)	(0.024)
ln(distance to precolon. empire)			0.0145	-0.00184			-0.0320	-0.00413
			(0.037)	(0.034)			(0.037)	(0.038)
ln(distance to precolon. state)			0.0195	0.0108			0.00716	-0.00176
			(0.021)	(0.020)			(0.019)	(0.020)
urban share (born < 1960)					-0.113***	-0.0572***	-0.132***	-0.0895***
					(0.016)	(0.020)	(0.026)	(0.028)
migrant share (born < 1960)					-0.0429**	-0.0276	-0.0147	-0.0118
					(0.018)	(0.017)	(0.020)	(0.021)
agri. empl. share (born < 1960)					0.0422	-0.0651	0.163*	0.0852
					(0.060)	(0.066)	(0.096)	(0.103)
manuf. empl. share (born < 1960)					0.00254	0.00679	0.0190	0.00948
					(0.020)	(0.018)	(0.031)	(0.032)
serv. empl. share (born < 1960)					-0.157***	-0.124**	0.0974	0.0997
					(0.056)	(0.062)	(0.088)	(0.094)
R2	0.670	0.706	0.601	0.640	0.686	0.708	0.622	0.637
within-R2	0.122	0.218	0.149	0.233	0.154	0.213	0.221	0.252
N	2725	2725	2266	2266	2332	2332	1849	1849

The dependent variable is the country-level share of illiterate kids of literate parents (estimated net of census year and old and young birth decade fixed effects). parental literacy = district-level share of literate parents (also estimated net of fixed effects). Coefficients are standardized. Standard errors clustered at the province-level in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

D Correlation between schooling and other outcomes with DHS and Afrobarometer

Appendix Section D reports graphical and regression evidence of a positive correlation between years of schoolings and various "good" outcomes using data using data from the Demographic and Health Surveys (DHS) and the Afrobarometer Surveys.

Section D.1 examines the association between years of schooling and DHS-based proxies of well-being, health, and public goods provision.
The DHS correlations are based on a sample of 3,516,848 individuals, drawn from 155 surveys in 41 countries. DHS provides a sub-national region identifier for all surveys (856 units). For 2,823,745 observations from 118 surveys, DHS reports geolocation information. This allows us to assign respondents to admin-1 and admin-2 administrative regions, 516 and 3,552 respectively.

Tables D.1-D.6 and Figures E.1-E.6 report the correlation analysis of years of schooling with:

i. DHS composite wealth index quintiles. Table D.1 and Figure D.1

ii. Child mortality. Table D.2 and Figure D.2

iii. A female bargaining power index. Table D.3 and Figure D.3

iv. An index capturing attitudes towards domestic violence. Table D.4 and Figure D.4

v. Fertility. Table D.5 and Figure D.5

vi. Age of first marriage. Table D.6 and Figure D.6

Tables D.1-D.6 have six columns.
Column (1) shows the unconditional correlation between the variable of interest and schooling.
Column (2) shows the correlation conditional on individual controls: age, age-squared, dummies for male individuals, male household head, urban residence, log of number of household members, and birth-decade dummies.
Column (3) adds to the set of controls in column (2) survey (= country$\times$year) constants.
Column (4) adds to the set of controls in column (3) DHS region fixed-effects.
Columns (5) and (6) restrict the analysis to geo-referenced observations. Column (5) reports DHS province (admin-1 unit) fixed-effects estimates.
Column (6) reports regional (admin-2 unit) fixed-effects estimates.

Figures D.1-D.6 visualize the corresponding correlations through binned scatter plots.

The DHS analysis shows that educational attainment and mean years of schooling correlate significantly with proxies of household wealth (positively), child mortality and fertility (negatively). The DHS analysis further shows that education correlates strongly with proxies of women empowerment.

Section D.2 examines the association between years of schooling and proxies of well-being, from the Afrobarometer Surveys.
The Afrobarometer correlations exploit information from rounds 3, 4 and 5 and cover 104,004 respondents, residing in 523 regions in 34 countries.

Tables D.7-D.11 and Figures D.7-D.11 report the correlation analysis of years of schooling with:

i. A living conditions index. Table D.7 and Figure D.7

ii. An index capturing how often the respondent goes without food. Table D.8 and Figure D.8

iii. An index capturing how often the respondent goes without food. Table D.9 and Figure D.9

iv. Interest in public affairs. Table D.10 and Figure D.10

v. Support for democracy. Table D.11 and Figure D.11

Tables D.7-D.11 have four columns.
Column (1) shows the unconditional correlation between the variable of interest and years of schooling.
Column (2) shows the correlation conditional on individual controls: age, age-squared, dummies for male individuals, urban residence, and birth-decade dummies.
Column (3) adds to the set of controls in column (2) survey (country-year) fixed-effects.
Column (4) adds to the set of controls in column (4) Afrobarometer region fixed-effects.

Figures D.7-D.11 visualize the corresponding correlations through binned scatter plots.

The analysis of the Afrobarometer Surveys shows a strong positive correlation between education and living conditions (positive) and measures of deprivation (negative). The Afrobarometer Surveys analysis further shows that education correlates strongly with proxies of political participation and support for democracy.

D.1 DHS

D.1.1 Proxies of wellbeing

Household wealth

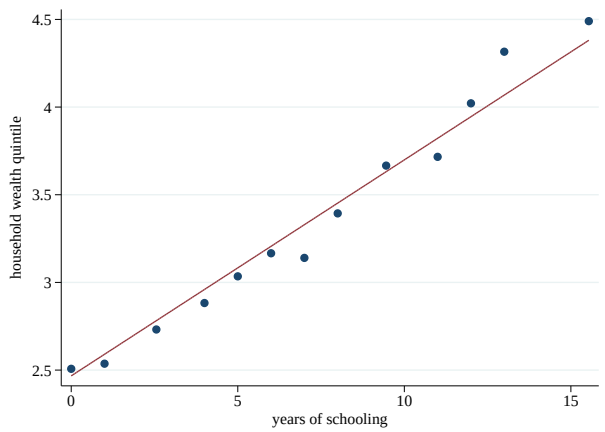
Table D.1: Household wealth quintile and years of schooling

	(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

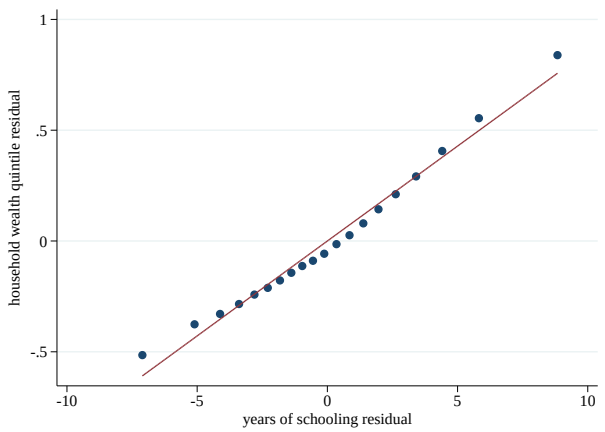
This table shows regression results of household wealth on years of schooling for individuals aged 18+. The dependent variable in all columns is the DHS household wealth quintile (computed for each survey, i.e. country-year) separately based on the DHS-computed wealth index). Individual controls are 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. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure D.1: Binned scatter plots, household wealth

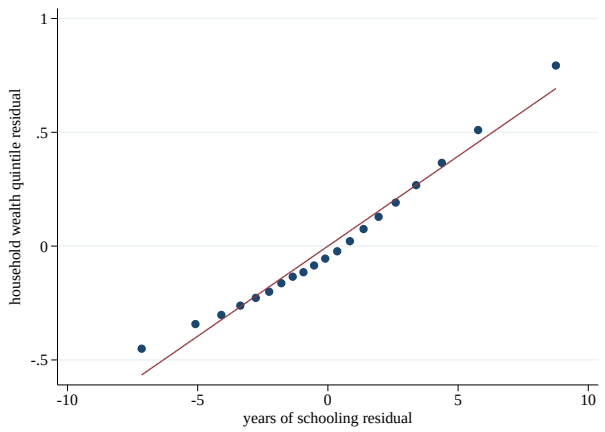
(a) wealth quintile, unconditional



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



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



Child mortality

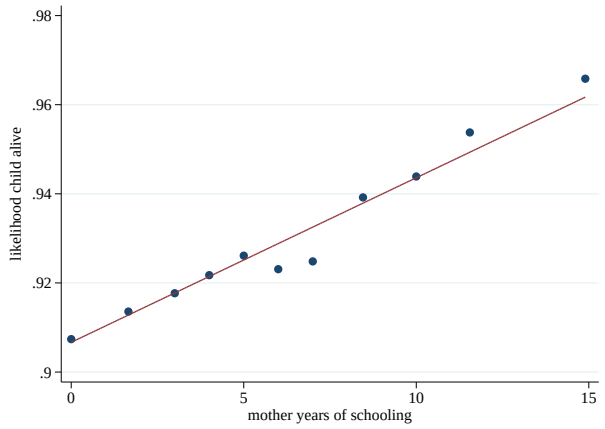
Table D.2: Probability that child survives and years of schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{I}$(child alive)	$\mathbb{I}$(child alive)	$\mathbb{I}$(child alive)	$\mathbb{I}$(child alive)	$\mathbb{I}$(child alive)	$\mathbb{I}$(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

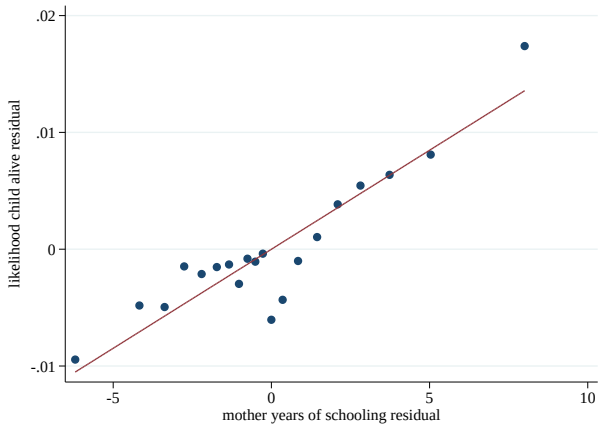
This table shows regression results for child mortality on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to 1 if a child is alive and zero otherwise. Individual controls are mother age, age squared, dummies for children born as twins, child-birth-year 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 D.2: Binned scatter plots, child mortality

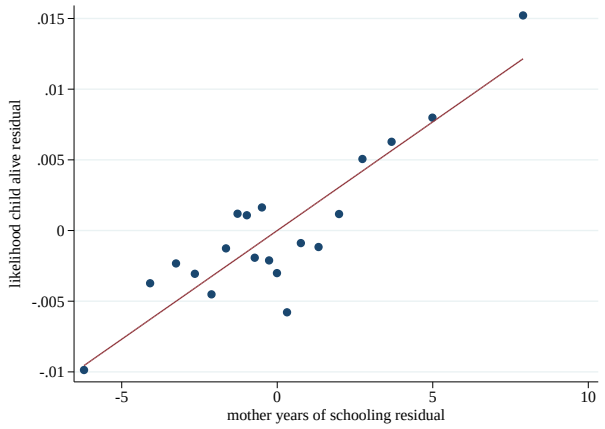
(a) $\mathbb{I}$(child alive), unconditional



(b) $\mathbb{I}$(child alive), conditional on controls and region FE



(c) $\mathbb{I}$(child alive), conditional on controls and admin-2 FE



D.1.2 Female empowerment

Female bargaining power

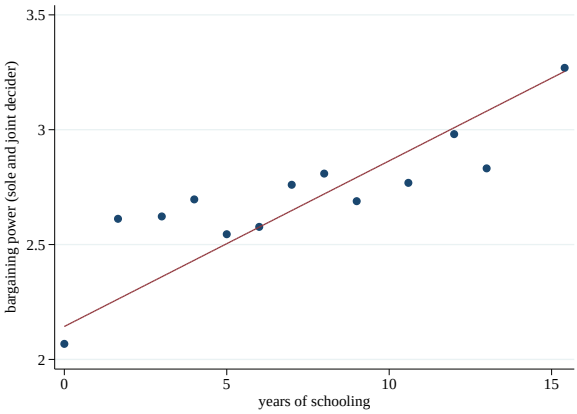
Table D.3: Bargaining power (sole and joint decider) on years of schooling

	(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

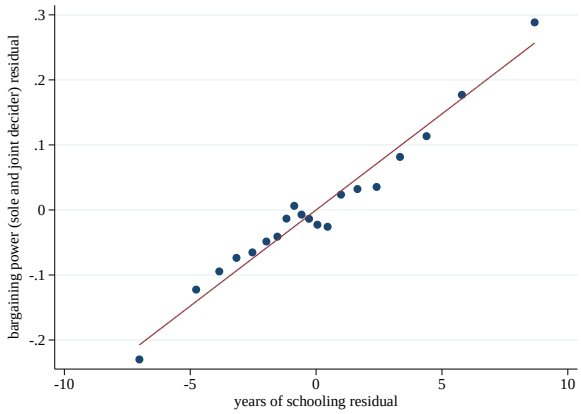
This table shows regression results for individual bargaining power on years of schooling for individuals aged 18+. The dependent variable in all columns is a measure of individual bargaining power. This measure is constructed as the sum of six indicators equal to 1 if an individual takes part (either 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 D.3: Binned scatter plots, bargaining power

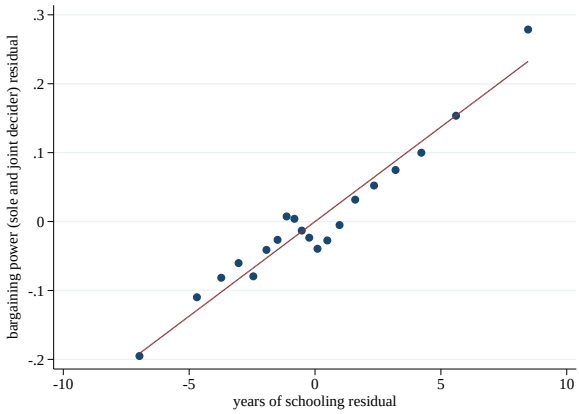
(a) Bargaining power, unconditional



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



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



Attitudes towards domestic violence

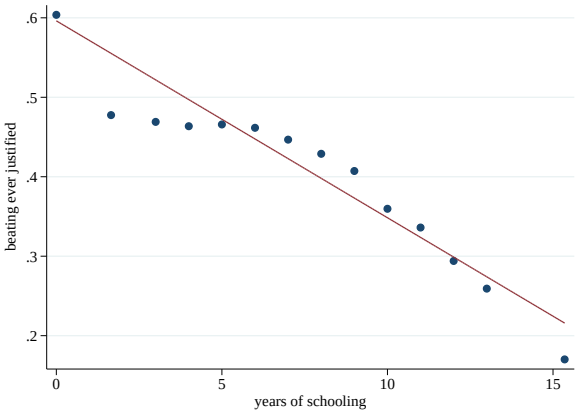
Table D.4: Attitudes towards domestic violence on years of schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{I}$(beating justified)	$\mathbb{I}$(beating justified)	$\mathbb{I}$(beating justified)	$\mathbb{I}$(beating justified)	$\mathbb{I}$(beating justified)	$\mathbb{I}$(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

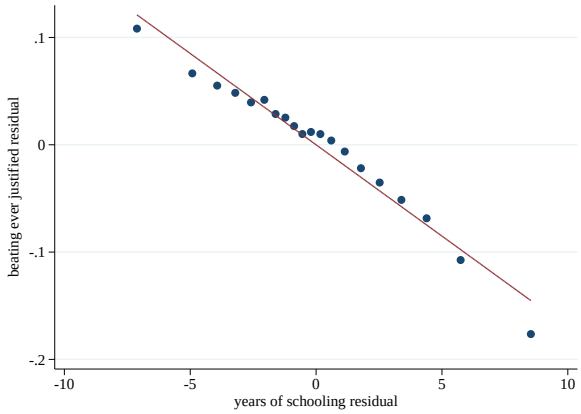
This table shows regression results for attitudes towards domestic violence on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to one if the respondent responds 'yes' to any of the questions of whether beating the wife is justified if she (a) goes 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 D.4: Binned scatter plots, attitudes towards domestic violence

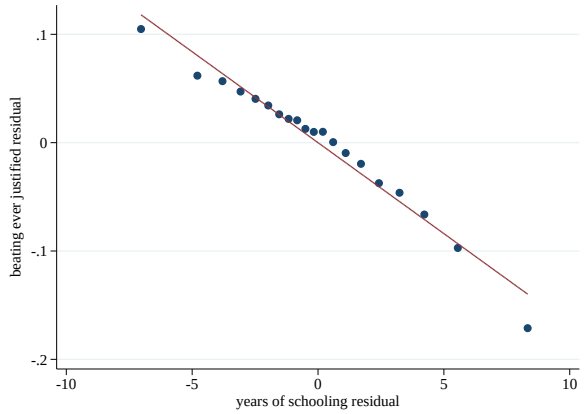
(a) Attitudes towards domestic violence, unconditional



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



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



Fertility

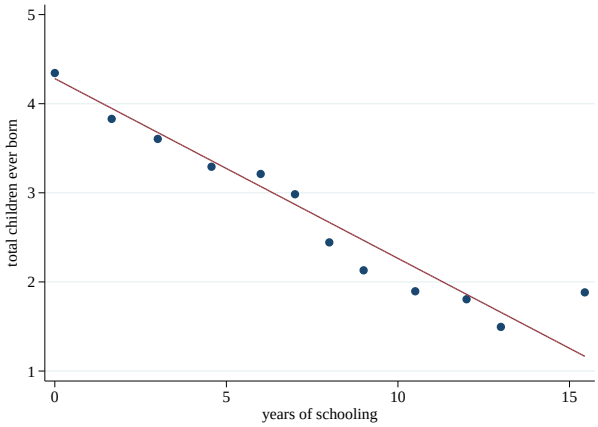
Table D.5: Fertility on years of schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	# children	# children	# children	# children	# children	# 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

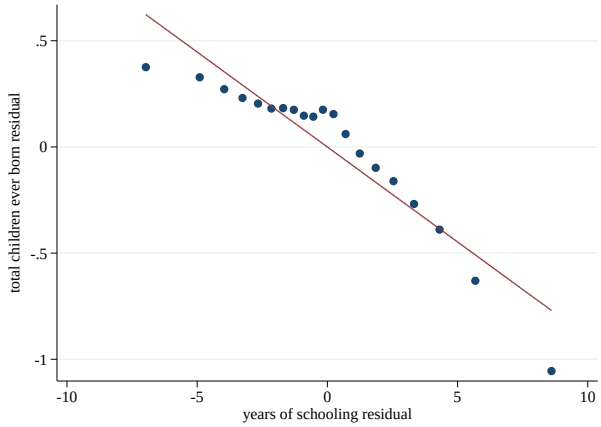
This table shows regression results for total number of children ever born on years of schooling for individuals aged 18+. The dependent variable in all columns is the total number of children ever born. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) 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 D.5: Binned scatter plots, fertility

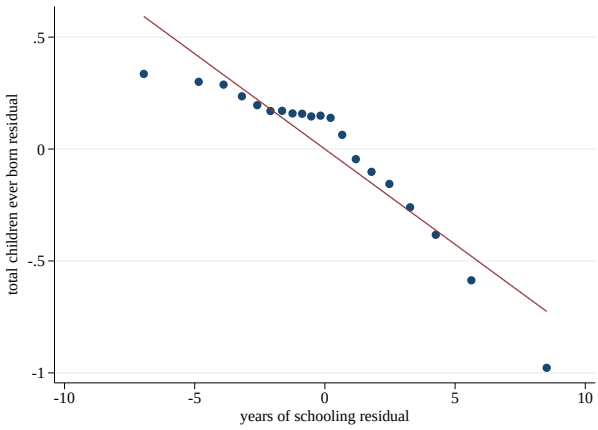
(a) Fertility, unconditional



(b) Fertility, conditional on controls and region FE



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



Age at first marriage

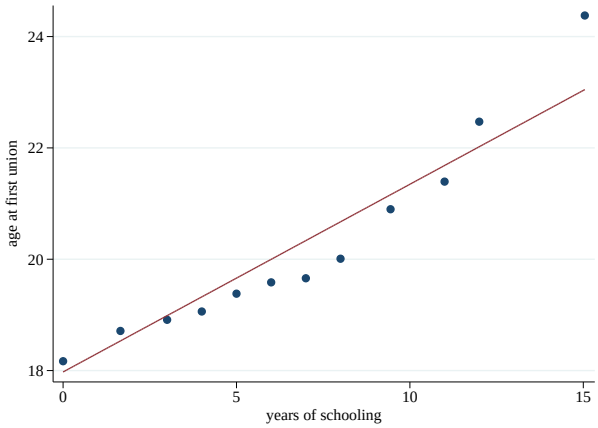
Table D.6: Age of first union on years of schooling

	(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

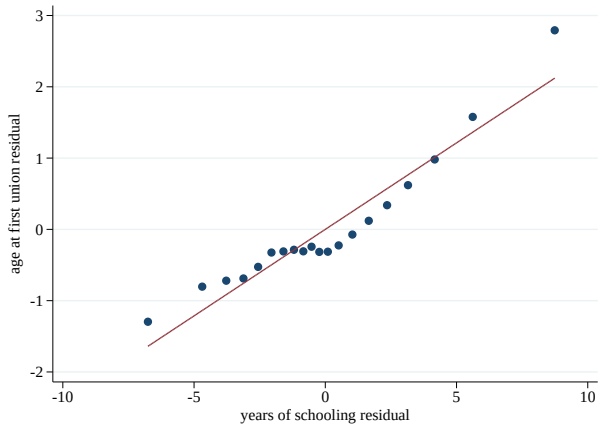
This table shows regression results for age at first union on years of schooling for individuals aged 18+. The dependent variable in all columns is the individual's age at first union / marriage. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the 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 D.6: Binned scatter plots, age at first marriage

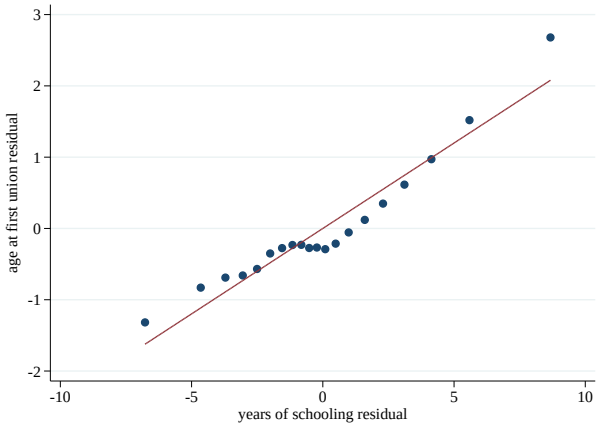
(a) Age at first marriage, unconditional



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



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



D.2 Afrobarometer

D.2.1 Living conditions and deprivation

Living conditions

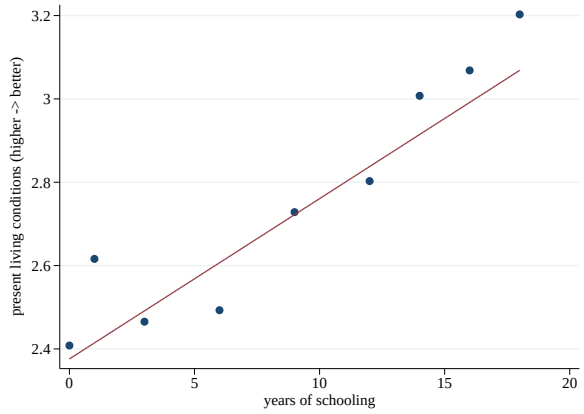
Table D.7: Present living conditions (higher → better) on years of schooling

	(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

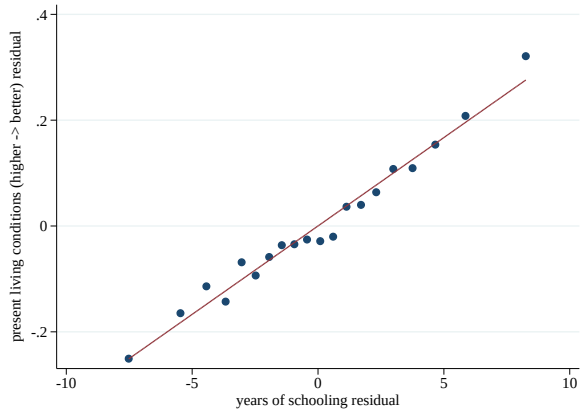
This table shows regression results for living conditions on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's present living conditions (higher → better). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (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 D.7: Binned scatter plots, living conditions

(a) Living conditions, unconditional



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



How often go without food

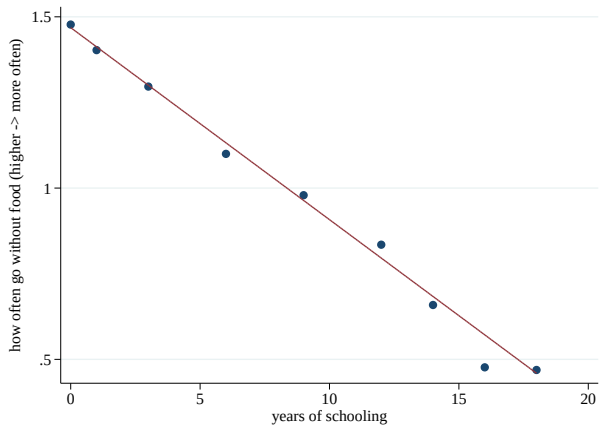
Table D.8: How often go without food (higher → more often) on years of schooling

	(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

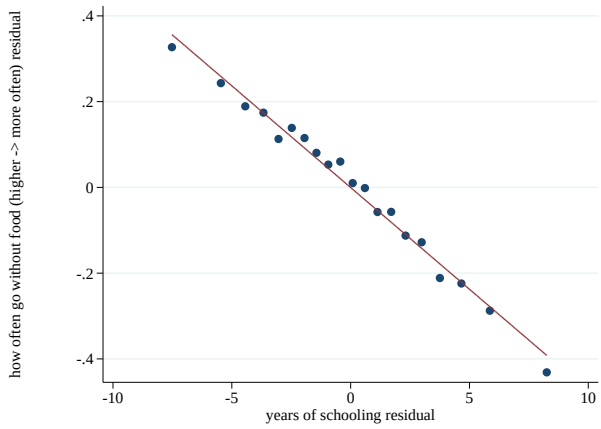
This table shows regression results for frequency of going without food on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without food (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed 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 D.8: Binned scatter plots, how often without food

(a) How often go without food, unconditional



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



How often go without water

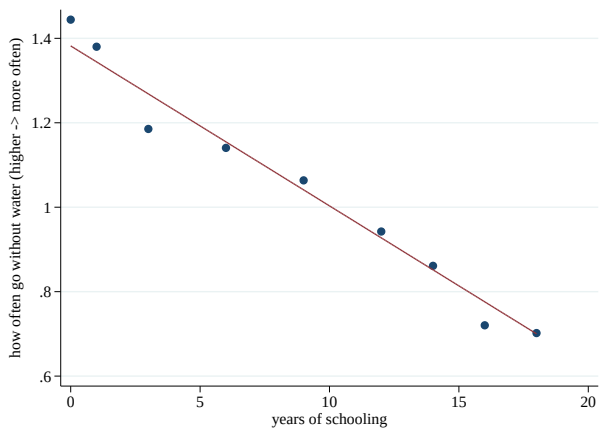
Table D.9: How often go without water (higher → more often) on years of schooling

	(1) freq. no water	(2) freq. no water	(3) freq. no water	(4) freq. no water
years of schooling	-0.0379***	-0.0299***	-0.0273***	-0.0255***
	(-8.37)	(-7.30)	(-13.18)	(-14.59)
individual controls	no	yes	yes	yes
fixed effects	no	no	survey	survey, region
R-squared	0.018	0.032	0.084	0.132
marginal R-squared	.018	.009	.007	.005
within R-squared		.032	.024	.018
N	104261	103213	103213	103213

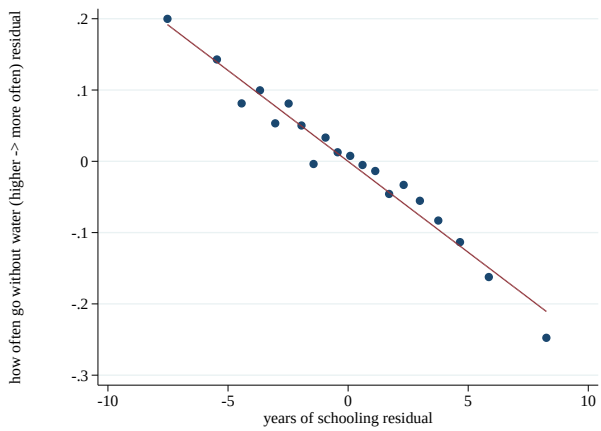
This table shows regression results for frequency of going without water on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without water (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed 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 D.9: Binned scatter plots, how often without water

(a) How often go without water, unconditional



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



D.2.2 Civicness

Interest in public affairs

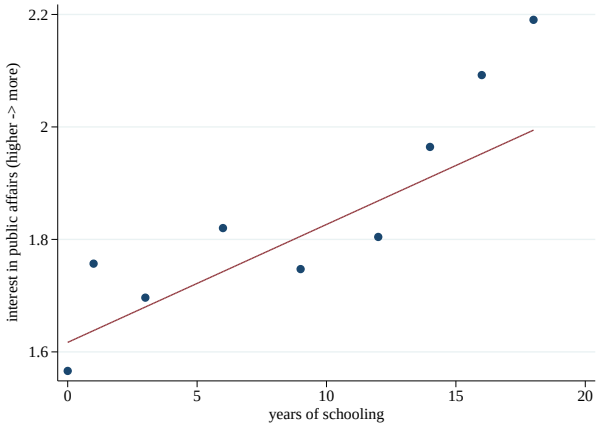
Table D.10: Interest in public affairs (higher → more) on years of schooling

	(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

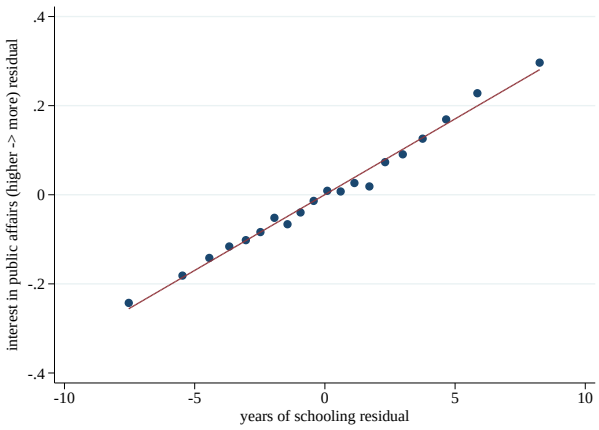
This table shows regression results for interest in public affairs on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's interest in public affairs (higher → more). Column (1) shows the simple bivariate relationship without controls or fixed effects. 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 D.10: Binned scatter plots, interest in public affairs

(a) Interest in public affairs, unconditional



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



Support for democracy

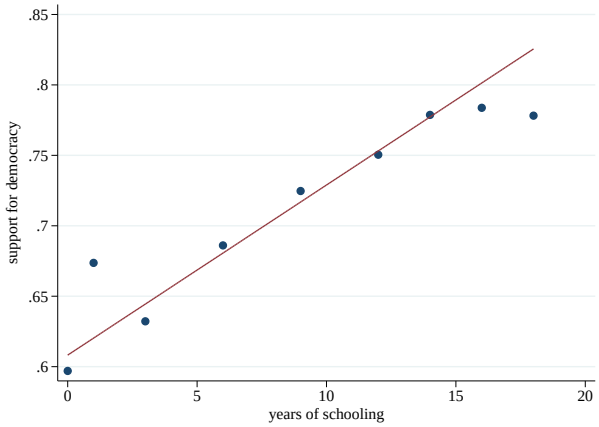
Table D.11: Support for democracy on years of schooling

	(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 D.11: Binned scatter plots, support for democracy

(a) Support for democracy, unconditional



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