Intergenerational Mobility in Africa
Supplementary Online Appendix

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

Sebastian Hohmann[†]

Stelios Michalopoulos[‡]
Brown University, CEPR and NBER

Elias Papaioannou[§]
London Business School and CEPR

This draft: March 2020

Abstract

The Online Appendix provides additional evidence and sensitivity checks. Section A presents summary statistics on educational attainment for each country-birth-decade. Section B details the sample construction process. In Section C we report validation checks of the IPUMS data on education with alternative cross-country and regional statistics. We also explore the relationship between educational attainment and various proxies of well-being across African regions using all available Demographic and Health as well as Afrobarometer Surveys. Section D provides details on family structure and cohabitation patterns across the millions of African households in our dataset. In Section E we report and discusses intergenerational mobility (IM) statistics, distinguishing between gender and rural-urban household residence. In Section F we present robustness checks –as well as LASSO estimates- of the correlational analysis where we associate educational mobility to at-independence, historical, and geographical features. In section G we describe sensitivity checks and additional results pertaining to the regional childhood exposure estimates.

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

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

[*]Harvard University, IGIER, CEPR and NBER aalesina@harvard.edu https://scholar.harvard.edu/alesina/home

[†]sebas.hohmann@gmail.com

[‡]Brown University, CEPR and NBER smichalo@brown.edu https://sites.google.com/site/steliosecon/

[§]London Business School and CEPR eliasp@london.edu https://sites.google.com/site/papaioannouelias/

A Education Data and Statistics

This section presents summary statistics on education in Africa without imposing any sample restrictions, related to observing younger and older generation in the same household. We include all individuals aged 25 and older at the time of the census, using all available censuses for a given country.

Table A.1: Average education by country-cohort

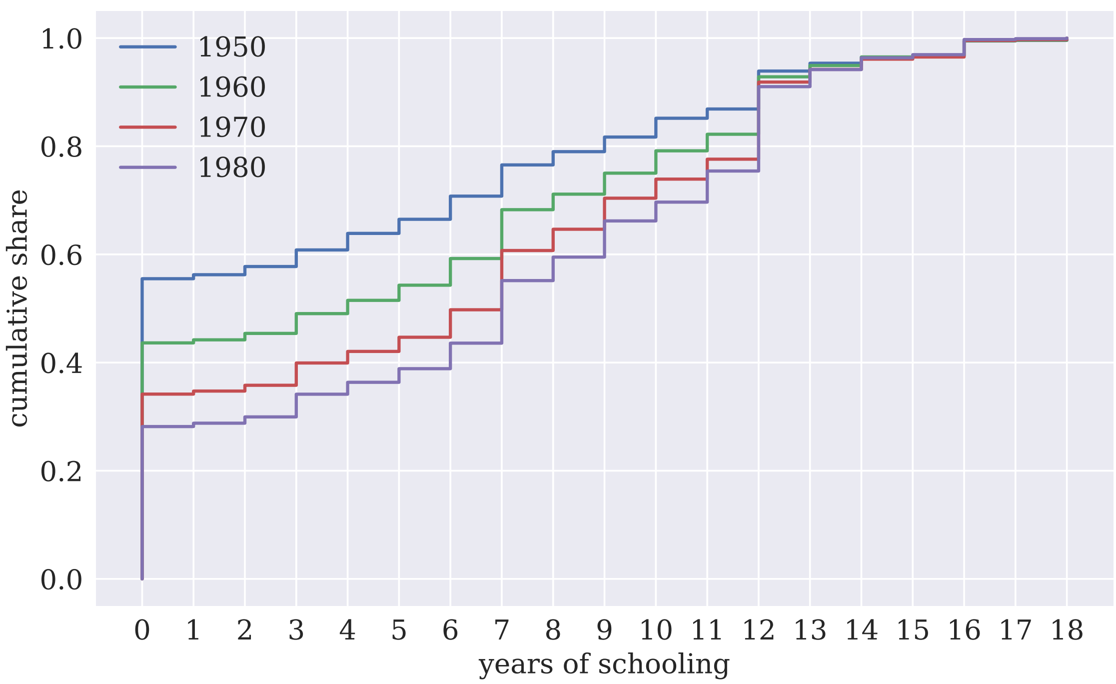
country	birth decade	(1) mean years of schooling	(2) % less than primary	(3) % primary	(4) % secondary	(5) % tertiary
Benin	1950	1.951	82.467	12.955	3.154	1.424
Benin	1960	2.757	75.158	18.878	4.605	1.359
Benin	1970	2.902	74.255	20.007	4.221	1.518
Benin	1980	3.93	67.926	19.665	9.674	2.735
Burkina Faso	1950	0.747	92.608	5.565	0.929	0.897
Burkina Faso	1960	1.011	90.135	6.756	1.798	1.312
Burkina Faso	1970	1.66	84.485	10.274	3.735	1.506
Burkina Faso	1980	1.933	81.843	12.99	4.407	0.76
Botswana	1950	4.67	54.884	35.411	5.68	4.025
Botswana	1960	6.714	33.495	49.888	9.922	6.695
Botswana	1970	9.011	15.077	56.931	19.111	8.881
Botswana	1980	10.477	7.991	56.116	26.589	9.304
Cameroon	1950	4.591	51.667	42.596	4.603	1.134
Cameroon	1960	5.765	41.248	48.831	7.797	2.124
Cameroon	1970	6.888	33.428	51.312	9.007	6.253
Cameroon	1980	6.213	40.941	42.713	14.293	2.053
Egypt	1950	4.091	67.667	5.414	16.954	9.966
Egypt	1960	5.531	56.294	5.844	26.476	11.385
Egypt	1970	7.471	40.305	8.401	36.76	14.534
Egypt	1980	8.559	32.425	7.622	42.158	17.795
Ethiopia	1950	1.106	93.15	4.177	2.356	0.317
Ethiopia	1960	2.039	84.867	9.666	5.143	0.324
Ethiopia	1970	2.186	82.437	11.445	5.541	0.576
Ethiopia	1980	2.397	80.105	13.51	5.769	0.616
Ghana	1950	5.985	48.506	38.501	11.513	1.479
Ghana	1960	6.361	46.122	39.72	12.2	1.958
Ghana	1970	6.249	43.395	38.872	15.761	1.972
Ghana	1980	7.086	35.715	38.94	23.07	2.275
Guinea	1950	1.78	85.741	6.212	5.892	2.155
Guinea	1960	1.736	84.813	10.857	3.066	1.264
Guinea	1970	1.476	86.472	10.394	2.687	0.447
Kenya	1950	4.961	52.067	31.459	15.117	1.357
Kenya	1960	6.721	33.647	39.214	25.066	2.072
Kenya	1970	7.669	24.287	46.098	27.319	2.296
Kenya	1980	7.897	22.664	45.813	29.3	2.222
Liberia	1950	4.009	66.331	13.273	16.623	3.773
Liberia	1960	4.687	59.328	18.476	19.002	3.193
Liberia	1970	4.665	58.34	22.811	17.104	1.744
Liberia	1980	4.804	55.81	28.218	15.289	0.683
Lesotho	1950	5.022	64.29	28.316	5.815	1.579
Lesotho	1960	6.243	46.829	41.504	10.018	1.649
Lesotho	1970	6.97	37.181	47.037	13.936	1.846
Lesotho	1980	7.352	34.083	47.227	17.289	1.401
Morocco	1950	2.774	80.898	11.074	5.881	2.148
Morocco	1960	3.846	72.96	12.202	10.658	4.18
Morocco	1970	4.826	63.277	21.542	10.688	4.493
Mali	1950	1.263	88.331	8.255	2.235	1.18
Mali	1960	1.42	87.435	9.756	1.772	1.036
Mali	1970	1.573	87.52	8.981	2.025	1.474
Mali	1980	2.111	83.042	11.108	3.666	2.184
Mozambique	1950	1.539	92.522	5.838	1.261	0.379
Mozambique	1960	2.158	87.326	10.195	2.015	0.465
Mozambique	1970	2.385	84.851	12.433	2.336	0.38
Mozambique	1980	2.852	79.623	16.678	3.387	0.312
Malawi	1950	3.435	79.163	16.394	4.028	0.415
Malawi	1960	4.275	73.281	19.762	6.302	0.655
Malawi	1970	5.234	66.508	21.125	11.554	0.813
Malawi	1980	6.228	57.99	26.712	14.552	0.746
Nigeria	1950	3.97	58.52	22.206	14.12	5.154
Nigeria	1960	5.017	49.657	23.042	22.158	5.143
Nigeria	1970	5.663	44.236	23.668	27.312	4.784
Nigeria	1980	6.559	38.211	22.276	34.153	5.361
Rwanda	1950	2.435	80.749	18.227	0.731	0.294
Rwanda	1960	3.404	68.46	29.026	1.82	0.694
Rwanda	1970	4.727	54.034	39.362	4.821	1.783
Rwanda	1980	4.745	64.986	23.806	8.518	2.69
Sudan	1950	1.375	89.11	7.335	1.082	2.473
Sudan	1960	1.836	83.667	12.004	1.093	3.235
Sudan	1970	2.137	81.228	13.169	1.001	4.602
Sudan	1980	2.369	79.706	14.009	1.025	5.26
South Sudan	1950	0.755	93.13	4.499	1.308	1.062
South Sudan	1960	1.131	88.875	7.792	2.114	1.219
South Sudan	1970	1.208	87.769	9.324	2.024	0.883

Table A.1: Average education by country-cohort, conitnued

country	birth decade	(1) mean years of schooling	(2) % less than primary	(3) % primary	(4) % secondary	(5) % tertiary
South Sudan	1980	1.476	85.013	11.725	2.433	0.829
Senegal	1950	2.281	77.667	16.691	4.05	1.592
Senegal	1960	2.444	75.789	18.553	4.464	1.194
Senegal	1970	2.756	72.353	21.039	5.563	1.045
Sierra Leone	1950	2.176	78.683	16.906	2.326	2.086
Sierra Leone	1960	2.441	75.452	21.13	1.878	1.539
Sierra Leone	1970	2.714	72.871	24.226	1.624	1.279
Tanzania	1950	3.896	60.099	32.915	6.185	0.801
Tanzania	1960	5.299	34.95	57.259	6.707	1.084
Tanzania	1970	5.9	26.606	64.117	7.549	1.728
Tanzania	1980	6.148	27.154	58.932	9.701	4.214
Uganda	1950	4.026	63.12	32.656	3.243	0.98
Uganda	1960	4.651	57.117	37.107	4.845	0.931
Uganda	1970	5.352	51.709	39.503	7.697	1.091
South Africa	1950	6.765	35.87	43.087	16.681	4.363
South Africa	1960	8.128	23.92	46.063	25.153	4.863
South Africa	1970	9.463	13.886	43.666	36.883	5.565
South Africa	1980	10.359	6.884	43.818	42.912	6.386
Zambia	1950	5.394	46.769	38.804	11.98	2.447
Zambia	1960	5.968	39.404	44.666	13.449	2.481
Zambia	1970	6.576	34.397	46.235	17.732	1.635
Zambia	1980	7.2	30.391	45.3	23.497	0.812
Zimbabwe	1950	5.6	52.109	41.752	4.033	2.107
Zimbabwe	1960	8.091	27.409	59.404	8.738	4.449
Zimbabwe	1970	9.613	10.033	77.086	9.724	3.157
Zimbabwe	1980	9.803	8.438	78.849	9.674	3.039
Togo	1950	3.471	68.507	24.177	5.594	1.723
Togo	1960	4.374	58.817	34.602	4.847	1.734
Togo	1970	4.76	56.813	35.289	6.27	1.629
Togo	1980	5.4	52.14	37.458	8.844	1.558

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

Figure A.1: Cumulative distribution of years of schooling by birth decade



This figure shows the cumulative distribution function for years of schooling for four birth decades since 1950 for ages 25+. Note that the sample is unbalanced in the sense that not all countries in the sample have a census such that each country is represented in every birth decade.

B Sample Construction

The table below gives the number of observations for each country-census, detailing how we compile the sample that we use in the main analysis from the raw IPUMS data (see please the Table notes).

Table B.1: Sample construction

(1) country	(2) year	(3) fraction	(4) N_{all}	(5) N_{age}	(6) N_{owned}	(7) $N_{owned}^{age\geq14}$	(8) $N_{owned}^{14\leq age\leq25}$	(9) $N_{owned}^{14\leq age\leq18}$	(10) $N_{olded}^{age\geq14}$	(11) $N_{olded}^{14\leq age\leq25}$	(12) $N_{olded}^{14\leq age\leq18}$	(13) $N_{olded,\ no\ mghh}^{age\geq14}$	(14) $N_{olded,\ no\ mghh}^{14\leq age\leq25}$	(15) $N_{olded,\ no\ mghh}^{14\leq age\leq18}$
Benin	1979	10	331,049	329,784	244,898	171,690	62,112	24,322	72,597	37,048	18,381	39,210	30,663	17,044
Benin	1992	10	498,419	498,107	435,652	256,763	101,543	44,735	109,234	65,432	35,000	57,957	56,106	33,190
Benin	2002	10	685,467	685,467	612,658	373,452	155,832	69,048	160,458	104,331	57,364	87,164	90,874	54,655
Benin	2013	10	1,009,693	1,009,693	911,604	559,525	240,049	108,694	244,182	170,580	93,329	141,282	148,917	88,090
Botswana	1981	10	97,238	96,187	72,951	50,399	20,258	9,533	19,054	12,165	6,847	12,048	10,821	6,413
Botswana	1991	10	132,623	132,623	113,172	78,814	32,680	15,830	26,842	18,264	10,979	16,608	15,573	10,013
Botswana	2001	10	168,676	168,134	159,257	109,649	44,806	20,616	42,629	29,159	16,102	26,541	25,017	14,834
Botswana	2011	10	201,752	201,235	190,212	138,375	48,926	20,677	47,177	28,874	14,302	29,888	24,861	13,288
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	250,977	157,907	95,710	119,954	127,332	88,869
Burkina Faso	2006	10	1,417,824	1,410,123	1,244,291	770,161	321,384	151,393	327,195	211,275	123,364	153,253	167,136	112,927
Cameroon	1976	10	736,514	736,320	605,749	413,814	157,287	72,886	153,664	93,046	52,858	94,677	81,910	50,900
Cameroon	1987	10	897,211	896,649	763,652	481,727	191,552	90,805	178,841	114,345	67,018	117,877	104,219	65,397
Cameroon	2005	10	1,772,359	1,772,359	1,542,200	1,018,632	438,407	199,054	433,774	299,540	164,047	261,360	257,094	154,003
Egypt	1986	14.1	6,799,093	6,794,386	5,418,332	4,262,426	1,609,719	722,024	1,931,495	1,345,174	693,323	1,397,036	1,276,970	679,214
Egypt	1996	10	5,902,243	5,901,839	4,453,382	3,810,835	1,471,285	718,874	1,616,808	1,230,963	695,795	1,302,668	1,211,167	692,249
Egypt	2006	10	7,282,434	7,282,434	5,739,722	5,096,618	1,977,932	785,619	2,046,232	1,590,965	759,450	1,706,597	1,568,687	756,807
Ethiopia	1984	10	3,404,306	3,398,027	2,733,575	1,800,650	620,022	303,780	556,877	360,470	234,377	397,880	344,904	230,979
Ethiopia	1994	10	5,044,598	5,044,597	4,201,616	2,833,214	1,224,762	614,179	1,034,238	788,117	498,607	743,852	765,183	494,430
Ethiopia	2007	10	7,434,086	7,434,086	1,097,614	744,744	331,544	161,226	259,645	200,774	128,818	187,293	190,833	126,197
Ghana	1984	10	1,309,352	1,309,351	1,050,813	747,642	302,953	142,526	340,180	219,918	121,103	174,954	182,110	110,717
Ghana	2000	10	1,894,133	1,894,133	1,730,902	1,152,128	434,882	200,000	489,201	309,485	167,556	309,535	282,693	161,277
Ghana	2010	10	2,466,289	2,466,289	2,262,894	1,575,528	603,020	270,162	636,999	424,323	229,128	404,083	380,287	217,963
Guinea	1983	10	457,837	457,778	364,805	275,065	99,816	44,129	120,722	68,409	33,806	61,363	57,736	32,395
Guinea	1996	10	729,071	727,246	551,619	397,137	148,064	69,165	202,166	113,623	58,386	90,316	87,334	52,520
Kenya	1969	6	659,310	659,310	659,310	394,835	167,003	67,260	107,690	73,634	42,565	74,096	68,169	41,487
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	220,398	167,963	103,453	168,343	163,583	102,768
Kenya	1999	5	1,407,547	1,407,547	1,191,268	832,083	378,922	176,867	293,668	229,678	137,385	237,572	224,347	136,227
Kenya	2009	10	3,841,935	3,841,935	3,402,695	2,246,737	955,548	432,424	779,027	599,114	354,687	583,560	560,842	344,093
Lesotho	1996	10	187,795	187,795	165,960	121,446	50,160	24,283	57,938	40,652	21,167	41,310	37,454	20,514
Lesotho	2006	10	180,208	180,208	171,947	123,644	50,609	22,361	52,223	37,257	18,405	39,656	34,511	17,796
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	86,523	60,197	32,411	58,878	55,981	31,437
Malawi	1987	10	798,669	798,193	657,998	447,247	176,370	81,029	122,477	89,555	55,837	101,627	88,444	55,755
Malawi	1998	10	991,393	991,393	826,197	582,694	251,873	114,846	170,361	131,703	83,465	140,075	128,509	82,947
Malawi	2008	10	1,341,977	1,341,046	1,161,773	736,175	307,167	135,833	216,020	170,362	108,154	179,356	166,624	107,806
Mali	1987	10	785,384	773,407	582,678	422,837	162,820	76,364	191,000	117,544	63,569	106,596	104,041	61,368
Mali	1998	10	991,330	986,822	734,156	519,001	207,852	102,961	239,952	151,320	85,538	126,625	132,131	81,840
Mali	2009	10	1,451,856	1,424,140	1,262,277	776,333	326,105	158,458	366,650	242,131	135,111	174,224	198,373	124,732
Morocco	1982	5	1,012,873	1,012,873	948,008	571,980	242,307	115,031	294,399	204,348	106,327	210,573	201,143	106,113
Morocco	1994	5	1,294,026	1,293,171	1,293,171	842,330	322,163	149,529	447,785	283,778	141,311	308,463	269,180	138,034
Morocco	2004	5	1,482,720	1,481,076	1,481,076	1,052,531	363,627	161,892	563,135	328,002	155,722	406,132	314,978	153,357
Mozambique	1997	10	1,551,517	1,550,505	1,248,483	879,255	370,427	167,753	276,001	205,248	126,675	200,488	188,002	121,938

Table B.1: Sample construction, continued

(1) country	(2) year	(3) fraction	(4) N_{all}	(5) N_{age}	(6) N_{owned}	(7) $N^{age\geq14}_{owned}$	(8) $N^{14\leq age\leq25}_{owned}$	(9) $N^{14\leq age\leq18}_{owned}$	(10) $N^{age\geq14}_{olded}$	(11) $N^{14\leq age\leq25}_{olded}$	(12) $N^{14\leq age\leq18}_{olded}$	(13) $N^{age\geq14}_{olded,\ no\ mghh}$	(14) $N^{14\leq age\leq25}_{olded,\ no\ mghh}$	(15) $N^{14\leq age\leq18}_{olded,\ no\ mghh}$
Mozambique	2007	10	2,047,048	2,047,048	1,616,853	1,103,596	439,299	193,512	340,331	250,911	150,964	252,302	231,567	145,429
Nigeria	2006	.06	83,700	83,700	82,740	49,282	18,063	8,803	20,096	13,901	7,991	11,172	12,308	7,667
Nigeria	2007	.06	85,183	85,182	84,122	49,102	18,013	8,811	20,104	13,875	7,990	11,727	12,389	7,709
Nigeria	2008	.07	107,425	107,425	105,944	62,151	23,183	11,453	26,700	18,548	10,570	15,708	16,985	10,279
Nigeria	2009	.05	77,896	77,880	77,650	45,988	16,676	8,050	17,545	12,180	6,815	10,302	10,547	6,482
Nigeria	2010	.05	72,191	71,991	58,973	41,830	15,485	7,534	17,736	12,400	6,857	10,988	11,639	6,748
Rwanda	1991	10	742,918	742,918	535,602	372,386	146,839	71,287	121,757	98,452	59,490	101,356	93,882	58,192
Rwanda	2002	10	843,392	843,392	629,146	473,714	221,106	109,367	160,426	139,716	87,351	137,927	132,864	85,472
Rwanda	2012	10	1,038,369	1,038,369	938,201	624,155	250,162	112,248	206,411	169,300	95,112	176,081	161,473	93,342
Senegal	1988	10	700,199	699,981	527,462	378,289	153,541	68,971	200,537	118,387	58,295	100,231	104,889	56,535
Senegal	2002	10	994,562	994,562	911,891	594,599	260,317	124,706	374,034	223,678	112,749	150,679	178,191	101,982
Sierra Leone	2004	10	494,298	492,922	395,788	291,916	120,773	55,346	148,389	92,052	48,263	71,802	72,943	43,093
South Africa	1996	10	3,621,164	3,578,019	3,055,995	2,328,067	840,077	376,601	883,678	594,898	308,241	614,616	538,065	295,656
South Africa	2001	10	3,725,655	3,725,655	3,353,684	2,598,672	915,973	421,066	1,014,988	673,692	351,871	686,654	607,967	336,252
South Africa	2007	2	1,047,657	1,047,657	842,103	665,305	233,345	105,048	268,806	177,267	88,805	181,994	158,798	84,647
South Africa	2011	8.6	4,418,594	4,418,594	3,845,633	3,101,908	1,020,126	422,182	1,104,821	709,491	347,343	790,453	641,430	331,670
South Sudan	2008	7	542,765	542,765	542,333	295,567	120,639	57,922	132,069	89,636	49,765	91,682	83,835	48,071
Sudan	2008	16.6	5,066,530	5,066,530	3,902,071	2,790,992	1,197,729	560,663	1,207,911	843,100	476,833	831,591	799,231	466,630
Tanzania	1988	10	2,310,424	2,304,474	1,911,308	1,322,841	556,836	278,218	483,701	336,938	210,110	349,716	325,632	208,887
Tanzania	2002	10	3,732,735	3,732,735	3,123,724	2,190,557	903,114	416,283	684,743	479,128	293,410	482,029	446,969	284,588
Tanzania	2012	10	4,498,022	4,498,022	3,918,823	2,603,099	1,036,707	491,497	897,469	641,322	382,527	622,879	586,677	366,998
Togo	1960	10	13,759	13,759	13,758	7,842	3,329	1,374	2,930	1,619	775	1,293	1,265	704
Togo	1970	1	23,680	23,617	23,609	12,262	4,146	1,627	4,186	2,216	1,113	2,440	1,901	1,042
Togo	2010	10	584,859	584,859	517,900	339,447	132,399	58,429	131,016	85,981	46,328	98,071	80,277	45,213
Uganda	1991	10	1,548,460	1,547,604	1,242,885	855,537	378,505	179,263	282,250	205,997	128,937	215,330	199,233	127,667
Uganda	2002	10	2,497,449	2,497,449	2,042,838	1,355,857	601,101	289,123	438,634	337,594	223,066	328,581	319,184	217,556
Zambia	1990	10	787,461	787,461	664,239	460,486	216,756	108,294	209,029	157,907	94,631	159,787	155,451	94,493
Zambia	2000	10	996,117	996,117	825,110	570,022	259,096	119,089	235,708	177,260	102,847	151,980	151,774	95,117
Zambia	2010	10	1,321,973	1,321,973	1,028,628	704,471	307,786	147,933	263,135	205,511	126,500	178,608	177,748	117,433
Zimbabwe	2012	5	654,688	653,276	587,010	397,356	157,602	74,305	120,052	86,481	52,191	89,333	79,337	49,874
total			117,902,003	117,764,561	93,086,552	66,791,915	26,575,313	12,238,046	25,805,626	18,096,111	10,085,172	18,018,282	16,819,190	9,788,011

The table shows how we proceed from the raw IPUMS data to the sample employed in our analysis. Columns (1) and (2) give the country and census year, respectively. Column (3) reports coverage, the fraction of the census obtained by IPUMS. Column (4) gives the total number of observations in the IPUMS data. Column (5) gives the number of observations with information on individual's age. Column (6) gives the number of observations for which IPUMS reports both the age and education. Column (7) gives the number of observations/individuals, older than 14 years old with available information on age and education. Column (8) gives the number of observations/individuals in the 14-25 age range with available information on age and education, while column (9) reports the corresponding number for individuals in the 14-18 age range. Columns (10)-(12) are similar to (7)-(9), but also require that IPUMS reports the age and education of at least one member of the older generation. Columns (13)-(15) are similar to (10)-(12), but omit individuals residing in multigenerational households.

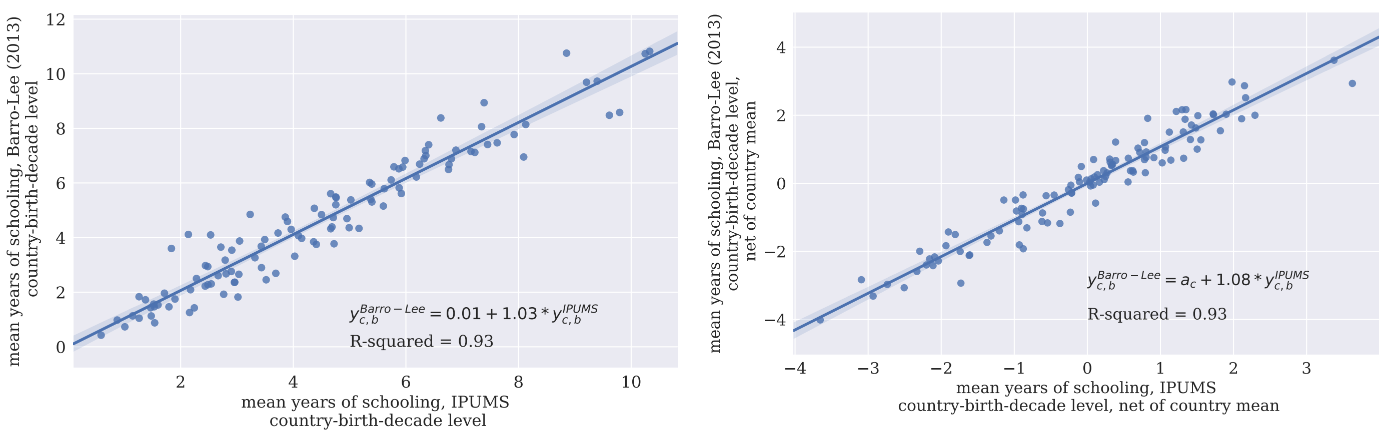
C Data Validation and Relevance of Education

Section C.1 validates the IPUMS data across country-cohorts with the widely-used Barro and Lee (2013) statistics and at the regional level using 109 geo-referenced Demographic and Health Surveys from 22 countries. Section C.2 reports the correlational analysis of a variety of "good outcomes" and educational attainment. The correlational analysis uses all georeferenced DHS surveys ($3,457,367$ observations from 134 surveys in 33 countries) and all geo-referenced rounds of the Afrobarometer surveys ($246,723$ observations from 37 countries).

C.1 Validation IPUMS Education Data

Figure C.1: IPUMS and Barro-Lee (2013) Years of Schooling

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



This figure plots years of schooling at the country-birth-decade level derived from Barro and Lee (2013) on the equivalent measure computed from the IPUMS data for all cohorts since 1940 for individuals aged 20+, as this is age reported by Barro and Lee (2013). The figure also reports the OLS regression fit.

Table C.1: IPUMS and DHS years of schooling, 1980s birth cohort

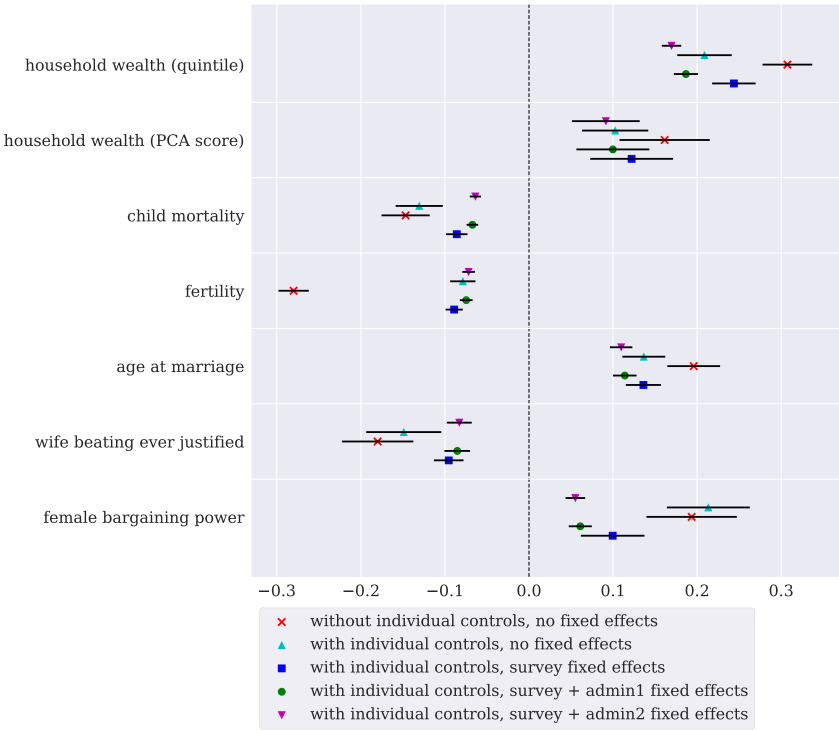
	(1)	(2)	(3)
	ipums	ipums	ipums
dhs	0.986***	0.997***	0.948***
	(0.0446)	(0.0256)	(0.0270)
level	country	province	district
R-squared	0.861	0.933	0.895
N	22	306	2335

This table shows results for regressions of mean years of schooling computed for individuals aged 18+ for three different levels of aggregation – country, province (admin-1) and district (admin-2). The LHS variable in all regressions is the variable computed from IPUMS census data. The RHS variable is the corresponding measure computed from DHS survey data. Column (1) shows the country-level, column (2) the province level, and column (3) the district-level result. Standard errors [clustered at the country-level in columns (2) and (3)] in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

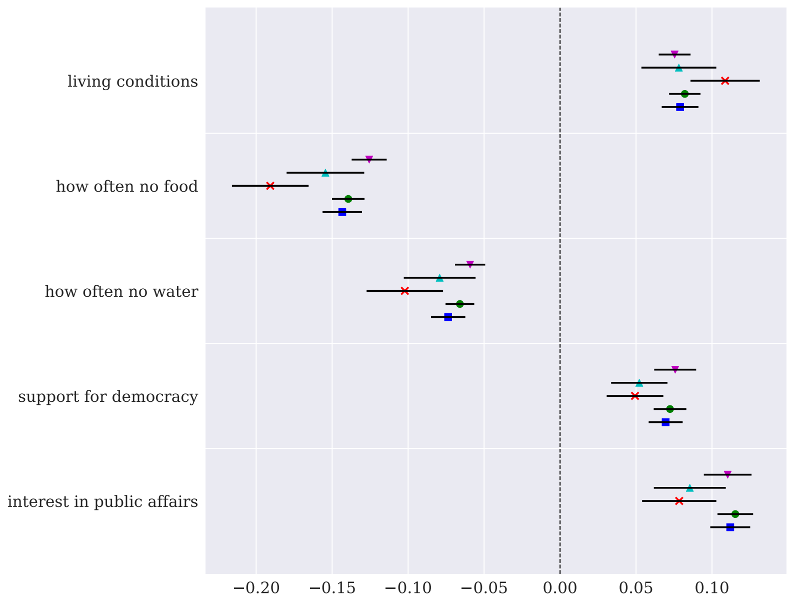
C.2 Education and "Good Outcomes"

Figure C.2: Correlation of at-least primary educational attainment and good outcomes

(a) DHS



(b) Afrobarometer



The figures report OLS regression results (standardized coefficients) associating various outcomes reported in DHS (panel (a), 3,457,367 observations, 134 surveys, 33 countries) and Afrobarometer surveys (panel (b), 246,723 observations, rounds 1-7, 37 countries) on an indicator variable that takes the value of one for individuals who have completed at least least primary educational attainment. The figure also plots 95% confidence bands, based on double-clustered standard errors at the survey (=country×year) and at the admin-1 levels. In the DHS specifications (panel (a)) the set of individual controls consists of age, age-squared, the natural logarithm of the number of household members, a gender dummy, an indicator for male household head, and an urban location dummy. In the Afrobarometer Survey specifications (panel (b)) the set of individual controls consists of age, age-squared, a gender dummy, and an urban location indicator.

D Household Structure and Cohabitation Selection

In this Section we describe the estimation of cohabitation rates and, alongside, we provide an overview of the household structure across countries and censuses. We estimate intergenerational mobility (IM) for boys and girls ("children") between ages 14 and 18. We choose this age interval because children in this age range (i) have most likely completed primary schooling –the relevant educational cutoff for Africa in the period we study (Figure A.1)– and (ii) still largely cohabitate with their parents and other older generation relatives (see Card, Domnisoru, and Taylor (2018) for a similar approach). We also compile IM statistics for (young) individuals aged 14-25, as this doubles the sample size, while cohabitation rates remain still reasonably high.

D.1 Household Structure

To estimate IM we need to observe the educational attainment of the young generation ($14 - 18$ or $14 - 25$-year-old) and that of the immediately older generation in a given household.

The starting point for the intergenerational matching is census information for each individual's "*relationship to household head*" (we denote the head's generation as $G(0)$). Across the different censuses this variable can take as many as three dozen distinct values including: children, parents, grandparents, nieces and nephews, grandchildren, siblings, in-laws, etc. Some censuses provide fine classifications, distinguishing between biological children, adopted, foster, and step-children, as, for example, the post-1995 censuses in South Africa, Cameroon, Burkina Faso, Botswana, and Zambia. However, most censuses have a single "children" category. Likewise, some censuses distinguish clearly between nieces/nephews, uncles/aunts, cousins, sibling in-laws, in-laws whereas others report a composite "*other relatives (not elsewhere classified)*" category that subsumes various family relationships.[1] Illustrative examples include the 1976 and the 1987 Censuses in Cameroon that, beside the household head, only identify clearly the spouse, children, parents, and then all other possible family relationships appear in a single "*other relatives*" category. In an even more extreme case, the 1991 census in Botswana and the 1988 census in Senegal report a single category of "*other relatives and non relatives*".

Using the "*relationship to household head*" we assign household members to generations.

- $G(-2)$, that we rarely observe, consists of the household head's grandparents and grand-uncles/aunts.
- $G(-1)$ consists of the household head's parents, parents-in-law, and uncles/aunts (in-law).
- $G(0)$ consists of the household head, typically male, as well as his spouse, siblings, cousins, siblings and cousins in-law (in the censuses with fine disaggregation).
- $G(1)$ consists of children (biological, adopted, etc.) as well as nieces and nephews.

[1] The level of disaggregation also differs among non-relatives who are part of the household. These individuals cannot be matched and so count as not co-resident.

- $G(2)$ consists of the head's grandchildren (and in a few occasions grandnephews/nieces).

Individuals classified as "*other relatives*" cannot be unequivocally assigned to generations; the prevalence of this category depends (partially) on the census-specific detail. One extreme case is that of the 2010 Census in Togo, where half of $14-18$ year old are classified as "*other relatives*". To avoid misclassification due to differences in census coarseness, we use the age of those classified as "*other relatives*" to assign them to generations. Specifically, we count individuals as being one generation apart if they are in-between 15 and 40 years older.

D.2 Household Arrangements for 14-18 Year Old

Table D.1 reports the relationship to household head for all $14-18$ years old individuals in each census. In total there are $13,005,949$ individuals. The first column reports the percentage of those between 14 and 18 years of age that we can assign to a generation using only censuses' classification on "*relationship to household head*", i.e., without using their age. The remaining columns give the breakdown across generations ($G(0)$, $G(1)$, $G(2)$), further distinguishing by the main categories of "*relationship to household head*". The last row gives the pan-African averages (across 13 million observations). The following patterns emerge.

First, about 70% are children of the household head, and so belong to generation $G(1)$. An additional 1% are nieces or nephews (of the household head).

Second, 4.8% are grandchildren of the household head, $G(2)$; approximately half of them cohabitate with one of their parents, who is the son or daughter of the household head (not reported). Some of the grandchildren also reside with an uncle, aunt, or another relative of the older generation, $G1$.

Third, a few, around 2%, of $14-18$ year old, almost exclusively men, appear as household heads. An additional 4.8% are spouses, reflecting the low age of female marriage in some countries, such as Mali, Burkina Faso, Mozambique, Cameroon, and Guinea. For about 10% of these individuals (who are in $G(0)$), at least one of their parents lives in the household, ($G(-1)$).

Fourth, 10% are classified as "other relatives". As such we cannot directly assign them to generations; they could, for example, be nieces/nephews or son/daughter in laws, but also young cousins, siblings, etc.

Fifth, around 2.5% are unrelated to the household head. 1% are "friends", "employees", "visitors", or live in "group quarters". 1.5% are assigned to a catch-all "non-relatives" category.

D.3 Cohabitation Rates

We then estimate cohabitation rates for each country-census. Table D.2 gives the estimates. Column (1) reports the number of observations. In total, we have about 13 million $14-18$ year old across the 68 censuses spanning 26 countries plus the 5 household surveys from Nigeria. The sample jumps to $28,154,990$ for the sample of $14-25$ years old. Column (2) shows that 94.5% of $14-18$ year olds cohabitate with some relative(s); the corresponding share is 92.6% in the larger sample.

Column (3) shows the share of $14-18$ year old individuals, who cohabitate with at least one relative of the immediately older generation (assigned either with the "relationship to head" classification or using the age of "other relatives"). 83.8% of $14-18$ year olds cohabitate with an older generation relative.[2] Coresidence rates exceed 90% in Egypt, Morocco, and Nigeria. For most censuses cohabitation rates hover between 80% and 90%. As the quality of census information is worse in the early decades, the lowest cohabiting rate is in Kenya in 1969 (63.3%); in the 1989, 1999, and 2009 Kenyan censuses cohabitation rates hover around 80%. For all other censuses cohabitation rates exceed 69%. Cohabitation rates in the $14-25$ sample are, unsurprisingly lower, on average 69%. Setting aside the Kenyan census of 1969, cohabitation rates for $14-25$ year old individuals range from about to 55% to 85%.

Appendix Figures D.1 (a)-(b) zoom on the family arrangements of young individuals who reside with at least one relative from an older generation, i.e., the sample on which IM rates are constructed. Roughly half (51% and 47%) cohabitate with both their mother and father and an additional $8\%-9\%$ coreside with both parents and other relatives of the immediately older generation (e.g., uncles and aunts). 25% cohabitate with one parent, mostly the mother, though often other older generation relatives are present. About $16\%-20\%$ live with other relatives only.

Appendix Figure D.2 (a) reports the distribution of cohabitation rates across the 27 countries for the $14-18$ age sample. Cross-country cohabitation rates average 82.5%; the standard deviation is quite small. The countries with the highest cohabitation rates are Egypt (96.5%) and Morocco (94.5%) while the countries with the lowest cohabitation rates are Zimbabwe (70.5%) and Botswana (72.9%).

Appendix Figure D.2 (b) plots the distribution of cohabitation rates across $2,846$ districts (mostly admin-2 units), pooling across all censuses. The mean (median) is 82.6% (82.5%). The standard deviation is less than a tenth of the mean, 0.09. While there are some outliers, mostly in regions with limited coverage, the $p10-p90$ range is $0.71-0.95$.

Appendix Figure D.2 plots the evolution of cohabitation rates across districts for the 1970, 1980, and 1990 cohorts. Panel (a) looks at an unbalanced sample of districts. Panel (b) looks at a balanced sample of $1,760$ districts from 17 countries for which we have data for all three cohorts. The average coresidence rate increases slightly from 0.804 for the 1970 cohort, to 0.82 for the 1980 cohort and to 0.855 for the 1990-cohort. The increase in the median is smaller (from 0.82 to 0.855), as in the earlier cohorts there are some outliers. The standard deviation of cohabitation rates also falls and over time the distribution becomes less skewed.

[2] As a reference point, Card, Domnisoru and Taylor (2018) report coresidence rates for African Americans and whites in the US 1940 census of about 78% and 89%, respectively.

Table D.1: Relationship to head shares by census and generation, ages 14-18

			G(0)						G(1)							G(2)		G(none)							
iso	year	assignable	Head	Spouse/partner	Sibling/sibling-in-law	Sibling	Sibling-in-law	Cousin	Child	Biological child	Adopted child	Stepchild	Child-in-law	Nephew/niece	Foster child	Grandchild	Grandchild or great grandchild	Other relative, n.e.c.	Other relative or non-relative	Non-relative, n.e.c.	Friend/guest/visitor/partner	Visitor	Domestic employee	Group quarters	Unknown
NGA	2010	99.3	0.3	5.2		1.5	0.3			87.4	0.1	0.6		0.7		3.0		0.2		0.1			0.3		0.2
NGA	2006	99.2	0.5	7.5		2.0	0.8			83.6		0.5		0.7		3.4		0.4		0.1			0.3		
NGA	2008	99.0	0.4	6.0		1.4	0.8			86.1		0.6		0.7		2.7		0.6		0.1			0.4		0.0
NGA	2009	98.9	0.4	7.5		1.5	0.5			84.9		0.7		0.8		2.6		0.5		0.1			0.4		
EGY	1996	98.9	0.6	2.0		2.8			91.5				0.6			1.4		0.8		0.2		0.0	0.1		
NGA	2007	98.6	0.7	6.5		2.0	0.9			83.3		0.7		1.1		3.4		0.6		0.1			0.5		0.2
EGY	2006	98.3	0.3	1.4		1.6			93.7				0.2			1.0		0.5		0.2		1.0	0.1		
EGY	1986	98.1	1.2	2.8		3.6			87.7				1.2			1.6		1.4		0.1		0.1	0.2		0.0
ZMB	2010	95.3	0.7	4.2		4.9		1.1		63.5		2.5	1.3	8.1		9.0		3.7		1.0					
MAR	2004	95.2	0.3	1.0	2.1				85.2				1.8			4.7		2.7		1.3			0.8		
MAR	1994	93.9	0.7	1.2		4.0			81.8				1.9			4.2		2.6		3.5					
SDN	2008	93.1	2.2	9.2	3.3				74.0				0.9	1.6		2.0		3.6		3.3					
ZMB	2000	92.4	1.0	5.4		6.2				62.0		2.5	1.7	7.1		6.5		6.2		1.4					
RWA	1991	92.2	1.9	2.9		3.2			75.5					1.7		6.9				7.5					0.3
ZAF	1996	91.1	2.7	1.0		4.9			66.7							14.8		4.2		1.5				2.3	0.9
ETH	2007	91.0	2.9	6.6	3.8				72.6					2.5		2.7		4.7		4.3					
GHA	1984	90.6	2.1	2.6		6.7		0.6	60.7				1.3	7.2		9.3		4.2	0.0	1.7				3.4	
MLI	2009	89.9	1.1	10.3		4.6			63.0				1.0	6.9		2.8		5.3		1.9			2.9		
RWA	2012	89.8	1.2	0.9		2.1			77.3						1.3	7.0		2.5		7.6					
ZAF	2001	89.6	3.0	0.5		5.6	0.7			58.7	1.9	1.1	1.0				17.1	7.5		0.9				2.0	
BFA	2006	88.2	1.5	11.0		6.4			59.4					6.5		3.0		6.4		3.9					1.5
KEN	2009	88.1	1.8	2.8		3.8				68.9				3.2		7.7		6.3		3.5				2.1	
BEN	2013	87.9	2.1	4.0		5.3				65.1	4.2		2.4			4.7		7.9		4.2					
CMR	2005	86.6	2.2	6.5		7.8	1.0		1.3	61.2		0.6	0.7	2.4		2.6		10.7		2.6					0.0
BWA	2001	86.3	5.3	0.3		8.8				46.9		0.5	0.1	7.5		16.9		9.4		3.8				0.5	
ZAF	2011	86.1	2.3	1.3		4.8	0.5			56.2	0.8	1.2	0.7				18.1	10.2		1.7				2.0	
GHA	2010	85.5	2.2	0.9		6.1				63.2	0.8	1.5	0.7			10.1		9.4		2.7				2.5	
ETH	1984	85.2	3.5	12.6	4.3					63.3		1.5						10.1		4.6					0.0
LSO	1996	85.0	1.1	0.6					69.8				2.8				10.5	9.6		3.1			2.3		
MAR	1982	84.9	0.8	2.2					79.1				2.8					12.6		2.5					0.0
MOZ	2007	84.6	3.7	9.4						58.5		3.6	2.2			7.2		13.9		1.6					
ETH	1994	84.4	1.6	7.9	3.9					69.8		1.2						9.4	0.0	6.2					
BFA	1996	84.3	1.4	11.4		6.6		1.6	57.7					5.5				11.6		0.3					3.8
ZAF	2007	84.1	1.8	0.4		4.4	0.6			54.0	1.6	1.2	0.7				19.4	7.2		0.7				8.0	
SLE	2004	83.9	1.2	4.4		7.8			53.5					9.8		7.0		12.9		3.1					
BWA	1991	82.7	6.5	0.4		8.8			46.3					7.0		13.7		1.6	15.7						0.0
MLI	1998	82.2	0.9	12.9		4.3			64.0									12.6		5.0					0.2
RWA	2002	82.1	2.1	1.6		4.7			67.5						0.7	5.5		5.8		5.6		5.2		0.4	0.9
LSO	2006	81.9	1.4	1.1					62.4				2.2				14.8	13.6		1.0			3.6		
SSD	2008	81.9	1.9	3.0	5.3				64.4				1.6	4.0		1.6		14.1		4.0					
MOZ	1997	80.6	4.4	10.6					57.5				2.3			5.9		17.1		2.3					
MLI	1987	79.6	1.2	14.2		4.1			60.0									16.7		3.6					0.1
GIN	1996	79.6	0.7	12.1		5.8			49.8					7.4		3.7		13.7		6.1					0.6
BWA	1981	78.0	1.8	0.6		5.6			52.0					5.6		12.4		10.9		8.6		1.1		1.4	
UGA	2002	77.8	2.6	6.0	5.9				59.2			4.1						17.7		2.8					1.6
BEN	2002	76.9	1.8	5.4					60.0						4.9	4.1		16.4		6.7					
TZA	2012	75.5	2.1	3.2					61.0							9.3		14.8		9.6					
KEN	1989	75.5	2.6	3.3						69.4								19.3		4.6				0.6	0.1
BWA	2011	75.3	3.9	0.3		6.6				39.6		0.5	0.5	7.7		15.9		5.9		1.6		0.5		16.8	
ZWE	2012	75.0	4.1	2.8		5.3			48.5							14.4		19.6		4.4				0.8	0.2
ZMB	1990	73.9	0.4	4.2						66.8		2.6						24.8		1.4					
KEN	1999	73.8	3.1	3.1		4.0				63.5								19.6		6.7					0.0
LBR	2008	73.5	1.6	3.2					62.4						1.1	5.2		21.2		3.1			0.5		1.8
BEN	1979	73.0	2.7	9.4		7.0			49.3									18.1			2.3		2.5	0.4	3.8
SEN	2002	71.3	0.3	3.4		4.6			55.8							7.2		22.5		6.2					
GHA	2000	71.0	2.0	1.7					57.7				1.0			8.6		24.3		4.5				0.2	
GIN	1983	70.4	0.6	12.8					57.0									23.8		5.7					0.2
BEN	1992	69.7	1.6	6.8					56.3							3.8		18.8		10.2				0.8	0.5
MWI	2008	69.6	1.7	5.7						62.2								24.5		5.9					
TZA	2002	69.2	3.2	3.8					54.7							7.5		16.7		14.1					
UGA	1991	68.7	3.3	7.6		4.0			53.8									24.6		3.8	1.5			1.3	0.0
MWI	1998	67.5	3.8	7.4						56.3								27.7		4.7				0.1	
SEN	1988	66.9	0.1	4.0		2.7			55.2				0.2	4.7				17.9	11.1	3.8					0.3
CMR	1976	66.6	1.9	14.6					50.1									24.2		6.7					2.5
CMR	1987	64.2	2.8	11.8					49.5									27.9		5.8				2.1	
MWI	1987	64.1	4.5	8.1					51.5									30.0		5.2				0.7	
TZA	1988	63.7	3.2	4.3					56.1									29.0		7.3					0.0
KEN	1969	60.9	5.7	6.6					46.8							1.8		20.2		18.9					
TGO	2010	45.9	3.2	3.0		6.9			29.9							2.9		48.7		5.4					
total		85.3	2.0	4.8	0.9	2.3	0.1	0.0	48.0	20.2	0.2	0.4	0.5	1.0	0.0	3.4	1.4	10.0	0.1	3.7	0.0	0.1	0.1	0.5	0.2

This table gives an overview of the values of the "relationship to household head" classification for individuals aged 14-18 across all censuses. The table groups the values classifications across "generations" within the household: "G(0)" refers to the head's generation (e.g., spouse, siblings); "G(1)" to the generation of the children of the head (some censuses distinguish explicitly biological and other children); and "G(2)" to the generation of the grandchildren of the head. The table also reports the fraction of individuals who cannot be placed on the "generation-ladder" ["G(none)"]; these consist of "relatives (not elsewhere classified)", "employees", and "non-relatives". The table sorts the censuses by the fraction of observations that can be "assigned" to a generation using solely the "relationship to household head" classification.

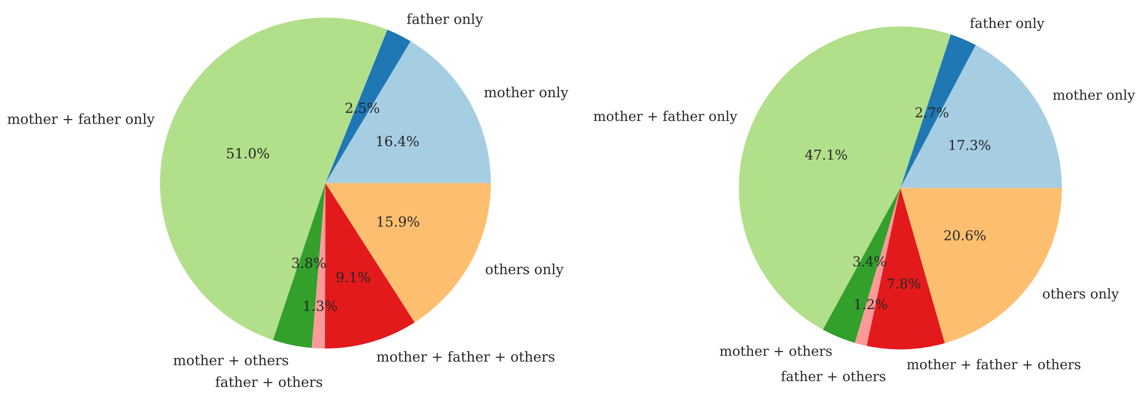
Table D.2: Cohabitation rates by census, ages 14-18 and 14-25

		Ages 14-18			Ages 14-25		
		(1)	(2)	(3)	(4)	(5)	(6)
iso	year	N	living with relative	living w/ old or fam mbr 15-40 yrs older	N	living with relative	living w/ old or fam mbr 15-40 yrs older
BEN	1979	24,439	90.0	75.9	62,680	88.0	59.9
BEN	1992	46,233	87.9	77.9	105,192	87.5	64.4
BEN	2002	70,033	92.6	83.0	158,251	90.7	67.0
BEN	2013	108,694	94.9	85.9	240,049	93.5	71.1
BFA	1996	122,125	95.7	84.6	250,759	95.8	71.8
BFA	2006	154,448	94.2	81.6	329,175	94.2	66.1
BWA	1981	9,798	87.9	75.0	21,736	82.2	63.9
BWA	1991	15,831	82.3	69.4	32,681	73.7	55.9
BWA	2001	20,650	93.9	78.3	44,939	87.2	65.2
BWA	2011	20,693	79.4	69.3	49,130	79.0	59.1
CMR	1976	75,676	87.5	73.3	163,629	85.6	60.0
CMR	1987	94,750	90.9	74.8	200,469	89.1	60.7
CMR	2005	203,450	95.6	82.8	449,515	93.6	68.7
EGY	1986	723,853	99.3	96.1	1,611,636	98.7	83.6
EGY	1996	718,875	99.5	96.8	1,471,286	98.9	83.7
EGY	2006	785,676	98.7	96.7	1,978,864	97.8	80.4
ETH	1984	309,641	93.8	76.9	633,338	92.6	58.3
ETH	1994	627,895	93.4	80.7	1,254,576	91.9	64.2
ETH	2007	969,639	94.6	81.5	1,949,874	92.7	63.3
GHA	1984	142,526	93.8	85.0	302,954	91.3	72.6
GHA	2000	200,000	94.2	83.8	434,882	91.8	71.2
GHA	2010	270,162	94.2	84.8	603,020	89.1	70.4
GIN	1983	44,316	84.6	76.9	100,250	84.4	68.8
GIN	1996	71,439	92.8	84.4	153,619	91.8	76.7
KEN	1969	67,260	78.3	63.3	167,003	70.2	44.1
KEN	1989	127,055	94.0	81.9	263,585	89.8	64.2
KEN	1999	176,867	92.5	77.7	378,922	90.0	60.6
KEN	2009	438,897	93.6	82.6	968,723	89.9	63.2
LBR	2008	38,854	94.1	83.4	87,459	92.2	68.8
LSO	1996	24,395	93.9	87.2	50,274	93.7	81.1
LSO	2006	22,361	94.8	82.3	50,625	92.9	73.7
MAR	1982	118,922	97.3	93.6	251,487	96.4	85.5
MAR	1994	149,529	96.4	94.5	322,163	95.3	88.1
MAR	2004	161,892	97.8	96.2	363,627	96.7	90.2
MLI	1987	79,031	95.9	84.0	168,635	95.6	73.0
MLI	1998	113,202	94.6	84.1	225,998	94.8	74.2
MLI	2009	159,705	94.9	85.4	328,411	95.3	74.4
MOZ	1997	172,377	96.7	75.8	380,792	95.5	56.0
MOZ	2007	199,970	97.6	78.4	453,004	96.6	57.8
MWI	1987	81,070	93.0	68.9	176,442	91.7	50.8
MWI	1998	114,846	94.3	72.7	251,873	93.6	52.3
MWI	2008	144,751	93.2	76.3	324,412	93.8	54.2
NGA	2006	8,868	99.2	91.2	18,236	97.1	77.3
NGA	2007	8,886	98.7	91.0	18,194	96.4	77.4
NGA	2008	11,572	99.2	92.9	23,462	97.5	80.6
NGA	2009	8,058	90.7	84.9	16,702	89.0	73.3
NGA	2010	7,914	99.1	93.1	16,452	97.5	81.0
RWA	1991	78,334	91.6	82.7	163,796	90.2	67.2
RWA	2002	117,490	87.2	76.2	235,909	85.7	60.6
RWA	2012	113,386	91.8	84.9	252,856	89.5	67.8
SDN	2008	591,921	96.3	86.4	1,258,800	95.4	71.7
SEN	1988	71,362	88.2	84.9	158,443	85.0	77.6
SEN	2002	124,706	93.8	90.4	260,317	92.8	85.9
SLE	2004	55,764	96.6	87.7	121,925	95.4	76.8
SSD	2008	57,942	95.5	85.9	120,722	94.3	74.3
TGO	2010	59,583	92.9	79.5	136,454	90.9	65.3
TZA	1988	278,539	91.5	75.5	557,552	88.8	60.5
TZA	2002	416,283	84.5	70.5	903,115	82.6	53.1
TZA	2012	491,497	89.7	77.9	1,036,707	87.6	61.9
UGA	1991	179,474	92.0	72.0	381,144	89.7	54.5
UGA	2002	289,123	94.4	77.2	601,101	92.0	56.2
ZAF	1996	384,919	93.3	83.1	865,987	88.8	71.7
ZAF	2001	421,066	96.1	83.6	915,973	92.4	73.6
ZAF	2007	114,829	90.5	78.8	256,019	87.7	70.5
ZAF	2011	431,062	94.2	80.7	1,052,039	88.1	67.6
ZMB	1990	108,294	98.5	87.4	216,756	97.1	72.9
ZMB	2000	119,089	98.4	86.4	259,096	97.3	68.5
ZMB	2010	159,684	98.8	87.6	333,273	97.5	69.4
ZWE	2012	74,478	92.9	70.5	158,021	90.3	55.2
total		13,005,949	94.5	83.8	28,154,990	92.6	69.0

Cohabitation rates by census for all individuals aged 14-18 and 14-25. Columns (1) and (4) report the number of individuals for each census. Columns (2) and (5) report the fraction of individuals living with a relative of any generation. Columns (3) and (6) report the fraction of individuals living with at least one relative of the immediately older generation; these are identified with the “relationship to household head” classification or with the age for “other relatives (not elsewhere classified).

Figure D.1: Family Structures. All Censuses

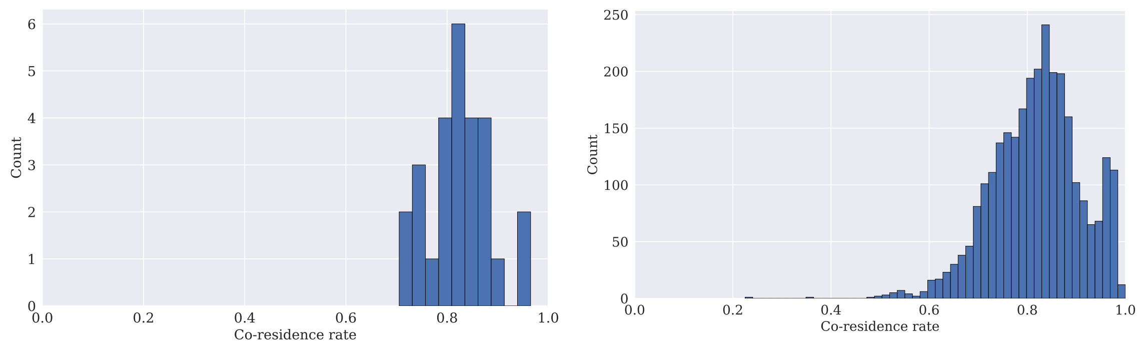
(a) Ages 14-18 **(b)** Ages 14-25



The chart plots the distribution of household structure for individuals aged 14-18 (panel (a)) and 14-25 (panel (b)), who reside with at least one relative from the immediately older generation.

Figure D.2: Country & District Co-habitation Rates, individuals aged 14-18

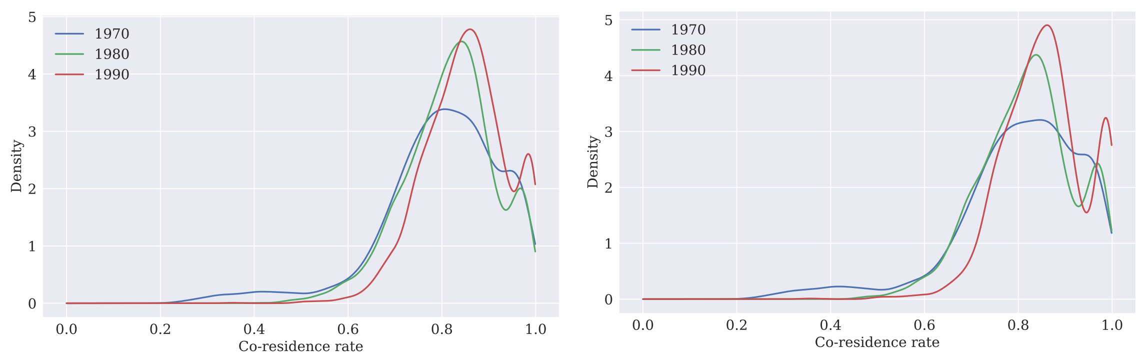
(a) Country **(b)** District



Distribution of co-habitation rates for all individuals aged 14-18 at the country (panel (a)) and at the district (panel (b)) level, pooling observations across all cohorts and census-years. Country: Min=0.705, q10=0.742, median=0.826, q90=0.894, max=965, std=0.06. District: Min=0.225, q10=0.706, median=0.826, q90=0.947, max=1, std=0.09

Figure D.3: District Co-habitation across Cohorts, individuals aged 14-18

(a) All districts **(b)** Balanced Sample



The figure plots the distribution of co-habitation rates for individuals aged 14-18 at the district level across birth-cohorts. Panel (a) uses all birth-cohort-district observations. Panel (b) uses observations from districts with data for all cohorts.

sample	cohort	N	min	q10	q20	median	mean	q80	q90	max	std	skew
all	1970	1,996	0.247	0.658	0.719	0.813	0.8	0.905	0.957	1.0	0.13	-1.246
all	1980	2,352	0.405	0.691	0.74	0.823	0.817	0.894	0.953	0.995	0.096	-0.45
all	1990	2,598	0.355	0.74	0.777	0.852	0.85	0.924	0.977	1.0	0.087	-0.383
balanced	1970	1,760	0.247	0.656	0.723	0.821	0.804	0.919	0.96	1.0	0.134	-1.308
balanced	1980	1,760	0.466	0.69	0.737	0.823	0.819	0.904	0.966	0.995	0.098	-0.305
balanced	1990	1,760	0.355	0.747	0.782	0.855	0.855	0.928	0.988	1.0	0.087	-0.338

E IM Across Countries and Regions

This Section reports additional measures of intergenerational mobility (IM), distinguishing across gender, rural-urban residence, and family structure. It also gives further descriptive evidence.

E.1 IM by Household Structure

We compute upward IM and downward IM for the 1990s cohort, which has the broadest country-coverage, for individuals who cohabitate with "older" generations of different types. In particular, we compute IM for children cohabiting (i) only with biological parents, (ii) only with older relatives other than biological parents (e.g., aunts, uncles), and (iii) with biological parent(s) and other older generation relative(s). For comparability, figure E.1 below also reports the baseline IM measure, estimated for young individuals residing in all types of households. Table E.1 gives the cross-country correlations of IM statistics across the different types of family structure.

Figure E.1: IM by Household Structure, individuals ages 14-18

(a) up **(b)** down

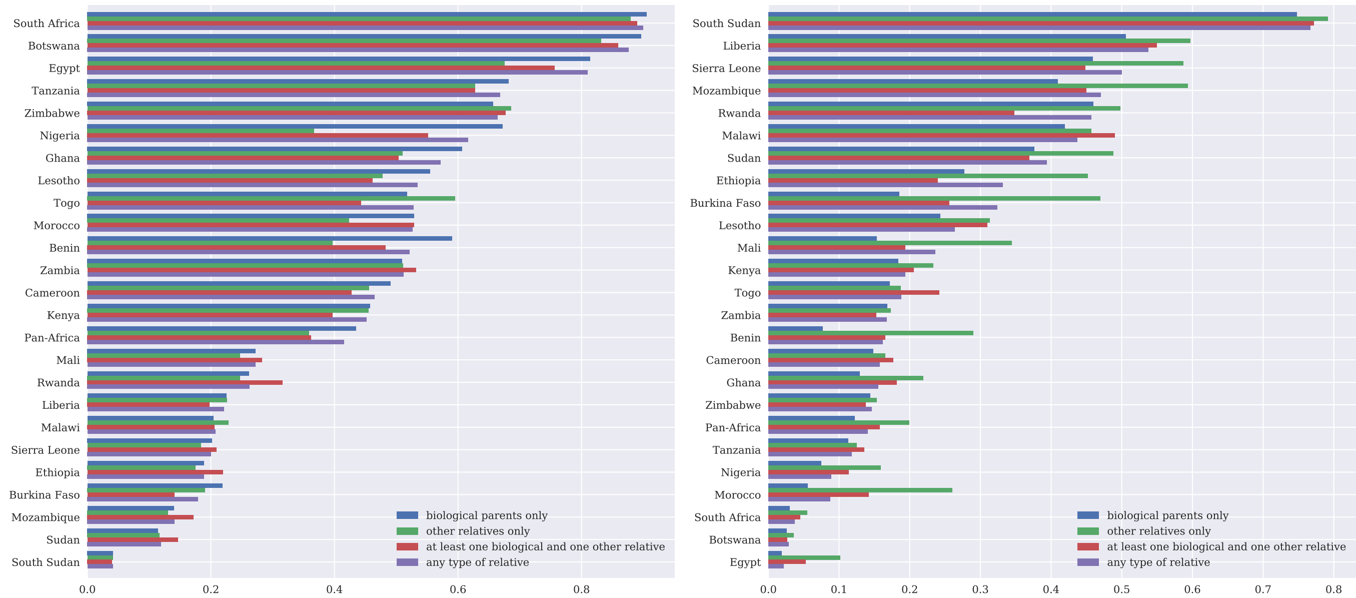


Table E.1: Correlation matrix for country-level IM, ages 14-18, different family structures

Panel A: upward IM

	other only	biogical only	other and biological	all (baseline)
other only	1.0	0.95	0.961	0.966
biogical only	0.95	1.0	0.981	0.997
other and biological	0.961	0.981	1.0	0.989
all (baseline)	0.966	0.997	0.989	1.0

Panel B: downward IM

	other only	biogical only	other and biological	all (baseline)
other only	1.0	0.92	0.927	0.968
biogical only	0.92	1.0	0.971	0.984
other and biological	0.927	0.971	1.0	0.973
all (baseline)	0.968	0.984	0.973	1.0

This table shows a correlation matrix for country-level IM, computed for individuals aged 14-18, 1990s birth cohort, who live in different family arrangements. Panel A shows the results for upward IM, panel B for downward IM.

E.2 IM across Rural and Urban Households

Appendix Table E.2 reports upward IM and downward IM for rural and urban households and the associated urban-rural gap. Appendix Figure E.2 plots the cross-country association between the urban-rural gap and economic development.

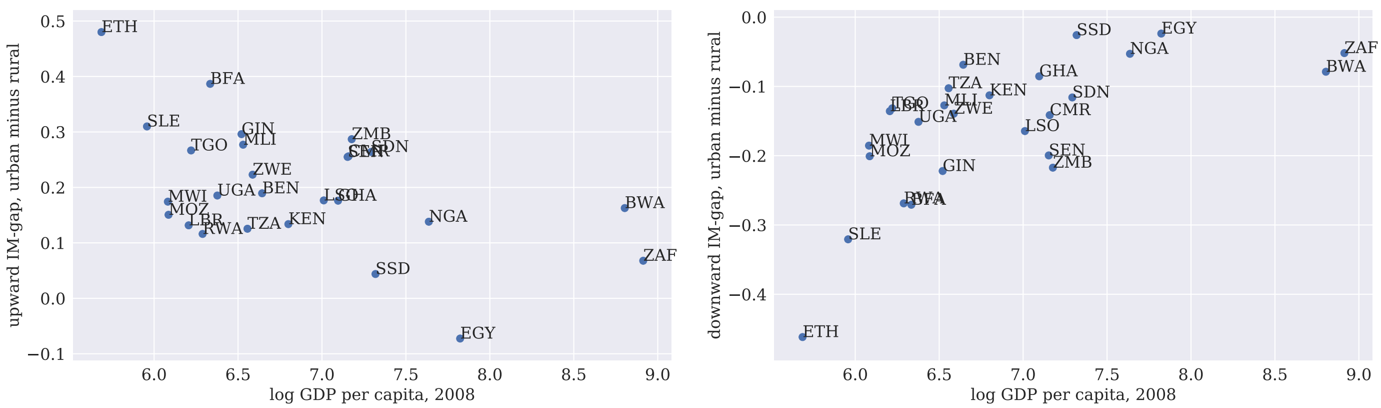
Table E.2: Country-level estimates of intergenerational mobility (IM), urban/rural

		(1)	(2)	(3)	(4)	(5)	(6)
mobility	census years	upward	upward	downward	downward	upward	downward
location		urban	rural	urban	rural	urban - rural	urban - rural
Egypt	1986, 1996, 2006	0.593	0.665	0.06	0.083	-0.073	-0.024
South Sudan	2008	0.076	0.032	0.754	0.78	0.044	-0.026
South Africa	1996, 2001, 2007, 2011	0.838	0.77	0.05	0.102	0.068	-0.052
Rwanda	2002, 2012	0.327	0.211	0.304	0.573	0.116	-0.269
Tanzania	2002, 2012	0.695	0.57	0.116	0.218	0.125	-0.103
Liberia	2008	0.305	0.174	0.502	0.638	0.132	-0.136
Kenya	1989, 1999, 2009	0.586	0.452	0.14	0.253	0.134	-0.113
Nigeria	2006, 2007, 2008, 2009, 2010	0.745	0.606	0.05	0.103	0.138	-0.053
Mozambique	1997, 2007	0.213	0.063	0.481	0.682	0.151	-0.201
Botswana	1991	0.769	0.606	0.11	0.189	0.163	-0.079
Malawi	1987, 1998, 2008	0.315	0.141	0.377	0.562	0.174	-0.185
Ghana	2000, 2010	0.681	0.505	0.133	0.218	0.176	-0.085
Lesotho	1996, 2006	0.595	0.418	0.182	0.346	0.177	-0.164
Uganda	1991, 2002	0.534	0.349	0.2	0.351	0.185	-0.151
Benin	1992, 2002, 2013	0.509	0.32	0.213	0.282	0.189	-0.069
Zimbabwe	2012	0.867	0.644	0.061	0.201	0.223	-0.139
Senegal	2002	0.415	0.16	0.236	0.435	0.255	-0.2
Cameroon	1987, 2005	0.707	0.451	0.076	0.217	0.256	-0.141
Sudan	2008	0.348	0.084	0.351	0.467	0.264	-0.116
Togo	2010	0.719	0.452	0.144	0.275	0.267	-0.132
Mali	1998, 2009	0.423	0.146	0.239	0.366	0.277	-0.127
Zambia	1990, 2000	0.689	0.402	0.148	0.365	0.287	-0.217
Guinea	1983, 1996	0.403	0.107	0.366	0.588	0.296	-0.222
Sierra Leone	2004	0.458	0.148	0.301	0.621	0.31	-0.321
Burkina Faso	2006	0.502	0.115	0.254	0.525	0.387	-0.271
Ethiopia	1984, 1994, 2007	0.545	0.065	0.233	0.695	0.48	-0.462
mean		0.533	0.333	0.234	0.39	0.2	-0.156

This table reports upward and downward IM measures (estimated without cohort effects) separately for individuals aged 14-18 and living in urban and rural locations at the time of the census. Columns (1) and (2) give the estimates for upward IM. They reflect the likelihood that children, 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, whose parents have completed primary schooling or higher will not manage to complete primary education. Columns (5) and (6) show the mobility gaps, subtracting the rural from the urban estimate. Countries are sorted from the lowest to the lowest to highest upward IM gap (column (5)). "mean" gives the simple average of the 26 country-estimates (note Morocco has no data on urban/rural status).

Figure E.2: Urban/Rural Gaps in IM and GDP per capita

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



The figures plot the gap in upward IM (panel (a)) and downward IM (panel (b)) between indviduals living in urban and rural households against the logarithm of real GDP per capita in 2008 using World Bank data (2008 is the earliest year for which data for all countries, including South Sudan, is available).

E.3 IM across Gender

Appendix Table E.3 reports upward IM and downward IM for girls and boys and the associated gender gap for all country-census-years. Appendix Figure E.3 plots the cross-country association between the gender gap in IM and economic development (GDP p.c.).

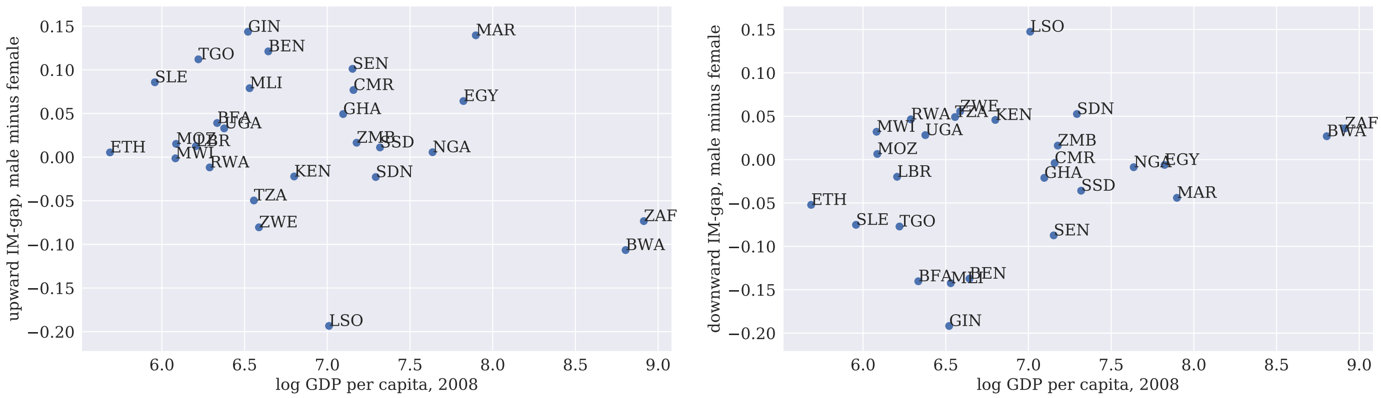
Table E.3: Country-level estimates of intergenerational mobility (IM), male/female

		(1)	(2)	(3)	(4)	(5)	(6)
mobility	census years	upward	upward	downward	downward	upward	downward
gender		male	female	male	female	male - female	male - female
Lesotho	1996, 2006	0.339	0.532	0.365	0.217	-0.193	0.147
Botswana	1981, 1991, 2001, 2011	0.649	0.756	0.083	0.056	-0.107	0.027
Zimbabwe	2012	0.627	0.707	0.174	0.118	-0.08	0.056
South Africa	1996, 2001, 2007, 2011	0.754	0.827	0.086	0.05	-0.073	0.036
Tanzania	1988, 2002, 2012	0.571	0.621	0.201	0.152	-0.05	0.049
Sudan	2008	0.109	0.132	0.42	0.367	-0.023	0.053
Kenya	1969, 1989, 1999, 2009	0.444	0.466	0.242	0.196	-0.022	0.046
Rwanda	1991, 2002, 2012	0.286	0.297	0.496	0.449	-0.012	0.046
Malawi	1987, 1998, 2008	0.154	0.156	0.497	0.465	-0.001	0.032
Ethiopia	1984, 1994, 2007	0.131	0.126	0.273	0.325	0.005	-0.052
Nigeria	2006, 2007, 2008, 2009, 2010	0.632	0.626	0.08	0.089	0.006	-0.009
South Sudan	2008	0.046	0.035	0.75	0.786	0.011	-0.036
Liberia	2008	0.227	0.214	0.528	0.547	0.013	-0.02
Mozambique	1997, 2007	0.118	0.103	0.515	0.509	0.015	0.006
Zambia	1990, 2000, 2010	0.494	0.478	0.208	0.192	0.017	0.016
Uganda	1991, 2002	0.373	0.34	0.325	0.297	0.033	0.028
Burkina Faso	1996, 2006	0.202	0.163	0.19	0.33	0.039	-0.14
Ghana	1984, 2000, 2010	0.589	0.54	0.148	0.169	0.049	-0.021
Egypt	1986, 1996, 2006	0.667	0.603	0.068	0.074	0.064	-0.006
Cameroon	1976, 1987, 2005	0.546	0.469	0.115	0.119	0.077	-0.004
Mali	1987, 1998, 2009	0.242	0.163	0.189	0.331	0.079	-0.142
Sierra Leone	2004	0.291	0.205	0.33	0.405	0.086	-0.075
Senegal	1988, 2002	0.307	0.206	0.198	0.286	0.101	-0.087
Togo	1960, 1970, 2010	0.561	0.449	0.151	0.229	0.112	-0.077
Benin	1979, 1992, 2002, 2013	0.432	0.311	0.16	0.297	0.121	-0.137
Morocco	1982, 1994, 2004	0.484	0.344	0.085	0.129	0.14	-0.044
Guinea	1983, 1996	0.262	0.118	0.304	0.496	0.144	-0.192
mean		0.39	0.37	0.266	0.285	0.02	-0.019

This table reports upward and downward IM measures (estimated without cohort effects) separately for male and female individuals aged 14-18. Columns (1) and (2) give the estimates for upward IM. They reflect the likelihood that children, 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, whose parents have completed primary schooling or higher will not manage to complete primary education. Columns (5) and (6) show the mobility gaps, subtracting the female from the male estimate. Countries are sorted from the lowest to the lowest to highest upward IM gap (column (5)). "mean" gives the simple average of the 27 country-estimates.

Figure E.3: Urban/Rural Gap in IM and GDP per capita

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

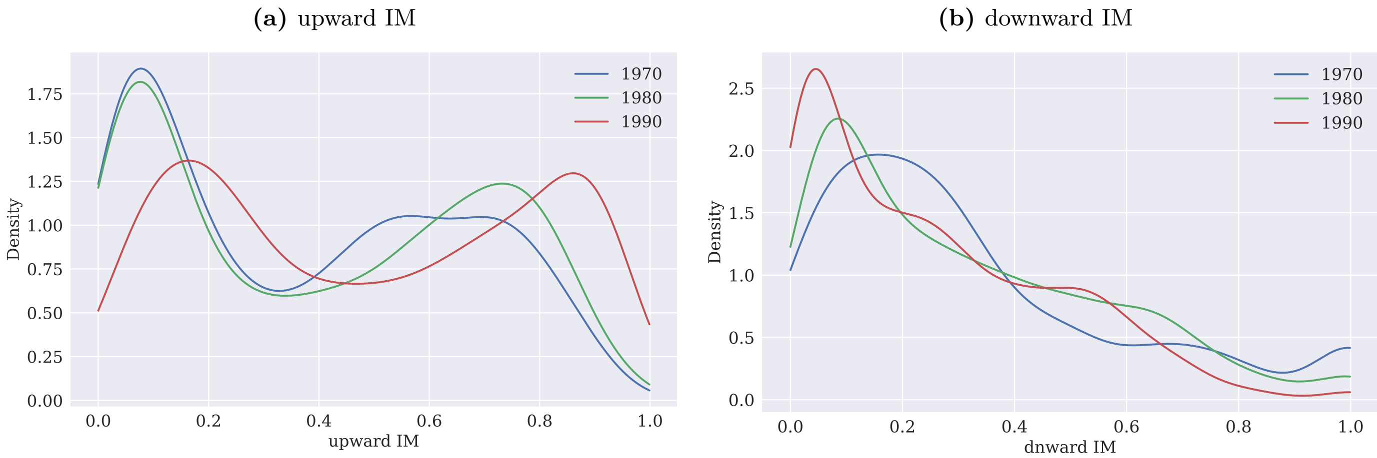


The figures plot the gap in upward IM (panel (a) and downward IM (panel (b)) between male and female young individuals against the logarithm of real GDP per capita in 2008 using World Bank data (2008 is the earliest year for which data for all countries, including South Sudan, is available).

E.4 IM across Cohorts

Appendix Figure E.4 plots the distribution of regional upward IM (panel (a)) and downward IM (panel (b)) for the 1970s, 1980s, and 1990-born cohorts. There are 17 countries with information on the three cohorts; Benin, Burkina Faso, Botswana, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Lesotho, Morocco, Mali, Mozambique, Malawi, Rwanda, Tanzania, South Africa, Zambia. Appendix Table E.4 reports summary statistics.

Figure E.4: Distribution district × cohort level IM, balanced sample



Distribution of district upward (panel (a)) and downward (panel (b)) IM for individuals aged 14-18 for different cohorts. Sample restricted to the districts for which IM is available for all cohorts.

Table E.4: Summary-statistics, district-cohort-level IM, individuals aged 14-18

direction	cohort	count	min	q10	q20	median	mean	q80	q90	max	std	skew
up	1970	1,759	0.0	0.04	0.076	0.363	0.378	0.683	0.769	0.962	0.282	0.22
up	1980	1,759	0.0	0.043	0.075	0.403	0.403	0.725	0.801	0.971	0.298	0.117
up	1990	1,759	0.006	0.098	0.164	0.488	0.49	0.833	0.898	1.0	0.304	0.026
down	1970	1,620	0.0	0.049	0.095	0.25	0.326	0.5	0.75	1.0	0.273	1.084
down	1980	1,620	0.0	0.047	0.077	0.247	0.312	0.545	0.667	1.0	0.252	0.838
down	1990	1,620	0.0	0.02	0.04	0.198	0.251	0.461	0.571	1.0	0.219	0.848

Balanced sample of 17 countries for which we have data for 1970, 1980, and 1990 birth cohort. The number of districts is different for the two directions since not every region in every cohort has recorded in the census an individual with the required configuration (right age range, literate or illiterate parents).

E.5 Heterogeneity. IM and Literacy of the Old Generation

Appendix Table E.5 reports un-weighted within-country regression specifications associating upward and downward IM with the old generation's literacy distinguishing between rural and urban households. The table also reports a Chow test of coefficient equality. The old's literacy - upward-IM correlation is considerately stronger for rural, 0.75 (in (2)), as compared to urban 0.525 (in (1)) households. Columns (5)-(6) replace the country fixed-effects with province constants to account for more localized features. The estimate in the urban sample is 0.40, while in the rural sample it is 0.62, a significant difference. The specifications with downward IM in columns (3)-(4), (7)-(8) yield similar patterns. The slope is twice as large in rural as compared to urban households (0.27 versus 0.53).

Table E.5: Old literacy and IM at the district-level, urban/rural, ages 14-18

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IM up	IM up	IM down	IM down	IM up	IM up	IM down	IM down
share literate old	0.525***	0.746***	-0.270***	-0.532***	0.402***	0.623***	-0.209***	-0.436***
	(0.044)	(0.052)	(0.026)	(0.049)	(0.048)	(0.055)	(0.039)	(0.038)
R-squared	0.750	0.863	0.700	0.764	0.825	0.918	0.774	0.821
N	1997	2633	1982	2618	1997	2633	1982	2618
sub-sample	urban	rural	urban	rural	urban	rural	urban	rural
country FEs	yes	yes	yes	yes	no	no	no	no
province FEs	no	no	no	no	yes	yes	yes	yes
p: coeff-equal	0.0000		0.0000		0.0000		0.0000	

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

Appendix Table E.6 associates – again in an unweighted manner – upward and downward IM to the literacy of the "old" generation at the regional level, distinguishing across boys and girls. The country (province) fixed-effects coefficient in the upward IM specification for boys is 0.66 (0.53), while for girls it is 0.78 (0.62). A Chow test of coefficient equality suggests that the difference is statistically different than zero. Likewise, the downward IM - "old" generation literacy association is somewhat steeper for girls, as compared to boys. The difference is 0.12 in the country fixed-effects estimation, but it drops to 0.047 in the province fixed-effects specifications.

Table E.6: Old literacy and IM at the district-level, male/female, ages 14-18

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IM up	IM up	IM down	IM down	IM up	IM up	IM down	IM down
share literate old	0.657***	0.782***	-0.400***	-0.521***	0.529***	0.623***	-0.397***	-0.444***
	(0.036)	(0.042)	(0.028)	(0.043)	(0.039)	(0.043)	(0.033)	(0.039)
R-squared	0.867	0.892	0.718	0.746	0.921	0.940	0.770	0.800
N	2846	2845	2829	2829	2846	2845	2829	2829
sub-sample	male	female	male	female	male	female	male	female
country FEs	yes	yes	yes	yes	no	no	no	no
province FEs	no	no	no	no	yes	yes	yes	yes
p: coeff-equal	0.0000		0.0028		0.0000		0.0474	

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

F Correlates of IM

This Section complements the correlational analysis (Section 5 of the paper). First, we provide variable definitions and sources for all correlates of regional mobility. Second, we report robustness checks. Third, we discuss and present LASSO estimates.

F.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, Rød, and Thieme (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, Gleditsch, and Gilmore (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, Beusen, and Janssen (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 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 pre-colonial states are obtained combining maps of pre-colonial ethnic homelands in Murdock (1959) with the levels of jurisdictional hierarchy beyond the local community level in Murdock (1967). Societies with 3 or 4 levels are classified as states.

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

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

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

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

F.2 Robustness

Table F.1 shows a table version of the baseline estimates (as in Figure 9) but conditions on province fixed effects instead of country fixed effects.

Table F.2 has country fixed effects as in the baseline estimates, but restricts estimation to districts with at least 80% cohabitation rate.

Table F.3 has country fixed effects as in the baseline estimates, but drops Egypt and Morocco.

Table F.1: District-level correlates of the share of literate old and 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 conditional on share literate old	N		IM	IM conditional on share literate old	N
Panel A: geography								
ln(distance to capital)	-0.318***	-0.267***	-0.079**	2786		0.202***	0.055*	2783
	(0.066)	(0.040)	(0.040)			(0.045)	(0.032)	
ln(distance to border)	0.006	-0.001	-0.005	2786		-0.002	0.001	2783
	(0.025)	(0.018)	(0.012)			(0.022)	(0.017)	
ln(distance to coast)	-0.181***	-0.155***	-0.045*	2786		0.143***	0.058**	2783
	(0.036)	(0.031)	(0.024)			(0.031)	(0.024)	
ln(1+malaria stability)	-0.254***	-0.167***	-0.013	2775		0.096***	-0.027	2772
	(0.060)	(0.064)	(0.032)			(0.030)	(0.020)	
ln(1+agricultural suitability)	0.064	0.104***	0.065***	2745		-0.050*	-0.020	2742
	(0.046)	(0.040)	(0.022)			(0.030)	(0.026)	
ln(terrain ruggedness)	0.025	0.031	0.016	2776		-0.044*	-0.032	2773
	(0.025)	(0.026)	(0.021)			(0.023)	(0.022)	
oil field dummy	-0.014	-0.040***	-0.032**	2761		0.016	0.009	2758
	(0.018)	(0.012)	(0.014)			(0.012)	(0.014)	
diamond mine dummy	-0.014	-0.012*	-0.003	2761		0.011	0.004	2758
	(0.011)	(0.006)	(0.009)			(0.008)	(0.008)	
Panel B: history								
ln(distance to railroad)	-0.251***	-0.247***	-0.076***	2276		0.215***	0.070***	2273
	(0.028)	(0.026)	(0.015)			(0.023)	(0.023)	
ln(distance to road)	-0.218***	-0.175***	-0.048***	2492		0.169***	0.059***	2489
	(0.022)	(0.019)	(0.009)			(0.017)	(0.016)	
ln(distance to cath. mission)	-0.344***	-0.243***	-0.035	2786		0.230***	0.071**	2783
	(0.039)	(0.031)	(0.024)			(0.032)	(0.031)	
ln(distance to prot. mission)	-0.295***	-0.207***	-0.031*	2786		0.167***	0.030	2783
	(0.037)	(0.030)	(0.016)			(0.022)	(0.020)	
ln(distance to precolon. empire)	0.056	-0.018	-0.052	2786		0.003	0.030	2783
	(0.054)	(0.037)	(0.056)			(0.038)	(0.049)	
ln(distance to precolon. state)	-0.036	-0.042*	-0.020	2786		0.024	0.007	2783
	(0.030)	(0.024)	(0.014)			(0.019)	(0.016)	
Panel C: contemporary								
ln(population density 1950)	0.211***	0.212***	0.089***	2774		-0.148***	-0.051**	2771
	(0.052)	(0.041)	(0.026)			(0.032)	(0.024)	
urban share (born < 1960)	0.338***	0.175***	-0.025	2583		-0.203***	-0.080***	2582
	(0.023)	(0.026)	(0.018)			(0.023)	(0.016)	
agri. empl. share (born < 1960)	-0.586***	-0.371***	-0.116***	2461		0.307***	0.105***	2460
	(0.036)	(0.025)	(0.033)			(0.024)	(0.024)	
manuf. empl. share (born < 1960)	0.195***	0.113***	0.011	2461		-0.098***	-0.016	2460
	(0.052)	(0.027)	(0.017)			(0.023)	(0.012)	
serv. empl. share (born < 1960)	0.556***	0.360***	0.120***	2461		-0.297***	-0.107***	2460
	(0.038)	(0.026)	(0.032)			(0.025)	(0.022)	

This is not a normal regression table. In the column entitled “share literate old” the dependent variable is the district share of parents with at least primary schooling (estimated net of country-year and country-birth-decade fixed effects for young and old). In the columns entitled “IM” it is the district-level share of children of parents with less than primary who complete at least primary (for upward IM, columns (2)-(4)) or the share of children of parents with at least primary who complete less than primary (for downward IM, columns (5)-(7)) (estimated net of country-year and country-birth-decade fixed effects for young and old), which is also the LHS in the columns entitled “IM conditional on 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 conditional on 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 F.2: District-level correlates of the share of literate old and IM, country fixed effects, all birth cohorts, ages 14-18, regions with 80+ percent cohabitation rates

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM conditional on share literate old	N	IM	IM conditional on share literate old	N
Panel A: geography							
ln(distance to capital)	-0.336***	-0.325***	-0.092***	1675	0.199***	0.065***	1673
	(0.038)	(0.042)	(0.026)		(0.034)	(0.025)	
ln(distance to border)	0.023	-0.019	-0.036**	1675	-0.013	-0.002	1673
	(0.046)	(0.037)	(0.015)		(0.029)	(0.020)	
ln(distance to coast)	-0.224***	-0.270***	-0.106***	1675	0.144***	0.047**	1673
	(0.066)	(0.054)	(0.024)		(0.039)	(0.024)	
ln(1+malaria stability)	-0.261***	-0.257***	-0.066**	1667	0.148***	0.039	1665
	(0.072)	(0.068)	(0.028)		(0.051)	(0.038)	
ln(1+agricultural suitability)	-0.007	0.042	0.048*	1647	-0.044	-0.047	1645
	(0.074)	(0.064)	(0.027)		(0.048)	(0.038)	
ln(terrain ruggedness)	0.177***	0.152***	0.020	1668	-0.126***	-0.052*	1666
	(0.065)	(0.048)	(0.022)		(0.034)	(0.030)	
oil field dummy	0.005	-0.006	-0.010	1653	0.020	0.022	1651
	(0.036)	(0.029)	(0.014)		(0.019)	(0.015)	
diamond mine dummy	-0.000	-0.001	-0.001	1653	0.016	0.016	1651
	(0.015)	(0.014)	(0.006)		(0.014)	(0.012)	
Panel B: history							
ln(distance to railroad)	-0.379***	-0.374***	-0.089***	1364	0.236***	0.039	1362
	(0.049)	(0.044)	(0.026)		(0.029)	(0.026)	
ln(distance to road)	-0.269***	-0.266***	-0.053***	1381	0.234***	0.100***	1379
	(0.045)	(0.042)	(0.018)		(0.029)	(0.025)	
ln(distance to cath. mission)	-0.394***	-0.403***	-0.120***	1675	0.254***	0.090***	1673
	(0.075)	(0.061)	(0.030)		(0.045)	(0.033)	
ln(distance to prot. mission)	-0.385***	-0.349***	-0.071***	1675	0.227***	0.069**	1673
	(0.054)	(0.043)	(0.027)		(0.029)	(0.027)	
ln(distance to precolon. empire)	-0.014	-0.008	0.002	1675	0.015	0.009	1673
	(0.045)	(0.040)	(0.020)		(0.047)	(0.043)	
ln(distance to precolon. state)	-0.048	-0.045	-0.008	1675	0.033	0.012	1673
	(0.045)	(0.039)	(0.017)		(0.031)	(0.025)	
Panel C: contemporary							
ln(population density 1950)	0.279***	0.253***	0.050**	1667	-0.156***	-0.041**	1665
	(0.065)	(0.059)	(0.024)		(0.032)	(0.021)	
urban share (born < 1960)	0.360***	0.279***	0.009	1537	-0.223***	-0.083***	1537
	(0.028)	(0.031)	(0.020)		(0.025)	(0.020)	
agri. empl. share (born < 1960)	-0.601***	-0.462***	-0.060*	1464	0.276***	0.036	1464
	(0.035)	(0.029)	(0.033)		(0.029)	(0.028)	
manuf. empl. share (born < 1960)	0.231***	0.140***	-0.028	1464	-0.067***	0.035**	1464
	(0.045)	(0.039)	(0.021)		(0.024)	(0.015)	
serv. empl. share (born < 1960)	0.599***	0.474***	0.084***	1464	-0.291***	-0.064**	1464
	(0.040)	(0.029)	(0.031)		(0.027)	(0.030)	

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 conditional on 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 conditional on 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 F.3: District-level correlates of the share of literate old and IM, country fixed effects, all birth cohorts, ages 14-18, excluding North Africa

		upward IM			downward IM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable	share literate old	IM	IM conditional on share literate old	N	IM	IM conditional on share literate old	N
Panel A: geography							
ln(distance to capital)	-0.301***	-0.305***	-0.092***	2552	0.251***	0.092***	2549
	(0.040)	(0.042)	(0.026)		(0.029)	(0.022)	
ln(distance to border)	0.059	0.037	-0.009	2552	-0.039	-0.004	2549
	(0.039)	(0.036)	(0.012)		(0.030)	(0.017)	
ln(distance to coast)	-0.212***	-0.247***	-0.089***	2552	0.176***	0.054***	2549
	(0.059)	(0.057)	(0.018)		(0.043)	(0.017)	
ln(1+malaria stability)	-0.260***	-0.256***	-0.062**	2542	0.193***	0.046	2539
	(0.049)	(0.053)	(0.025)		(0.045)	(0.032)	
ln(1+agricultural suitability)	-0.027	0.013	0.034**	2518	-0.015	-0.030	2515
	(0.053)	(0.048)	(0.017)		(0.037)	(0.024)	
ln(terrain ruggedness)	0.084*	0.125***	0.061***	2542	-0.104***	-0.056***	2539
	(0.049)	(0.041)	(0.017)		(0.035)	(0.022)	
oil field dummy	0.038	0.035	0.005	2545	-0.001	0.022	2542
	(0.031)	(0.029)	(0.009)		(0.027)	(0.017)	
diamond mine dummy	-0.011	-0.013	-0.004	2545	0.019*	0.012	2542
	(0.012)	(0.009)	(0.007)		(0.011)	(0.010)	
Panel B: history							
ln(distance to railroad)	-0.325***	-0.350***	-0.090***	2336	0.254***	0.064***	2333
	(0.042)	(0.043)	(0.019)		(0.027)	(0.017)	
ln(distance to road)	-0.272***	-0.239***	-0.035**	2552	0.221***	0.073***	2549
	(0.028)	(0.027)	(0.015)		(0.021)	(0.018)	
ln(distance to cath. mission)	-0.316***	-0.301***	-0.066***	2552	0.225***	0.044**	2549
	(0.050)	(0.050)	(0.020)		(0.038)	(0.022)	
ln(distance to prot. mission)	-0.380***	-0.355***	-0.078***	2552	0.285***	0.078***	2549
	(0.047)	(0.043)	(0.019)		(0.033)	(0.022)	
ln(distance to precolon. empire)	-0.001	-0.041	-0.040	2552	0.020	0.020	2549
	(0.042)	(0.030)	(0.027)		(0.037)	(0.035)	
ln(distance to precolon. state)	-0.045	-0.083**	-0.048***	2552	0.058**	0.032*	2549
	(0.040)	(0.035)	(0.016)		(0.028)	(0.019)	
Panel C: contemporary							
ln(population density 1950)	0.235***	0.247***	0.076***	2541	-0.177***	-0.046**	2538
	(0.039)	(0.035)	(0.017)		(0.026)	(0.020)	
urban share (born < 1960)	0.391***	0.255***	-0.049**	2349	-0.255***	-0.063***	2348
	(0.023)	(0.029)	(0.022)		(0.023)	(0.020)	
agri. empl. share (born < 1960)	-0.555***	-0.394***	-0.008	2189	0.348***	0.077***	2188
	(0.030)	(0.028)	(0.038)		(0.026)	(0.027)	
manuf. empl. share (born < 1960)	0.235***	0.144***	-0.023	2189	-0.121***	0.009	2188
	(0.058)	(0.047)	(0.019)		(0.040)	(0.017)	
serv. empl. share (born < 1960)	0.540***	0.392***	0.023	2189	-0.349***	-0.090***	2188
	(0.032)	(0.028)	(0.040)		(0.025)	(0.026)	

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 conditional on 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 conditional on 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. $*p < 0.1, **p < 0.5, ***p < 0.01$. Red/green lines indicate that variables remain significantly correlated with IM when we control for the share of literate parents.

F.3 LASSO

Many of the variables that we consider in Section 5 to characterize the spatial distribution of educational mobility contain measurement error and are correlated with each other. We thus apply LASSO, a machine learning method that via regularization selects the variables (shrinkage) most predictive of upward and downward IM (Tibshirani (1996)).[3] Before presenting the results, we should stress that LASSO does not aim to identify causal effects nor provide unbiased estimates of linear regression coefficients. The objective is providing parsimonious empirical models that predict well regional IM (see Mullainathan and Spiess (2017) and Varian (2014) for overviews and Hastie, Tibshirani, and Friedman (2009) for a textbook treatment). While OLS coefficients solve an unconstrained minimization problem (minimizing the sum of squared residuals), LASSO solves a constrained problem, where there is a trade-off between non-zero coefficients (that reduce the sum of squared residuals) and a penalty incurred having a flexible empirical model with non-zero coefficients. LASSO can be expressed in Lagrangian form:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \underbrace{\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2}_{OLS} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{\ell^1} \right\}$$

The first term is the usual OLS minimization element and the second term (ℓ^1) reflects the regularization (penalty). The larger the (Lagrangian) hyper-parameter λ, the greater the penalty for having non-zero coefficients. When λ is 0, then there is no penalty and LASSO equals OLS. As the parameter increases, variables "drop out", as the empirical model gains parsimony. At the limit, the coefficients of all regressors go to]. The hyper-parameter is typically chosen by cross-validation (re-sampling) techniques that weigh model over-fitting with parsimony.

We compute the LASSO path using Least Angle Regression (LAR, Efron et al. (2004)), allowing λ to range from 0 (OLS) to infinity (all coefficients are set to zero).[4]

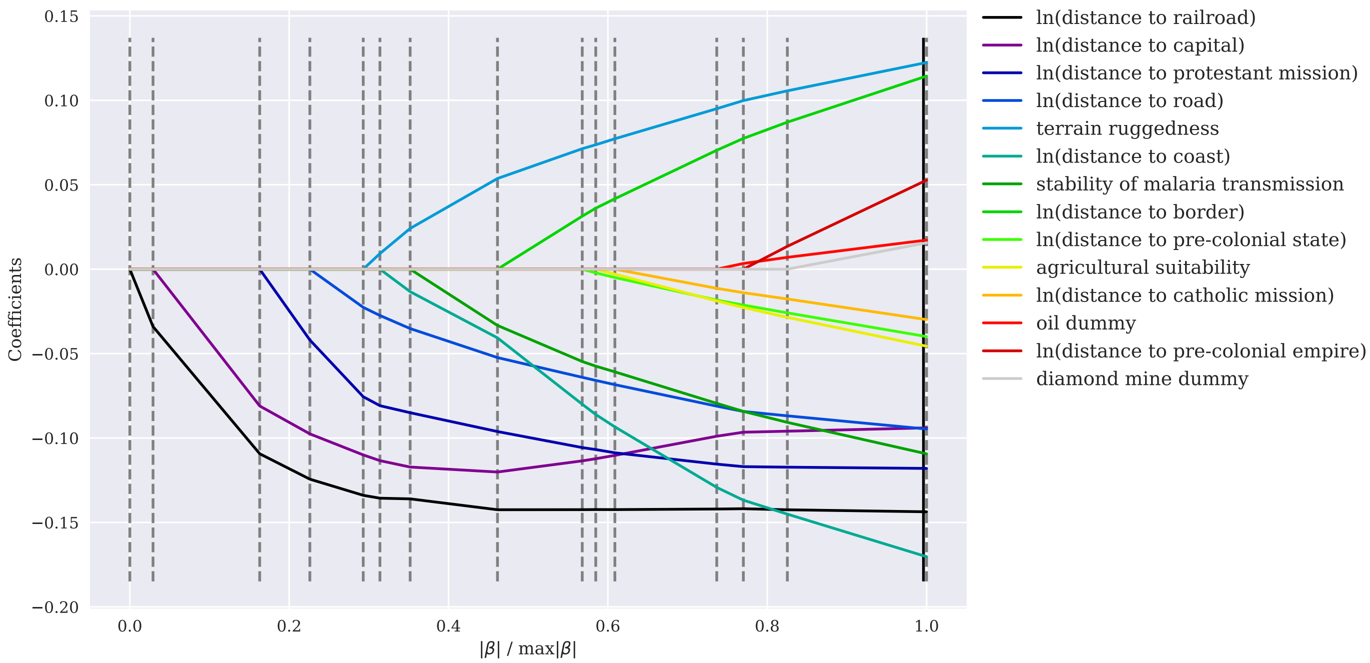
Figures F.1 (a)-(b) plot the entire LASSO path for upward and downward IM, looking at the role of the geographic and historical features. The vertical axis plots the coefficients of the standardized variables. The vertical axis gives the ratio of the sum of the absolute vales of the coefficients for a given λ to the OLS sum (that takes the maximum value). The graphs should be read from right, no regularization (OLS), to left (regularization results in all parameters to be zero). Moving from right to left, more variables drop from the model; the last variables that drop out are the "strongest" predictors of mobility.

[3] LASSO that can be given a Bayesian interpretation when the estimated parameters are independently Laplace distributed. See Xu, Caramanis, and Mannor (2010) for a link between LASSO and robust regression.

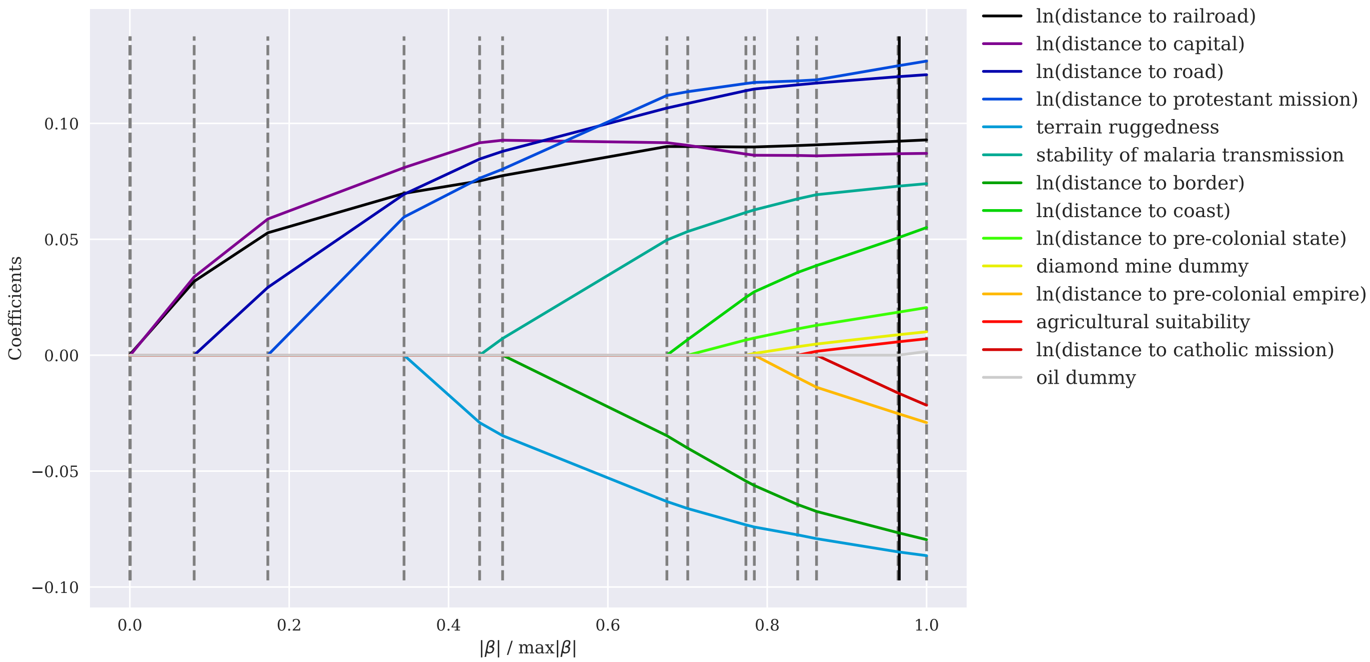
[4] We also computed the "optimal" degree of regularization through K-fold cross-validation. We split the data into 10 (5) equal size bins of districts; for a given value of λ we fit (train) the model in 9 (4) bins, then take out-of-sample fits in the remaining bin, and compute the forecast accuracy with the mean square error criterion. Repeating this process for a grid of 100 values of λ, we take the value that yields the lowest error. The cross-validation routine yields $\beta / \max|\beta|$ around 0.8.

Figure F.1: Lasso estimates

(a) Upward



(b) Downward



While the number of observations is quite small for "machine learning" type algorithms, LASSO is useful pinpointing the relatively more important predictors of mobility. The LASSO analysis with upward IM yields some noteworthy features. First, in line with the univariate correlational analysis, natural resource indicators and proximity to precolonial states are the variables with the least explanatory power. Second, proximity to colonial railroads is the most important predictor of upward mobility, hinting that these investments, though overall small, had sizable lasting consequences. Third, proximity to the capital is also an important predictor of upward IM. Fourth, proximity to Protestant missions is a robust predictor of upward IM, way more than proximity to Catholic missions that drops out of the empirical model once regularization increase. While these results do not have a causal interpretation, they add to the literature assessing the legacy of Protestant and Christian missions. Note that the univariate correlations are quite similar, showing the usefulness of complementing OLS with LASSO. Fifth, ecological features, malaria stability, terrain ruggedness, and distance to coast are in-between suggesting some non-negligible explanatory power.

The patterns when we run LASSO on downward IM are similar. Distance to colonial railroads and the capital carry the highest prediction power, while natural resources, soil suitability for agriculture, and distance to precolonial states are the least robust predictors of downward IM. The graph also illustrates the difference Protestant and Catholic missions; the former being among the most important predictors of downward IM, the latter the (second) least important.

G Exposure Effects

This Section's results complement the analysis in Section 5 of the paper. The sub-sections of this appendix follow the sub-sections of the main paper. In particular, section G.1 reports tests of coefficient equality for the age-specific regional effects and some sensitivity checks of the semi-parametric estimates in section 5.1. Section G.2 complements the parametric estimates of section 5.2, presenting the heterogeneity analysis with respect to gender and the direction of the move and further sensitivity checks. Section G.3 reports additional results that exploit displacement shocks and use historical migration to predict moves supporting the evidence in section 5.3.

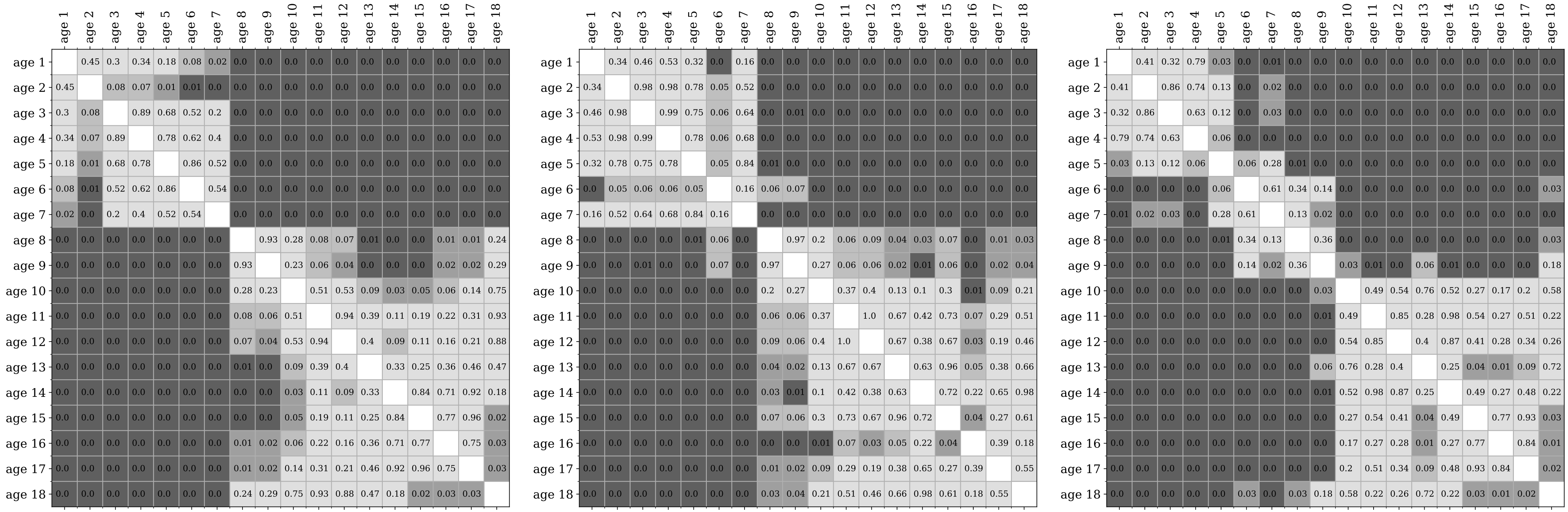
G.1 Baseline Semi-Parametric Estimates

Figure G.1: Visualization of coefficient equality pairwise F-tests

(a) Ages 14-25, all observations, no HH FEs

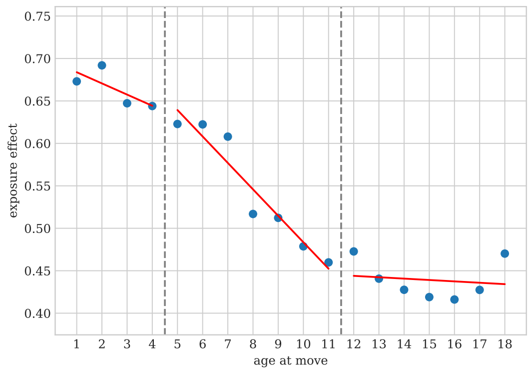
(b) Ages 14-25, HH FE sample, no HH FEs

(c) Ages 14-25, HH FEs



This figure visualizes p-values of pairwise coefficient equality F-tests for all coefficients in Figures 11 and 12. Cells are shaded according to significance level: $p < 0.01$, $0.01 < p < 0.05$, $0.05 < p < 0.1$, $p > 0.1$.

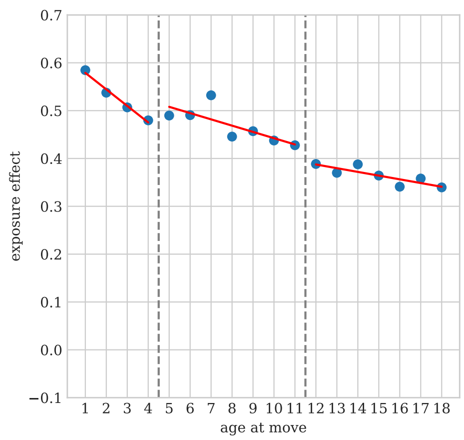
Figure G.2: Instrumenting Δ_{odb} with sample split



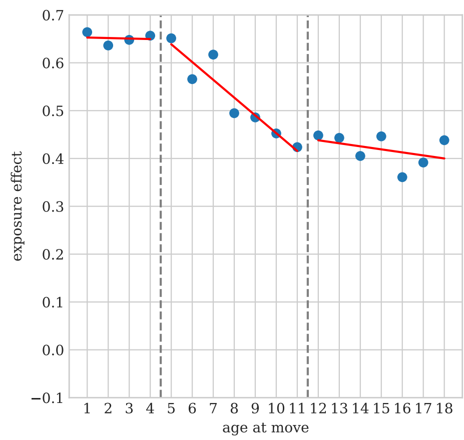
This figure is based on the same specification as Figure 11 except that instead of using Δ_{odb}, we split the sample of non-movers in every cohort-district in two at random and construct two measures of Δ_{odb}, one per sample split. We then re-estimate the specification using all terms involving Δ_{odb} from one of the halves as in instrument for the other half.

Figure G.3: Graphs of semi-parametric exposure effects, rich/poor countries

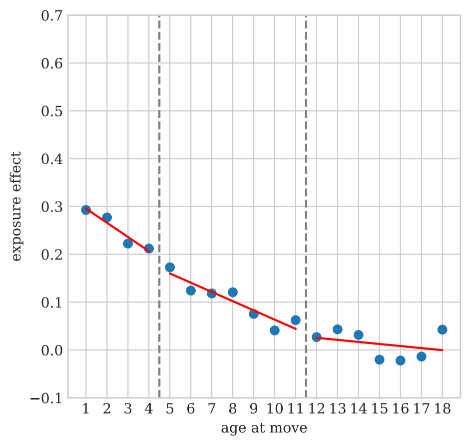
(a) Rich: HH FE sample, no HH FEs



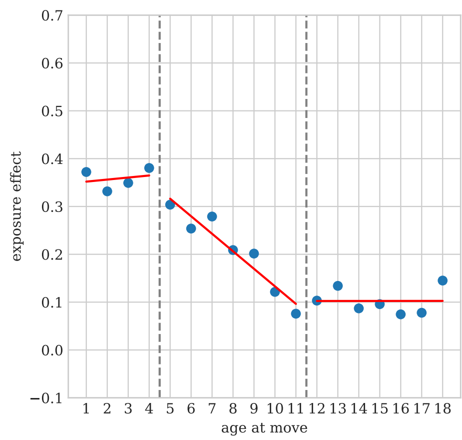
(b) Poor: HH FE sample, no HH FEs



(c) Rich: HH FEs



(d) Poor: HH FEs



This figure plots the estimates from the baseline semi-parametric exposure effects specification estimated, separately for "rich" and "poor" countries. The left column shows the estimates for rich countries, the bottom two those for poor countries. We use World Bank data for real GDP per capita in 1986 (the earliest year for which data for all countries in sample used for the exposure effects analysis are available) and use the median to make the rich/poor split. Accordingly, rich are, from least to most, Sudan, Ghana, Kenya, Zambia, Egypt, Morocco, Cameroon, South Africa. Poor are, from least to most, Benin, Togo, Guinea, Mali, Rwanda, Malawi, Uganda, Ethiopia.

G.2 Parametric Estimates

Table G.1: Parametric exposure effects estimates, heterogeneity

	(1) IM	(2) IM	(3) IM		(4) IM	(5) IM	(6) IM
β: 1-4, Δ^m_{odb}	0.00973	0.00672	-0.00504	β: 1-4, Δ^+_{odb}	0.0141	0.0288	0.0180
	(0.018)	(0.024)	(0.022)		(0.019)	(0.030)	(0.027)
γ: 5-11, Δ^m_{odb}	0.0233***	0.0255***	0.0257***	γ: 5-11, Δ^+_{odb}	0.0339***	0.0311***	0.0271***
	(0.006)	(0.006)	(0.006)		(0.008)	(0.010)	(0.009)
δ: 12-18, Δ^m_{odb}	-0.00966	-0.00324	-0.00272	δ: 12-18, Δ^+_{odb}	0.00788	0.00365	-0.0000797
	(0.007)	(0.008)	(0.006)		(0.009)	(0.011)	(0.006)
β: 1-4, $\Delta\Delta^f_{odb}$	0.0175	0.0126	0.0262	β: 1-4, $\Delta\Delta^-_{odb}$	0.0116	-0.0429	-0.0309
	(0.020)	(0.026)	(0.025)		(0.032)	(0.047)	(0.042)
γ: 5-11, $\Delta\Delta^f_{odb}$	0.0168***	0.00999	0.0133	γ: 5-11, $\Delta\Delta^-_{odb}$	-0.00758	-0.00523	0.00820
	(0.006)	(0.008)	(0.008)		(0.013)	(0.017)	(0.015)
δ: 12-18, $\Delta\Delta^f_{odb}$	0.0180**	0.0125*	0.0115	δ: 12-18, $\Delta\Delta^-_{odb}$	-0.0220	-0.00237	0.00513
	(0.009)	(0.007)	(0.007)		(0.014)	(0.017)	(0.012)
R-squared	0.150	0.123	0.683	R-squared	0.143	0.120	0.679
N	406175	226739	226739	N	406175	226739	226739
age at mig FE	yes	yes	yes	age at mig FE	yes	yes	yes
birth decade FE	yes	yes	yes	birth decade FE	yes	yes	yes
hh FE	no	no, hhfe sample	yes	hh FE	no	no, hhfe sample	yes
age range	14-25	14-25	14-25	age range	14-25	14-25	14-25

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

Table G.2: Parametric exposure effects estimates, robustness

	(1) IM	(2) IM	(3) IM	(4) IM	(5) IM	(6) IM	(7) IM	(8) IM	(9) IM
beta: 1-4	0.0174	0.0169	-0.000658	0.0186*	0.0109	0.00980	0.0144	0.000246	-0.000107
	(0.013)	(0.022)	(0.019)	(0.011)	(0.017)	(0.015)	(0.012)	(0.014)	(0.018)
gamma: 5-11	0.0236***	0.0225***	0.0218***	0.0302***	0.0274***	0.0307***	0.0313***	0.0323***	0.0331***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)
delta: 12-18	0.00147	0.00493	0.00473	-0.000455	0.00139	0.00196	-0.000639	0.00228	-0.00319
	(0.006)	(0.007)	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.007)
R-squared	0.167	0.148	0.675	0.142	0.115	0.683	0.099	0.084	0.692
N	297189	158438	158438	392079	213863	213863	276864	164071	164071
age at mig FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
birth decade FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
hh FE	no	no, hhfe sample	yes	no	no, hhfe sample	yes	no	no, hhfe sample	yes
age range	14-25	14-25	14-25	14-25	14-25	14-25	14-25	14-25	14-25
inc. Egypt + Morocco	no	no	no	yes	yes	yes	yes	yes	yes
incl mg hhs	yes	yes	yes	no	no	no	no	no	no
old for IM	all	all	all	all	all	all	biolog. only	biolog. only	biolog. only

The dependent variable in all regression is a dummy = 1 if the child has completed at least primary, and zero otherwise (i.e. a dummy for IM). The independent variables comprise a linear origin-average-IM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move dummies, birth-decade*imes*destination dummies interacted with destination-minus-origin opportunity differences (to capture differences in measurement error across locations and cohorts), all of which not reported, as well as three linear terms for destination-minus-origin differences in relevant-birth-cohort-non-mover average IM for move-ages 1-4, 5-11, and 12-18. Columns (1)-(3) exclude Egypt and Morocco. Columns (4)-(6) exclude individuals in multigenerational households. Columns (7)-(9) exclude individuals in multigenerational households and use only the education of biological parents to assess individual IM status. Standard errors clustered at origin- and destination-levels in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

Table G.3: Parametric exposure effects estimates for average IM: parents working in agriculture and migrants in rural destination

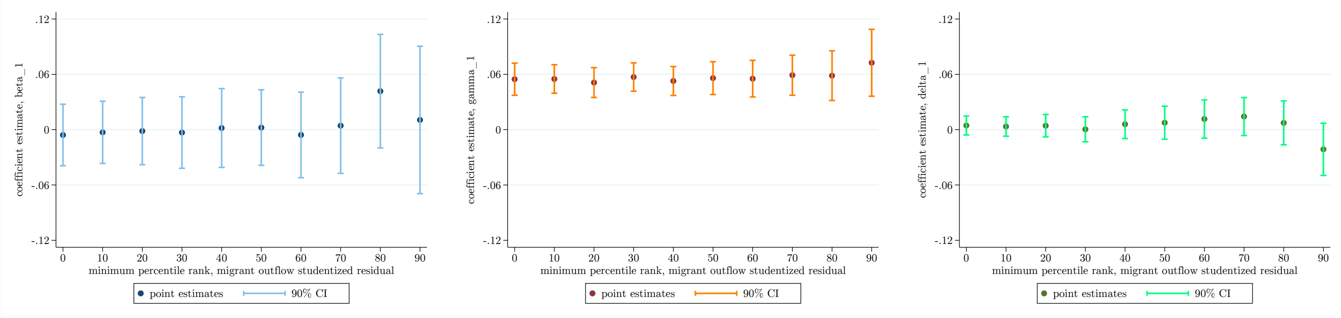
	(1) IM	(2) IM	(3) IM	(4) IM	(5) IM	(6) IM	(7) IM	(8) IM
beta: 1-4	-0.00574	-0.0260	-0.00296	-0.0101	-0.00364	-0.0344	-0.0159	-0.0463
	(0.030)	(0.028)	(0.022)	(0.021)	(0.039)	(0.040)	(0.038)	(0.039)
gamma: 5-11	0.0437***	0.0345***	0.0432***	0.0373***	0.0431***	0.0407***	0.0433***	0.0424***
	(0.011)	(0.010)	(0.008)	(0.008)	(0.014)	(0.015)	(0.013)	(0.014)
delta: 12-18	0.0112	0.0159*	-0.000312	0.00330	0.00721	0.00372	0.00731	0.0103
	(0.009)	(0.008)	(0.008)	(0.007)	(0.010)	(0.011)	(0.010)	(0.012)
R-squared	0.677	0.675	0.692	0.685	0.671	0.668	0.664	0.661
N	48871	57952	105224	97388	36234	31899	43234	38917
age at mig FE	yes	yes	yes	yes	yes	yes	yes	yes
birth decade FE	yes	yes	yes	yes	yes	yes	yes	yes
hh FE	yes	yes	yes	yes	yes	yes	yes	yes
age range	14-25	14-25	14-25	14-25	14-25	14-25	14-25	14-25
ind/occ	industry	occupation	all obs	all obs	industry	industry	occupation	occupation
rural	all obs	all obs	household	region	household	region	household	region

The dependent variable in all regression is a dummy = 1 if the child has completed at least primary, and zero otherwise (i.e. a dummy for IM). The independent variables comprise a linear origin-average-IM (calculated for the birth-cohort relevant to the individual among non-movers) term, age-at-move dummies, birth-decade*imes*destination dummies interacted with destination-minus-origin opportunity differences (to capture differences in measurement error across locations and cohorts), all of which not reported, as well as three linear terms for destination-minus-origin differences in relevant-birth-cohort-non-mover average IM for move-ages 1-4, 5-11, and 12-18. The first two columns focus on children of agricultural workers defined by industry (column (1)) or occupation (column (2)). The next two columns focus instead on children of of parents who migrated to rural areas defined by rural residence of the household (column (3)) or by a greater than sample median rural share (column (4)). Columns (5)-(8) focus more narrowly on the intersections of these two types of sample restrictions: Children of parents working in agriculture (as defined by industry) and having migrated to rural areas as defined by household (column (5)) or greater than median rural resident share (column (6)); and children of parents working in agriculture (as defined by occupation) and having migrated to rural areas as defined by household (column (7)) or greater than median rural resident share (column (8)). Standard errors clustered at origin- and destination-levels in parentheses. $*p < 0.1, **p < 0.5, ***p < 0.01$.

G.3 Endogeneity

Figure G.4: Combined strategies, region-level outflow anomalies, household fixed effects, 2SLS

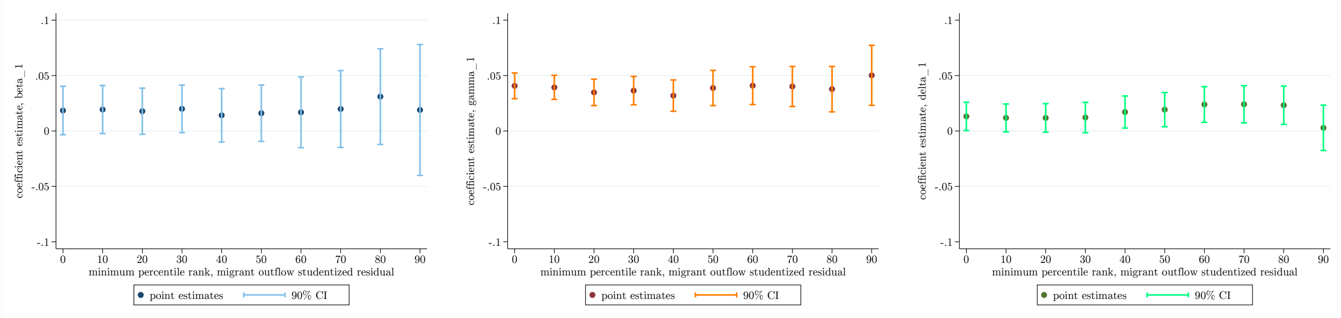
(a) β_1 **(b)** γ_1 **(c)** δ_1



This figure shows point estimates from parametric regression estimates of exposure effects concentrating, successively, on district-years that experienced anomalously larger and large migration outflows and instrumenting the treatment Δ_{odb} with the predicted treatment $\hat{\Delta}_{ob}$ from the shift-share instrument (reduced form regressions). Each point represents the marginal effect of one additional year of treatment (in the relevant age-at-move range) with $\hat{\Delta}_{ob}$=1. Panel (a) shows the coefficient estimates for ages-at-move 1-4, panel (b) those for 5-11, and (c) those for 12-18. To identify outflow-years as anomalous, we contruct a district-year migration outflow panel. For each district, we then regress these migration outflows on a constant and a linear time trend and obtain the studentized residuals from the regression. We then rank these residuals, within a district, from lowest to highest. District-years with high positive residuals experienced migration outflows above trend, and those with negative residuals experienced outflows below trend. We then run the baseline parametric regressions but successively focus on only those district-year observations that have a minimum percentile rank of at least X. The left-most observation for each graph shows the baseline estimates where no observations are dropped. The next observation requires the observations to come from district-years with studentized outflow residuals ranked at least at the 10th percentile within a district etc. All regressions include household fixed effects. 90% confidence bands contstructed from standard errors clustered at origin- and destination-levels are shown.

Figure G.5: Combined strategies, region-level outflow anomalies, observational estimates, 2SLS

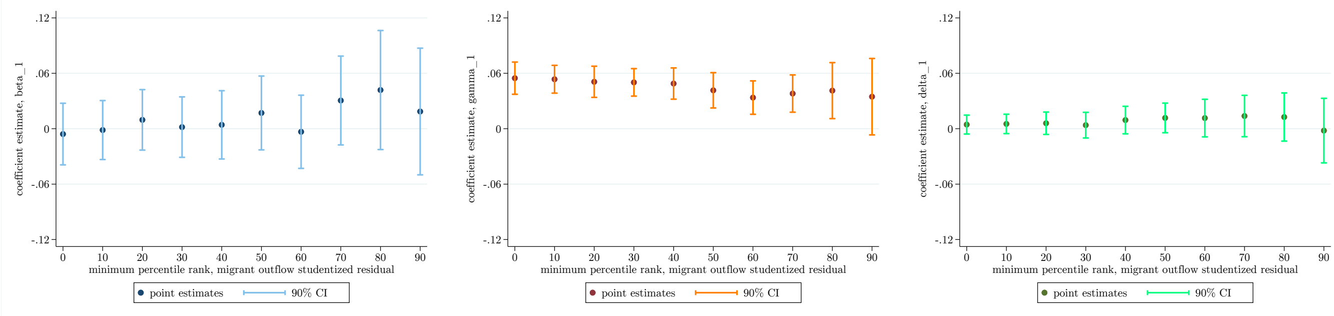
(a) β_1 **(b)** γ_1 **(c)** δ_1



This figure shows point estimates from parametric regression estimates of exposure effects concentrating, successively, on district-years that experienced anomalously larger and large migration outflows and instrumenting the treatment Δ_{odb} with the predicted treatment $\hat{\Delta}_{ob}$ from the shift-share instrument (reduced form regressions). Each point represents the marginal effect of one additional year of treatment (in the relevant age-at-move range) with $\hat{\Delta}_{ob}$=1. Panel (a) shows the coefficient estimates for ages-at-move 1-4, panel (b) those for 5-11, and (c) those for 12-18. To identify outflow-years as anomalous, we contruct a district-year migration outflow panel. For each district, we then regress these migration outflows on a constant and a linear time trend and obtain the studentized residuals from the regression. We then rank these residuals, within a district, from lowest to highest. District-years with high positive residuals experienced migration outflows above trend, and those with negative residuals experienced outflows below trend. We then run the baseline parametric regressions but successively focus on only those district-year observations that have a minimum percentile rank of at least X. The left-most observation for each graph shows the baseline estimates where no observations are dropped. The next observation requires the observations to come from district-years with studentized outflow residuals ranked at least at the 10th percentile within a district etc. 90% confidence bands contstructed from standard errors clustered at origin- and destination-levels are shown.

Figure G.6: Combined strategies, country-level outflow anomalies, household fixed effects, 2SLS

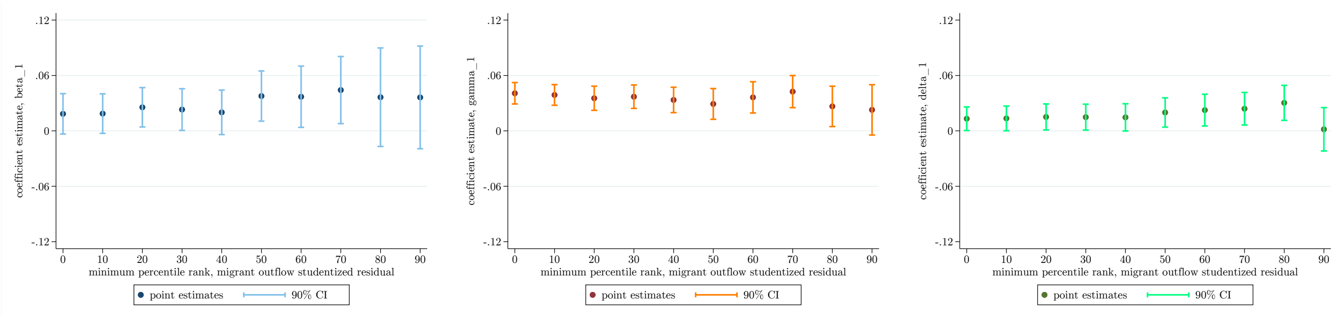
(a) β_1 **(b)** γ_1 **(c)** δ_1



This figure shows point estimates from parametric regression estimates of exposure effects concentrating, successively, on country-years that experienced anomalously larger and large migration outflows and instrumenting the treatment Δ_{odb} with the predicted treatment $\hat{\Delta}_{ob}$ from the shift-share instrument (reduced form regressions). Each point represents the marginal effect of one additional year of treatment (in the relevant age-at-move range) with $\hat{\Delta}_{ob}$=1. Panel (a) shows the coefficient estimates for ages-at-move 1-4, panel (b) those for 5-11, and (c) those for 12-18. To identify outflow-years as anomalous, we contruct a country-year migration outflow panel (we sum, country-by-country, for each year, the migration outflows from all districts). We hen regress these migration outflows on a constant and a linear time trend and obtain the studentized residuals from the regression. We then rank these residuals, within a country, from lowest to highest. Country-years with high positive residuals experienced migrant movements above trend, and those with negative residuals experienced movements below trend. We then run the the baseline parametric regressions but successively focus on only those country-year observations that have a minimum percentile rank of at least X. The left-most observation for each graph shows the baseline estimates where no observations are dropped. The next observation requires the observations to come from country-years with studentized outflow residuals ranked at least at the 10th percentile within a country etc. All regressions include household fixed effects. 90% confidence bands contstructed from standard errors clustered at origin- and destination-levels are shown.

Figure G.7: Combined strategies, country-level outflow anomalies, observational estimates, 2SLS

(a) β_1 **(b)** γ_1 **(c)** δ_1



This figure shows point estimates from parametric regression estimates of exposure effects concentrating, successively, on country-years that experienced anomalously larger and large migration outflows and instrumenting the treatment Δ_{odb} with the predicted treatment $\hat{\Delta}_{ob}$ from the shift-share instrument (reduced form regressions). Each point represents the marginal effect of one additional year of treatment (in the relevant age-at-move range) with $\hat{\Delta}_{ob}$=1. Panel (a) shows the coefficient estimates for ages-at-move 1-4, panel (b) those for 5-11, and (c) those for 12-18. To identify outflow-years as anomalous, we contruct a country-year migration outflow panel (we sum, country-by-country, for each year, the migration outflows from all districts). We hen regress these migration outflows on a constant and a linear time trend and obtain the studentized residuals from the regression. We then rank these residuals, within a country, from lowest to highest. Country-years with high positive residuals experienced migrant movements above trend, and those with negative residuals experienced movements below trend. We then run the the baseline parametric regressions but successively focus on only those country-year observations that have a minimum percentile rank of at least X. The left-most observation for each graph shows the baseline estimates where no observations are dropped. The next observation requires the observations to come from country-years with studentized outflow residuals ranked at least at the 10th percentile within a country etc. 90% confidence bands constructed from standard errors clustered at origin- and destination-levels are shown.