Intergenerational 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) $N_{owned}^{age\geq 14}$	(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) $N_{olded,\ no\ mghh}^{14\leq age\leq 18}$
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

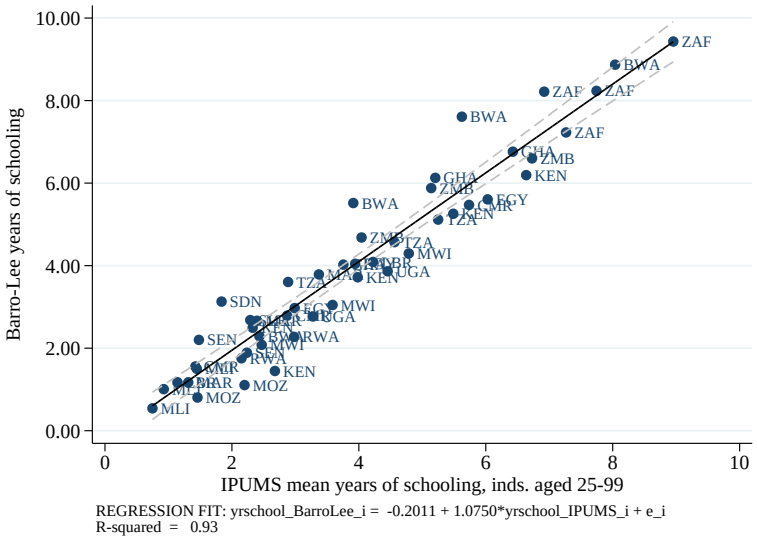
This table shows co-residence rates for individuals aged 8, individuals aged 14-18 and invidiuals aged 14-25. To compute co-residence rates, we start with a sample of individuals for whom their own eduactional attainment as well as relationship to household head is observed. The latter does not exclude 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

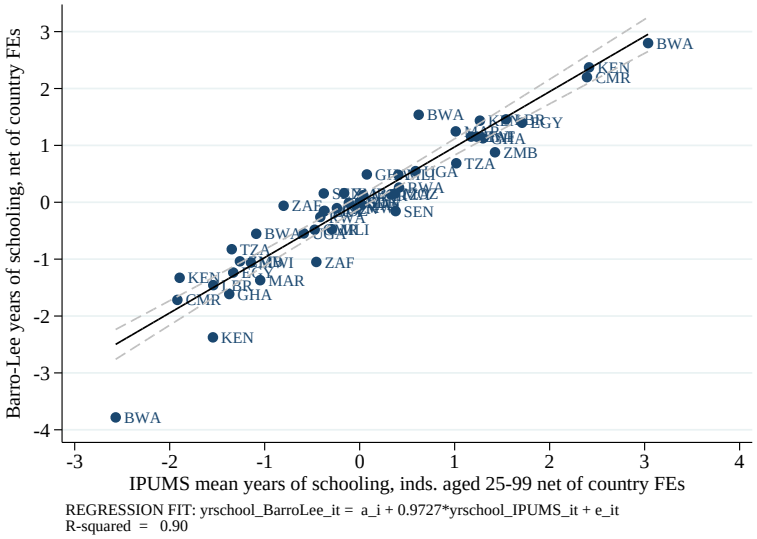
To gauge the quality of our education data, we correlate mean years of schooling for individuals aged 25-99 to the data from Barro and Lee (2013) who also report figures for years of schooling for this age range As Barro and Lee (2013) only report their estimates of years of schooling at 5-year intervals, 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

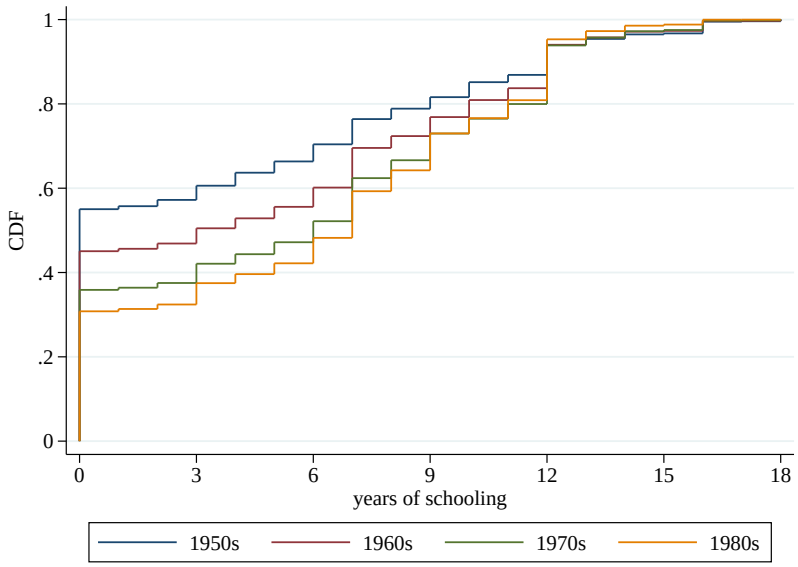
Table A.7: Average education by country-cohort, continued

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.

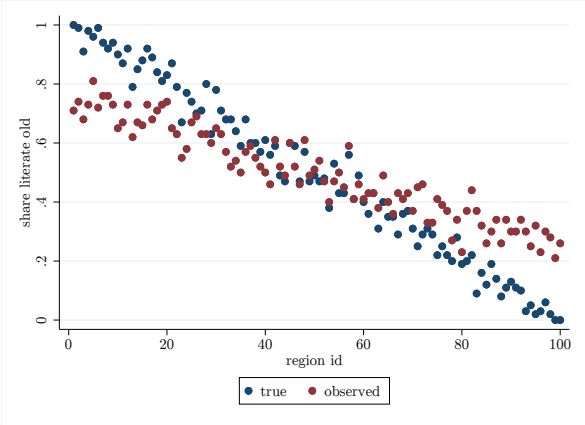
Consider a country with $R = 100$ regions. Each region $r \in \{1, \ldots, 100\}$ is populated by $n_r = 100$ old individuals who each have one child. Old literacy is determined by the value of a random variable $u \sim \mathcal{U}[0,1]$ drawn independently across all individuals in all regions. Individuals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

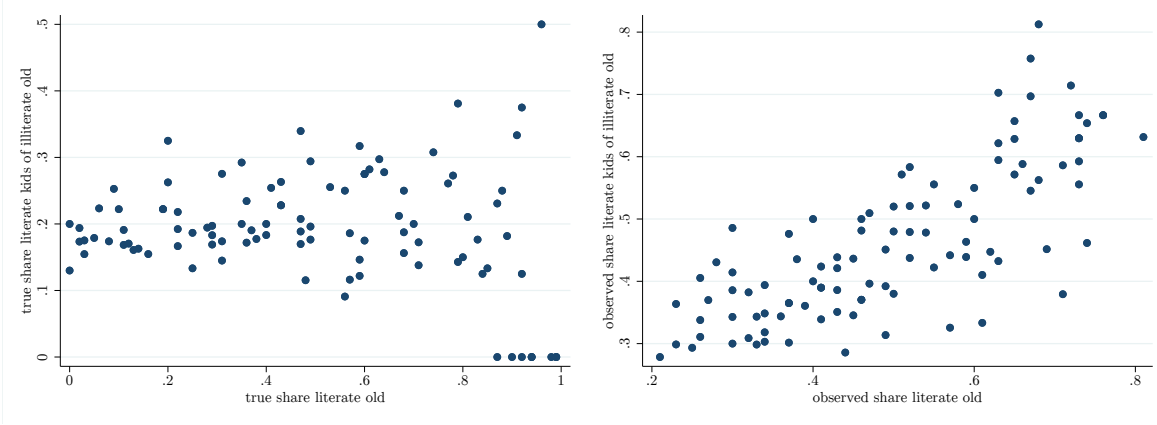


Child literacy is determined by a random variable $v \sim \mathcal{U}[0,1]$ drawn independently across all individuals in all regions. Children of truly literate parents have a 0.8 chance of being literate and children of truly illiterate parents have a 0.2 chance of being literate. As with the old, 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.

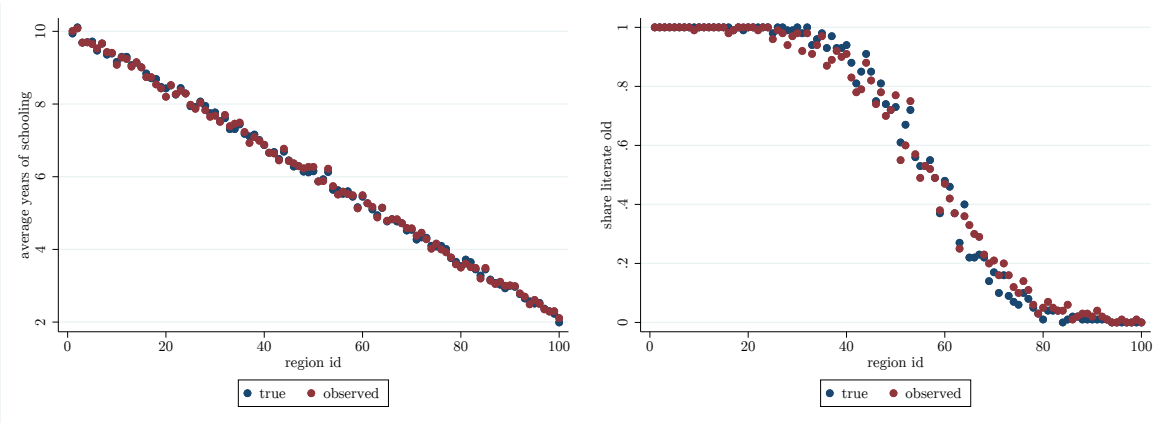
Case 2: Mobility and literacy defined in terms of primary schooling, measurement error affects schooling.

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.

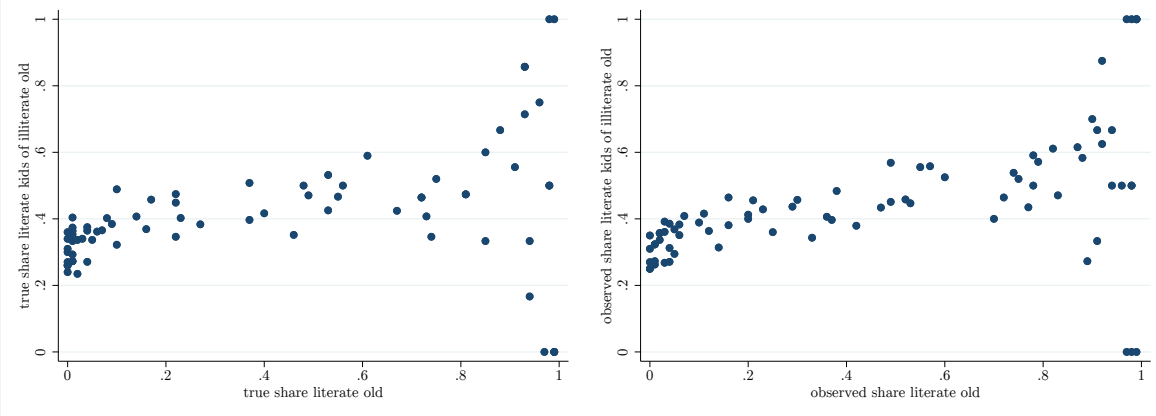
3) normalize by rows so that the transition probabilities for every parent sum to 1.

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

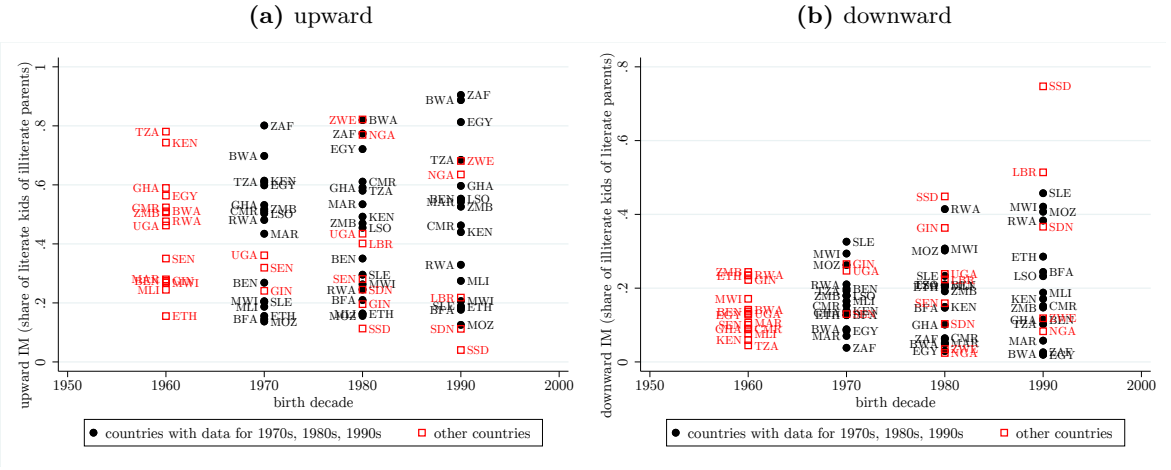
This table shows estimates of IM (likelihood that children of illiterate parents become literate, higher numbers → higher upward IM; likelihood that children of parents with at least primary, who complete less than primary, higher numbers → higher downward IM). Columns (1)-(4) show estimates for individuals 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



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. $*p < 0.1, **p < 0.5, ***p < 0.01$.

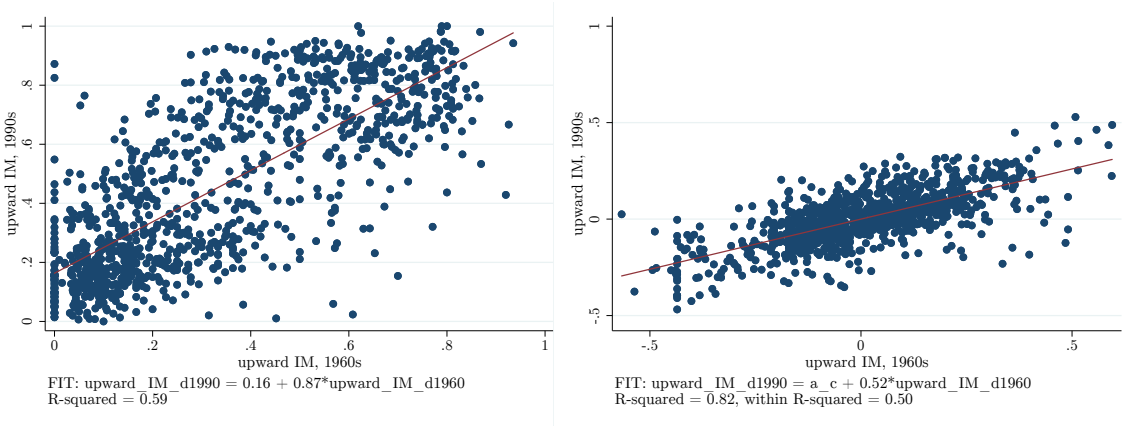
Table B.7: IM at the district × cohort level, trends, balanced sample

	(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)

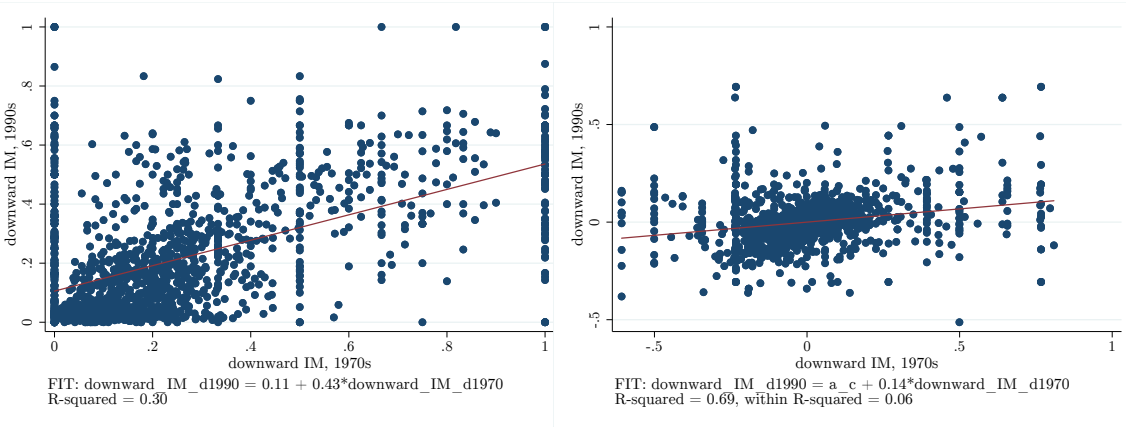
(a) OLS **(b)** country fixed effects



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×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

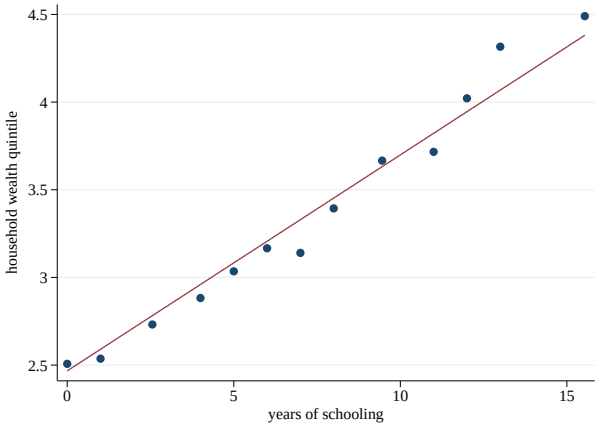
Table D.1: Household wealth quintile and years of schooling

	(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

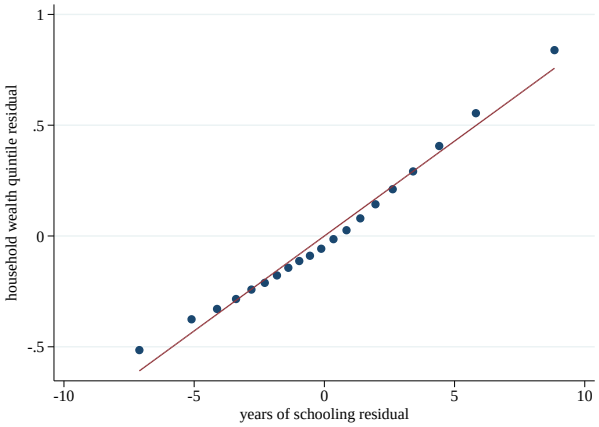
This table shows regression results of household wealth on years of schooling for individuals aged 18+. The dependent variable in all columns is the DHS household wealth quintile (computed for each survey, i.e. country-year) separately based on the DHS-computed wealth index). Individual controls are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

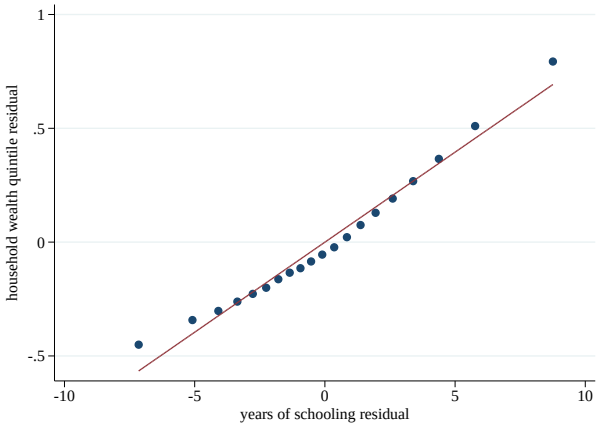
(a) wealth quintile, unconditional



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



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



Child mortality

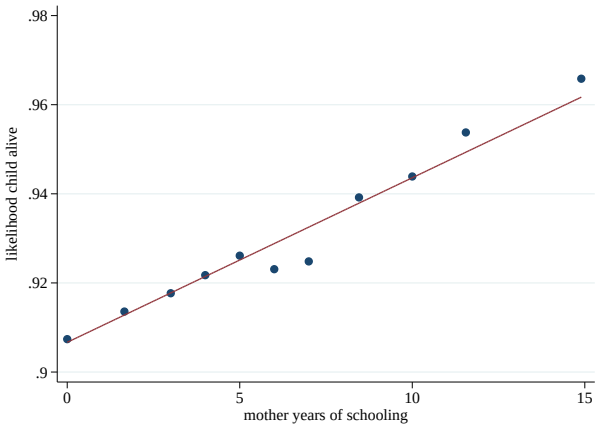
Table D.2: Probability that child survives and years of schooling

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

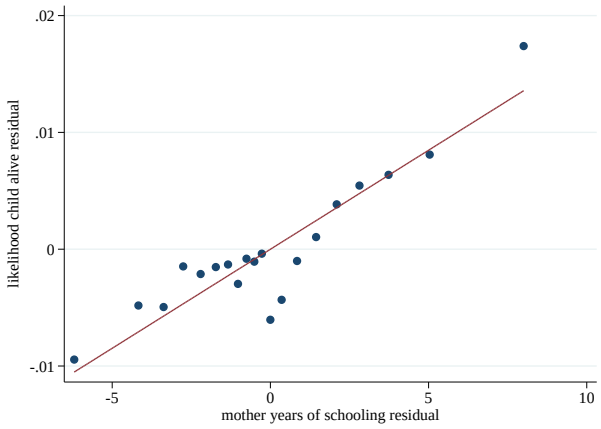
This table shows regression results for child mortality on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to 1 if a child is alive and zero otherwise. Individual controls are mother age, age squared, dummies for children born as twins, child-birth-year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

Figure D.2: Binned scatter plots, child mortality

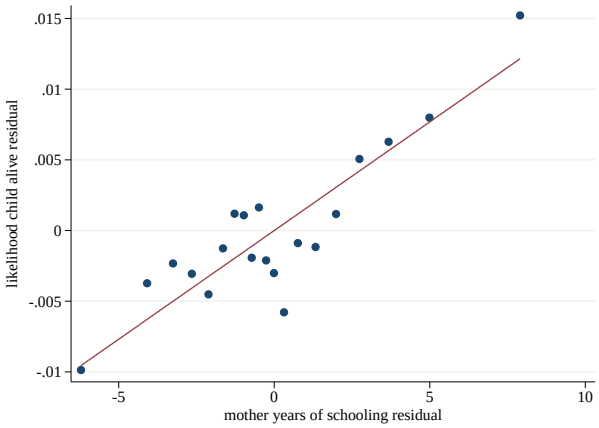
(a) 𝕀(child alive), unconditional



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



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



D.1.2 Female empowerment

Female bargaining power

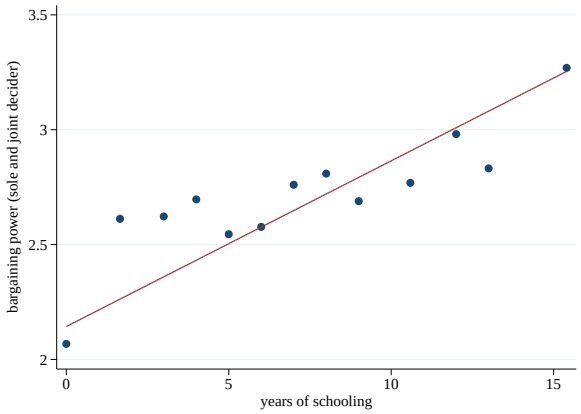
Table D.3: Bargaining power (sole and joint decider) on years of schooling

	(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

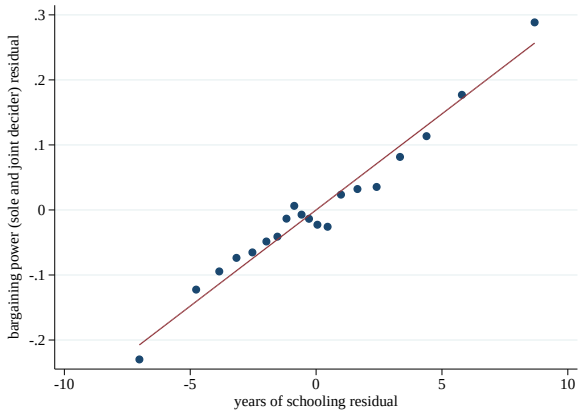
This table shows regression results for individual bargaining power on years of schooling for individuals aged 18+. The dependent variable in all columns is a measure of individual bargaining power. This measure is constructed as the sum of six indicators equal to 1 if an individual takes part (either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

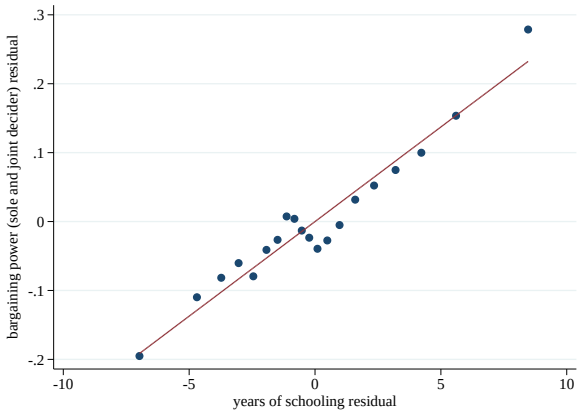
(a) Bargaining power, unconditional



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



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



Attitudes towards domestic violence

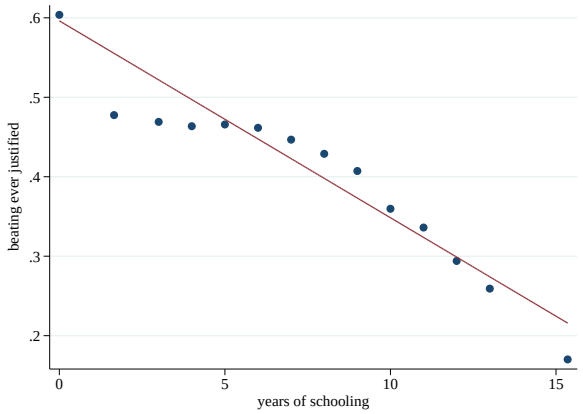
Table D.4: Attitudes towards domestic violence on years of schooling

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

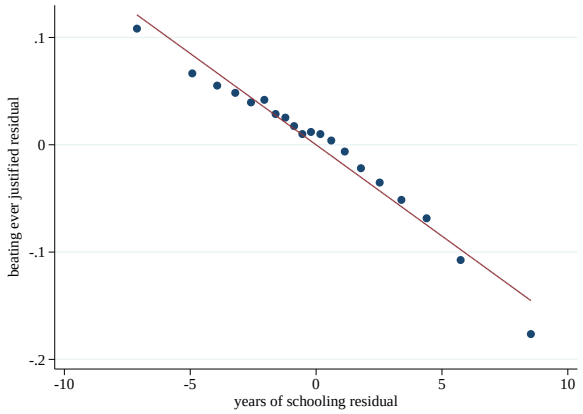
This table shows regression results for attitudes towards domestic violence on years of schooling for individuals aged 18+. The dependent variable in all columns is an indicator equal to one if the respondent responds 'yes' to any of the questions of whether beating the wife is justified if she (a) goes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

Figure D.4: Binned scatter plots, attitudes towards domestic violence

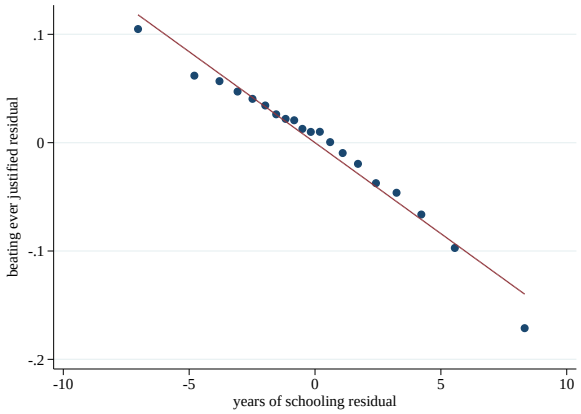
(a) Attitudes towards domestic violence, unconditional



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



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



Fertility

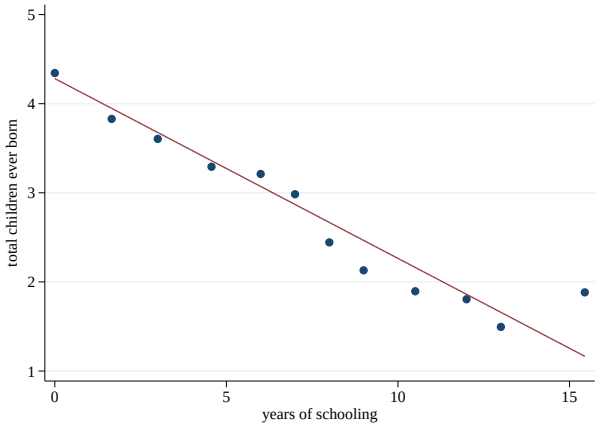
Table D.5: Fertility on years of schooling

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

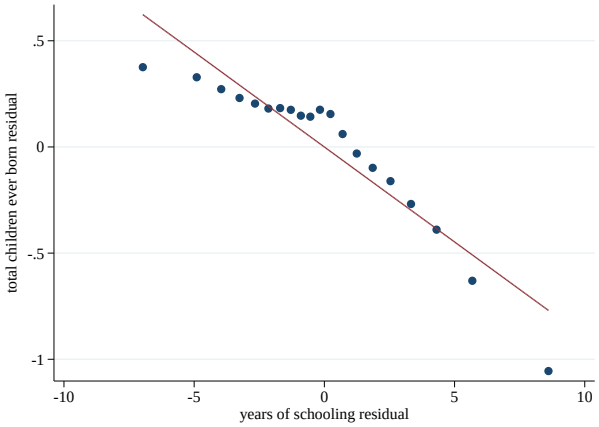
This table shows regression results for total number of children ever born on years of schooling for individuals aged 18+. The dependent variable in all columns is the total number of children ever born. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2)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

Figure D.5: Binned scatter plots, fertility

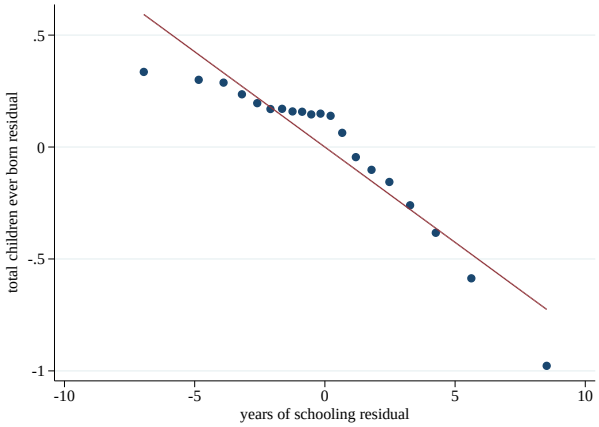
(a) Fertility, unconditional



(b) Fertility, conditional on controls and region FE



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



Age at first marriage

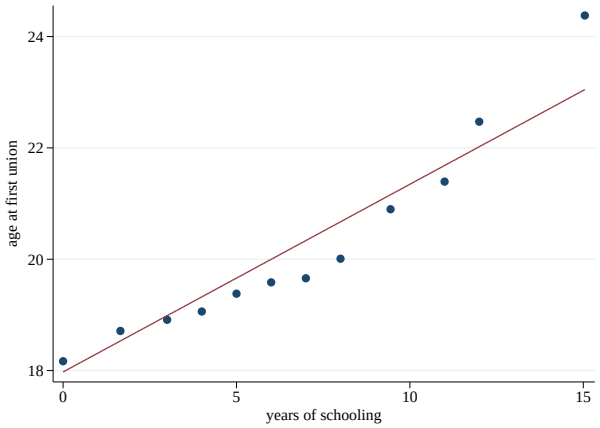
Table D.6: Age of first union on years of schooling

	(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

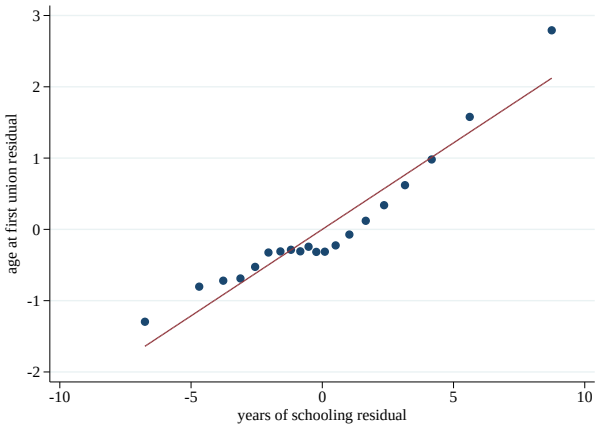
This table shows regression results for age at first union on years of schooling for individuals aged 18+. The dependent variable in all columns is the individual's age at first union / marriage. Column (1) shows the simple bivariate relationship without controls or fixed effects. Column (2) shows the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

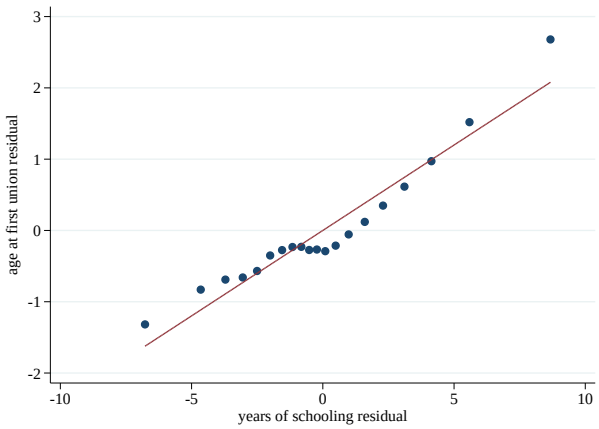
(a) Age at first marriage, unconditional



(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

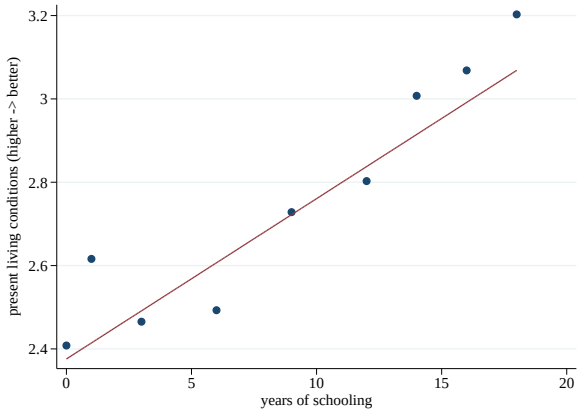
Table D.7: Present living conditions (higher → better) on years of schooling

	(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

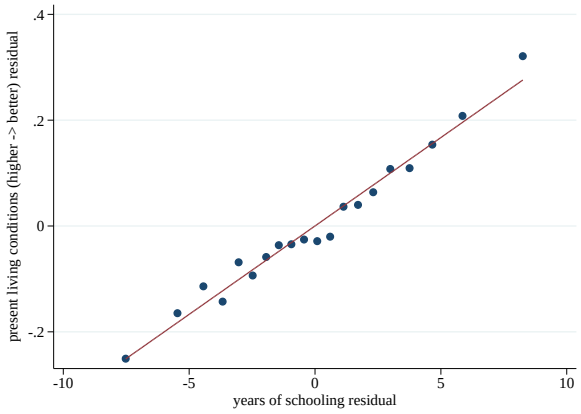
This table shows regression results for living conditions on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's present living conditions (higher → better). Column (1) shows the simple bivariate relationship without controls or fixed effects. Column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

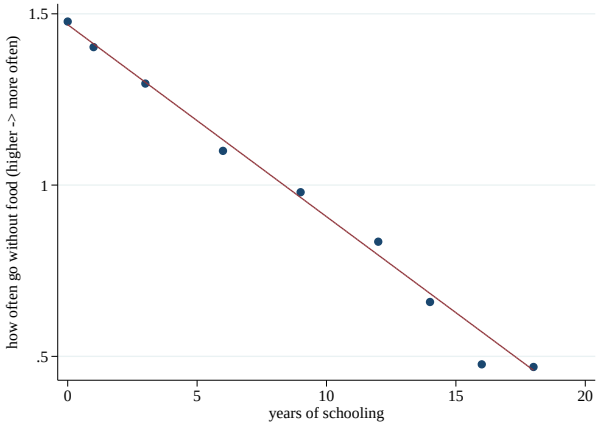
Table D.8: How often go without food (higher → more often) on years of schooling

	(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

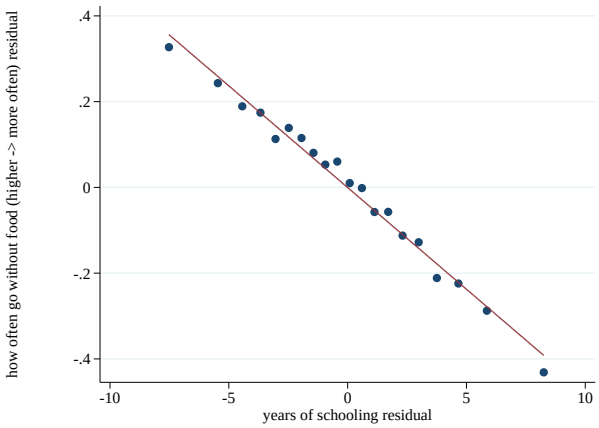
This table shows regression results for frequency of going without food on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without food (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed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

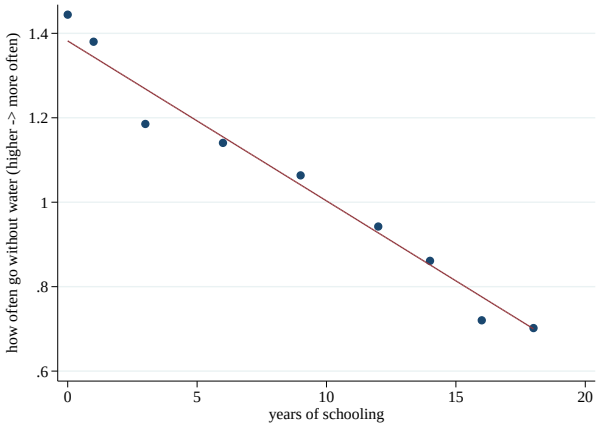
Table D.9: How often go without water (higher → more often) on years of schooling

	(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

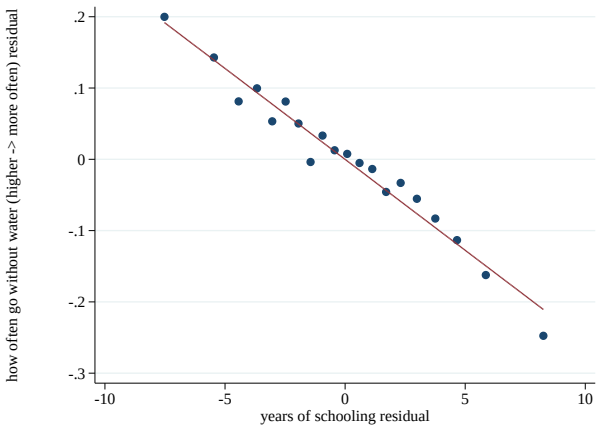
This table shows regression results for frequency of going without water on years of schooling for individuals aged 18+. The dependent variable in all columns is how often the repondent goes without water (higher → more often). Column (1) shows the simple bivariate relationship without controls or fixed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

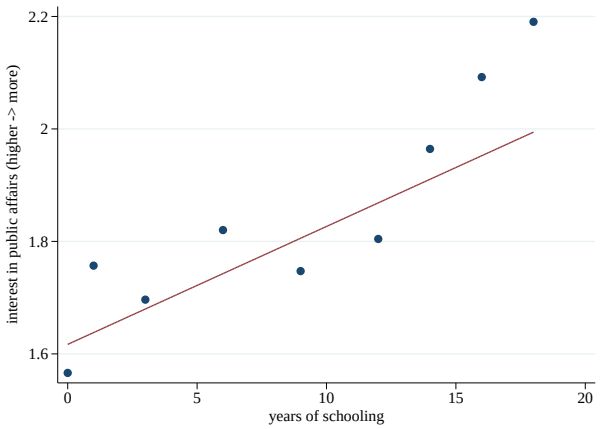
Table D.10: Interest in public affairs (higher → more) on years of schooling

	(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

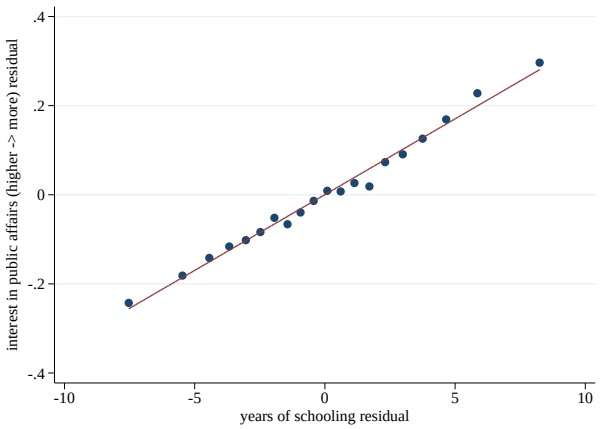
This table shows regression results for interest in public affairs on years of schooling for individuals aged 18+. The dependent variable in all columns is the respondent's interest in public affairs (higher → more). Column (1) shows the simple bivariate relationship without controls or fixed effects.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

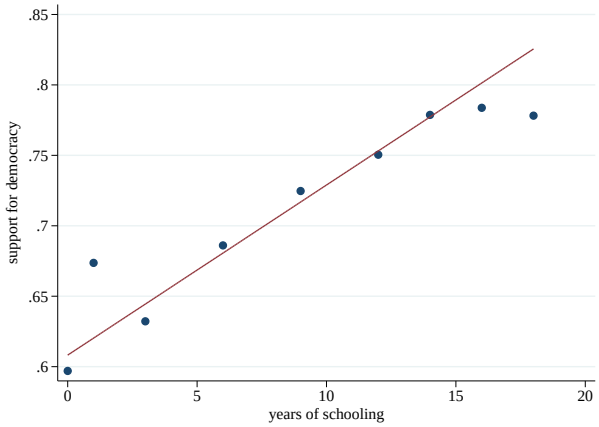
Table D.11: Support for democracy on years of schooling

	(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