Connectivity and Local Opportunity. Road Building in Benin, Cameroon, and Mali
SUPPLEMENTARY APPENDIX

Luke Heath Milsom*

September 2025

*KU Leuven and the IFS, luke.heathmilsom@kuleuven.be

Contents

1 Additional tables and figures 2

2 Descriptive statistics 3

3 Calculating expected travel times 6

4 Variation in connectivity over time and space 8

5 Calculating locality-year specific incomes 10

6 Digitizing Maps 12

7 Reduced form results robustness and validity checks 13

8 Local clientelism and public good provision 15

9 Top-coding 17

10 Koranic schools and Medersas 18

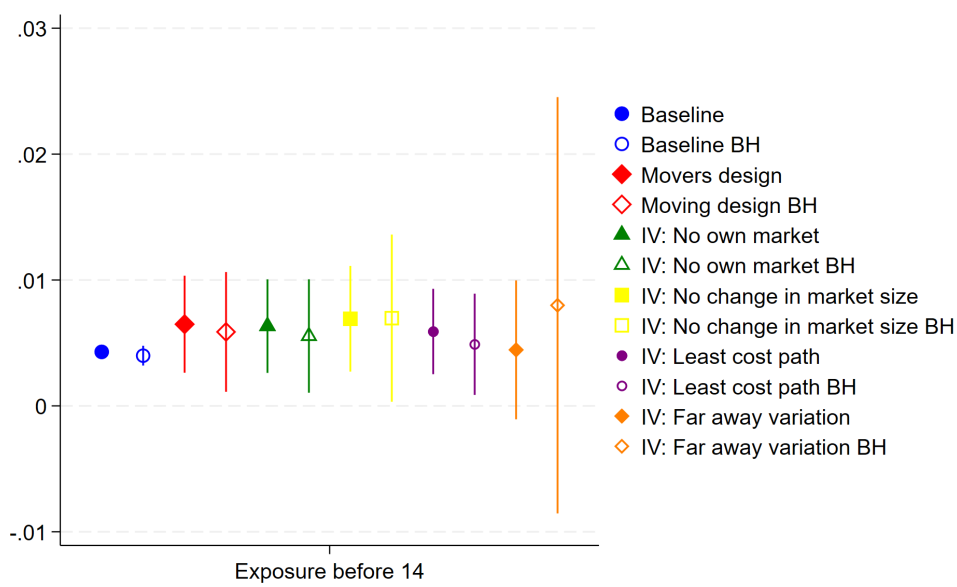
1 Additional tables and figures

Table 1 Estimating $\tilde{\theta}$ for each country in each year

Country	Year	Estimate	Standard Error
Benin	1992	-1.357822	.0552952
Benin	2002	-1.430456	.0672103
Benin	2013	-1.5607	.0950492
Cameroon	1976	-1.069	.0733561
Cameroon	1987	-1.091219	.0716346
Cameroon	2005	-1.165362	.0720341
Mali	1998	-1.282255	.0534792
Mali	2009	-1.316857	.0442214

Notes: This table shows the estimated elasticity of migration with respect to travel time at the country-year level, that is of estimating the migration gravity equation separately for each sample.

Figure 1 Main results controlling for expected market access



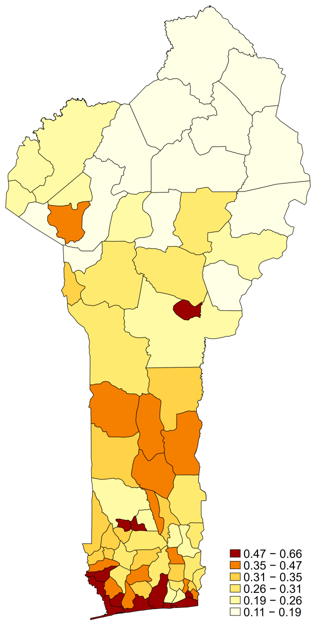
Notes: This figure shows the robustness of the main result to controlling for expected market access. This procedure is described in detail in the main text.

2 Descriptive statistics

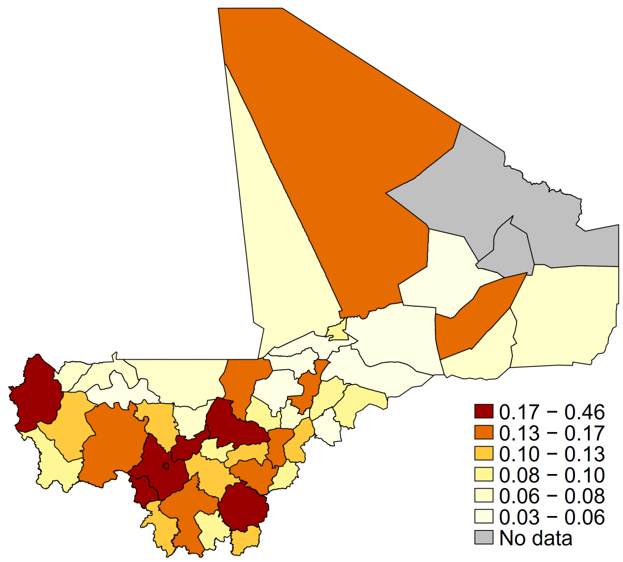
There is substantial variation in primary education completion within country across localities as can be seen by figures 2a, 2b and 2c. In Benin in 2013 the proportion of individuals who had completed primary education in an area varied from 11% in the north to as high as 66% on the coast close to the capital. In addition, Parakou, a large city in the centre of the country had high completion rates. Mali, has consistently lower primary completion rates as compared to Benin or Cameroon. Figure 2b also displays significant cross-locality variation with some areas completion rates as low as 3% and some, especially round the capital, closer to 50%. Cameroon shows a similar pattern, the most educated areas are around the capital, or close to the large coastal city of Douala. Cameroon also has the highest completion rates over all varying from 20% to almost 90%.

Figure 2 Locality level primary completion rates

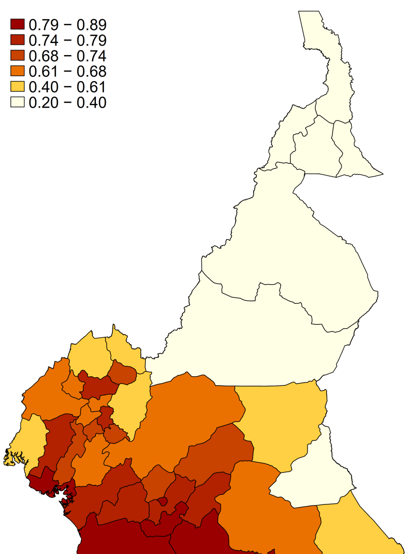
(a) Benin (2013)



(b) Mali (2009)



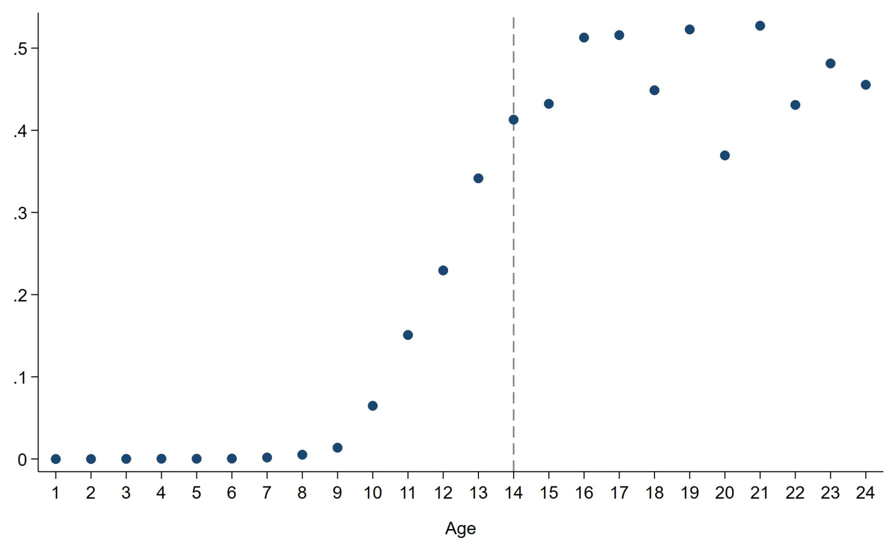
(c) Cameroon (2005)



Notes: this figure shows the spatial distribution of primary education completion rates for all those above the age of 12 in each of Benin, Cameroon and Mali. Each figure has it's own scale and corresponding legend where darker orange/red indicates higher completion rates. The data for Benin comes from the 2013 census, for Mali the 2009 census and for Cameroon the 2005 census.

Although not crucial for my analysis, it's helpful for the interpretation of results later that indeed most individuals who will at some point receive primary schooling do so by aged 14. This is evidence from figure 3, which shows the age fixed effects from a regression of primary completion against age and year fixed effects. This figures clearly shows, as anticipated, that most primary education in indeed completed by 14. However, it's worth noting that many children who haven't completed primary education by 12, which is the year officially primary schooling ends in the countries I study, go on to do so in the next few years.

Figure 3 Proportion of the population who have completed primary school by age

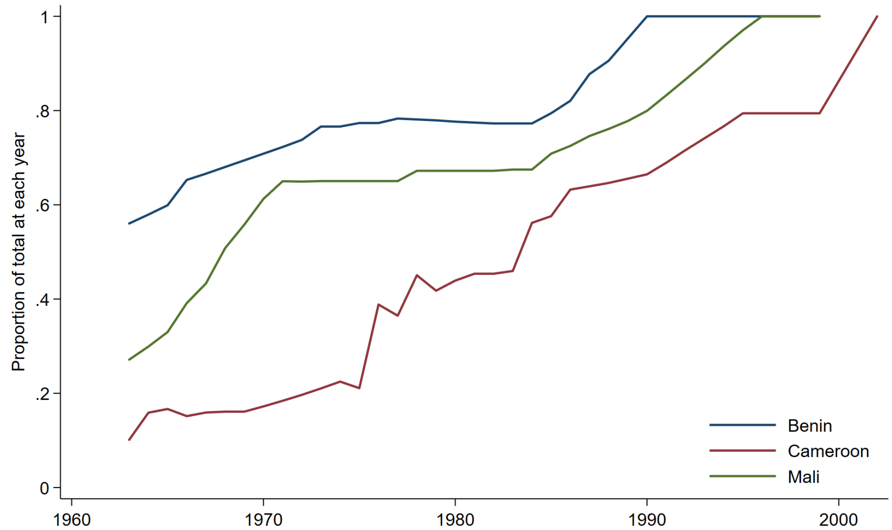


Notes: this figure shows the proportion of the sample who report having completed primary education at the time of each census against their age. This figure uses data from the full sample Benin (1992, 2002, 2013) Cameroon (1976, 1987, 2005) and Mali (1998, 2009).

Figure 4 uses data from Canning and Pedroni [2008] to calculate the proportion of the completed paved road network existing in a given year over my study period. From this figures it's clear that, unlike railways in Benin, Cameroon, and Mali, roads were mainly a post-colonial technology displaying significant variation even in the recent past. In figure 4 it's clear that Cameroon has seen the most intensive increase in road stock since 1960 when it had less than 20% of the length it does today. Mali and Benin, however, are not too far behind with less than 30% and less than 60%

respectively of their modern road stock in place by 1960.

Figure 4 Variation in paved roads



Notes: This figure shows the proportion of the 2000 total paved road stock in place in each given year for Benin, Cameroon, and Mali. It uses data from Canning and Pedroni [2008].

3 Calculating expected travel times

A key object required to calculate market access terms is the iceberg style movement and trade costs. Both of these are based on the fastest path from i to j in period t along the national transport network of the corresponding country, and so I turn first to estimating these travel times, which I denote by t_{ijt}. However, my data is available at the locality level, which means that t_{ijt} is an aggregate measure of travel times across regions. In order to fully utilise the available variation and keep as close to the actual road network as possible, I don't just rely on centroid-to-centroid measures of distance across large localities. Instead, I take the interpretation that transport costs from i to j are measured as the expected cost of a randomly chosen individual in i travelling to a randomly chosen individual in j. That is, consider individuals $p \in i$ and $q \in j$ and denote their travel time as d_{pqt}. Then I estimate, t_{pqt} as the following where $|i|$ and $|j|$ denote the population size of i and j respectively.

$$t_{ijt} = \frac{1}{|i|} \sum_{p \in i} \frac{1}{|j|} \sum_{q \in j} d_{pqt} \tag{1}$$

However, in order to estimate t_{ijt} in this manner, I would need to observe the exact within locality distribution of the population. To focus on variation in road building rather than potentially endogenous changes in the population distribution, I estimate the within locality population distribution in the pre-sample year of 1970 and keep it fixed. To do this, I introduce a new data source, Africapolis, which maps all agglomerations in Sub-Saharan Africa that will achieve a population of at least 10,000 in 2015 and backdates each agglomerations' population to 1970. I take all such available agglomerations, their exact coordinates and 1970 populations. To this, I add the backdated approximate remaining population of each locality using census data and assign this to the locality centroid. This gives the best estimated within-locality and within-country population distribution in 1970 using available data.

Having completed the above steps, I have a set of locations p within each locality i that is $p \in i$. For each location, I associate a 1970 population $P_{p,1970}$ and time of travelling along the observed road network to each other point $q \in j$ for each j location in the same country, d_{pqt}. Then the expected travel time of a randomly chosen household in i travelling to a randomly chosen household in j in year t is given by the following.

$$t_{ijt} = \sum_{p \in i} \frac{P_{p,1970}}{P_{i,1970}} \sum_{q \in j} \frac{P_{q,1970}}{P_{j,1970}} d_{pqt} \tag{2}$$

This can be seen as a coarse discretisation of equation 1, the best that can be done with the data available. Travel times d_{pqt} are calculated along the digitised actual road network, distinguishing between types of road as described in the main text.

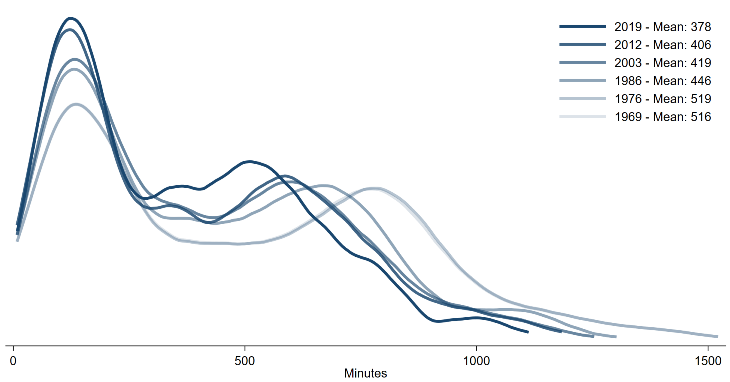
4 Variation in connectivity over time and space

In figure 5, I plot the distribution of expected travel times for each location, in each year maps are available over the digitised road network.[1] The expected travel time for a given location is defined as the length of time an individual should expect to travel for if they were to pick a person at random to travel to from the rest of the country. To calculate this over time, I fix the population distribution to 1970 levels and calculate each locality's expected travel time using the road network in each year. All three figures show considerable leftward shifts in the distribution of travel times over the study period, with mean travel times decreasing by 27%, 41%, and 44% in Benin, Cameroon, and Mali, respectively.

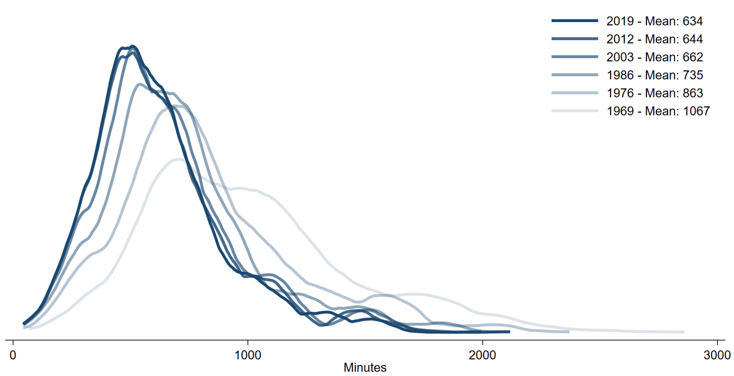
[1] Over my study period, in Benin, Cameroon, and Mali, other forms of transport such as railways or waterways exhibited little variation and are not modelled.

Figure 5 Distribution of expected pairwise travel times between localities

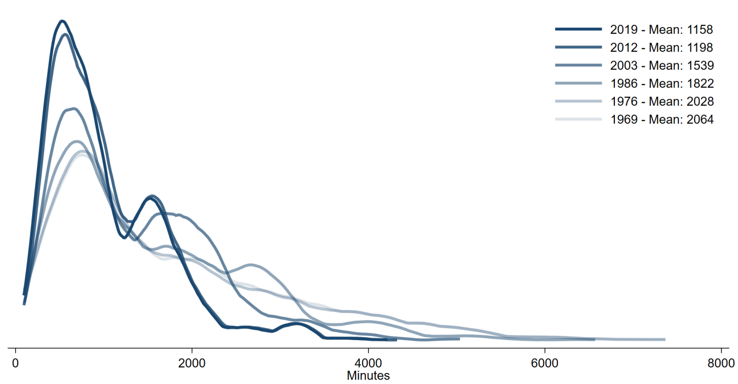
(a) Benin



(b) Cameroon



(c) Mali



Notes: These figures show the density of expected travel times from each locality in a given country-year. Expected travel time in a given location is defined as the time an individual in the location should expected to be traveling if one chooses an individual at random in the same country to travel to. The population distribution is kept fixed at 1970 levels, but the road network is allowed to vary. More recent years are denoted in a darker shade of blue. Population-weighted means across localities for each year are given in the top right.

5 Calculating locality-year specific incomes

Census data does not provide information about wages or the total income/ output of localities. However, it does provide some limited information on the assets households own, such as flooring material, sanitation and electricity. I can use this information, coupled with auxiliary regressions using income data from development health surveys (DHS), to impute approximate income at the locality-year level. Intuitively, this approach is similar to that of Young [2012] in that I use auxiliary Engle curve regressions to uncover parameters which are then used in a second stage with richer data to impute the outcome of interest at a broader and more granular geographic level. This approach requires some assumptions which are difficult to test; however, given the paucity of data available on wages/ incomes at sufficient geographic and temporal disaggregation, I believe that this is a good approximation. Additionally, due to high informality rates, it's unclear whether wages would be the most appropriate measure even if they were available.

Postulate that the (real) demand for an asset a by household h in locality i in year t is given by the following equation.

$$\ln(Q_{ahit}) = \alpha_a + \eta_a \ln(C^N_{hit}) + \xi_a \ln(P_{it}) + \beta X_{hit} + \varepsilon_{ahit} \tag{3}$$

Where α_a are product constants, η_a is the (quasi) income elasticity of demand, C^N_{hit} is nominal household consumption expenditure which is equal to household income in our setting, ξ_a is a vector of own and cross-price (quasi) elasticities of demand, $\ln(P_{it})$ is a vector of regional prices, X_{hit} and β are vectors of household characteristics and their coefficients. Finally ε_{ahit} is a white noise household-product preference shock. Elasticities are referred to as quasi above as for all assets considered, I use an indicator variable rather than a logarithm. To estimate this equation, I use data from the available DHS waves in Benin, Cameroon, and Mali that report income. Sadly, this is only two waves: Benin in 1995 and Mali in 1996. Additionally, these surveys don't include information on prices, and so I estimate equation 3 using product-locality-year fixed effects (although year fixed effects are redundant given that localities are only observed once), which absorbs price variation. Results from running regressions are given in table 2.

Table 2 Asset demand equations using DHS data

	(1)	(2)	(3)	(4)
Conrete floor	0.00550***	0.00557***	0.00479***	0.00496***
	(0.000914)	(0.000923)	(0.000820)	(0.000915)
Electricity	0.00212***	0.00218***	0.00205***	0.00158**
	(0.000653)	(0.000668)	(0.000593)	(0.000665)
Sanitation	0.00312***	0.00319***	0.00313***	0.00258***
	(0.000890)	(0.000898)	(0.000814)	(0.000893)
Asset × Region FE	X	X	X	X
Age polynomial		X	X	X
Asset × Region × Urban FE			X	
HH members control				X
R^2	0.402	0.403	0.498	0.404
N	22586	22586	22586	22586

Notes: This table shows the results from running regressions of the form given in equation 3 using DHS data.

Focusing on column (4), these results suggest that a 1% increase in income is associated with a 0.5pp. increase in the probability of having a concrete floor, a 0.16pp increase in the probability of having access to electricity and a 0.26pp. increase in having accessible sanitation.

In the second step, I use the inverted estimated coefficients from table 2 to approximate income differences by assets households own as indicated in census data. Imputed average income in a locality-year cell is then given by $\tilde{Y}_{it} = 1/N_{it} \sum_{h\in\{i,t\}} \sum_a \frac{1}{\hat{\eta}_a} Q_{hait} + base$ where *base* is the average income calculated from DHS data. Intuitively, if we see households in an area with more assets than those in a different area, we infer that those in the first area have more income with which to purchase such assets. The Engle curve auxiliary regressions in the first stage allow me to approximate how much more income owning an asset may signal, and thus translate regional differences into a money-metric form.

6 Digitizing Maps

Data on the changing connectivity of place comes from digitised historical Michelin road maps accessed from the Bodleian Library at the University of Oxford. In this section, I detail the digitising procedure taken. Figure 6 gives examples of the original maps used for Mali. Throughout, I used the geographical mapping software ArcGIS. The procedure taken is detailed in the steps below.

1. Download the Open Street Maps shapefiles for Benin, Cameroon, and Mali, which representing the current road network in each country.

2. Remove minor roads, or other roads not represented on the most recent (2019) Michelin road maps.

3. Categorise all remaining roads as in the most recent Michelin road maps.

4. Add all settlements from Africapolis that includes all agglomerations that have a population of at least 10,000 in 2015.

5. Using settlement-level population estimates from Africapolis and back-dated locality level population estimates from Census data, calculate the remaining locality-level population not covered by the Africapolis settlements. Add this population to a location at the centre of each locality.

6. Make small adjustments to the 2019 road network so that roads hit the centroid of each settlement and form a connected network. To do this I used the topology tool in ArcGIS.

7. Iteratively delete or downgrade roads using maps increasingly in the past. In this way, for each year a map is available, I create the complete road network.

Figure 6 Original maps

(a) Mali 1969

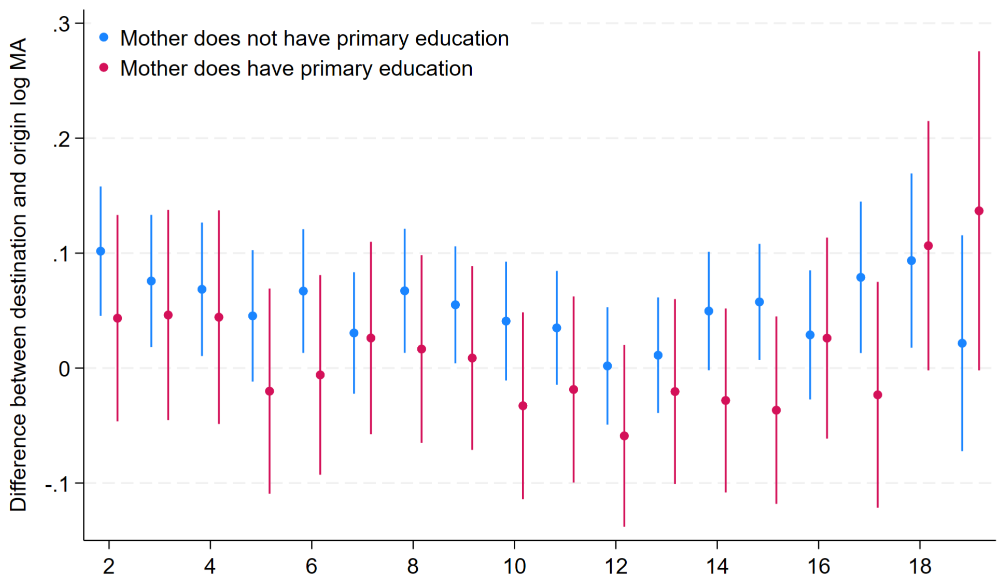
(b) Mali 2019

Notes: This figure shows pictures of the original Michelin road maps for Mali in 1969 on the left hand side and in 2019 on the right hand side.

7 Reduced form results robustness and validity checks

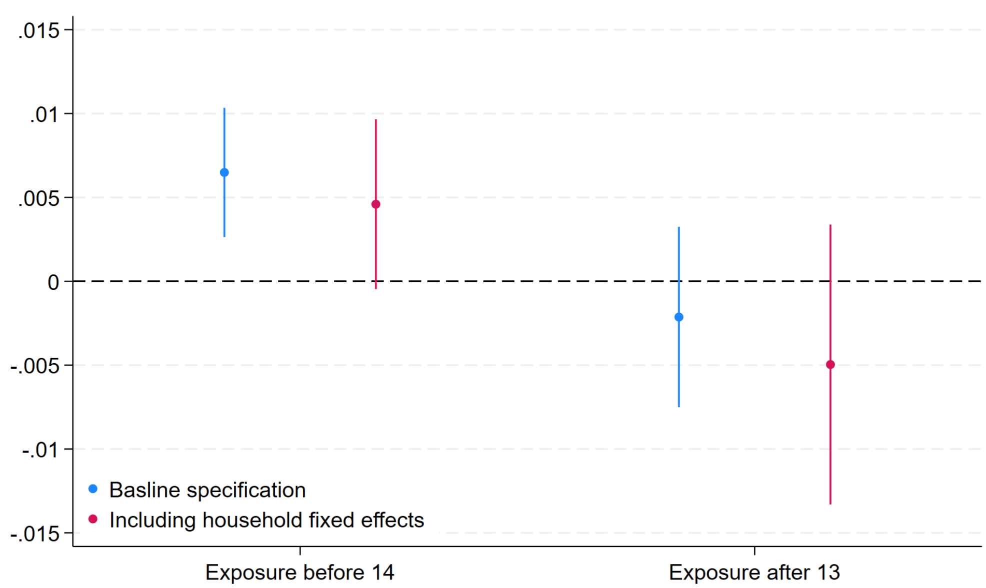
In this section, I report the results from the robustness and validity checks discussed in the main text.

Figure 7 Age-at-move varying selection by mothers education



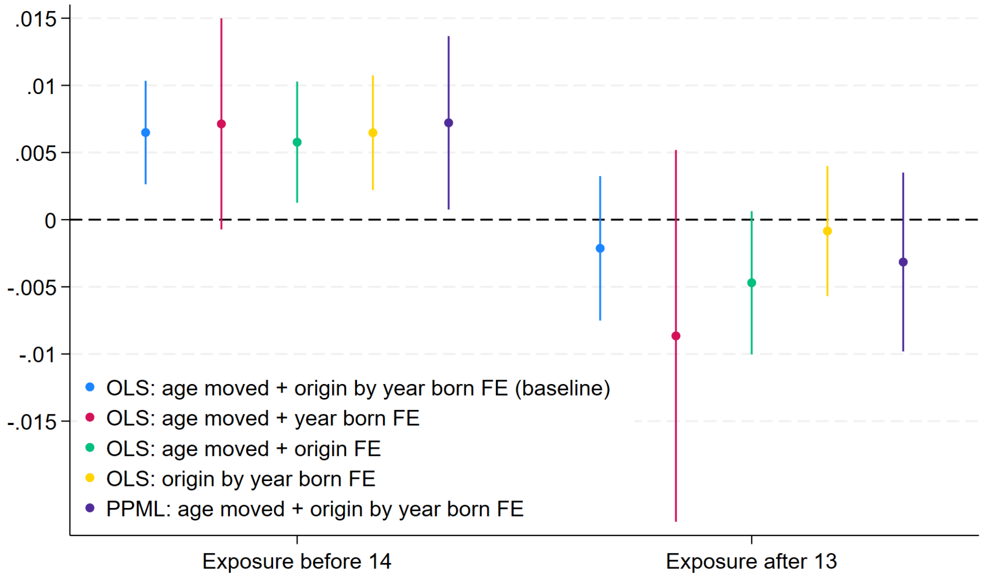
Notes: This figure shows the change in market access between destination and origin location by age at move for those whose mothers have and have not completed primary education.

Figure 8 Including household fixed effects



Notes: This figure compares the estimated β_1 and β_2 from the main specification given in equation **??** in the main text to the same coefficients estimated with the inclusion of household fixed effects.

Figure 9 Specification robustness



Notes: This figure shows the robustness of the main specification given in equation **??** in the main text to specification changes. Each colour refers to a different specification as indicated in the legend. As in the main results figure **??** coefficients β_1 and β_2 are plotted.

8 Local clientelism and public good provision

One potential threat to the identification strategy used in this paper is that the provision of government services and public goods may vary over time and space. That is, if a government comes into power and builds roads and schools so as to benefit a given location in a potentially complex way that is not nullified by the not-on-least-cost-path identification strategy, this could bias the estimated coefficients. In my setting, this concern is most manifest when considering the interaction between local ethnic groups and that of the current political leader, as discussed in the Kenyan context by Burgess et al. [2015]. However, the situation in Benin, Cameroon, and Mali is very different to that in Kenya. In Cameroon, Paul Biya has been in power since 1982, and thus, in Cameroon, there has been no temporal variation over my study period. In Mali, although there has been considerable variation in presidents since the 80s, ethnic favouritism or clientelism has been found to play only a minor, or perhaps even non-existent role [Dunning and Harrison, 2010, Basedau et al., 2011, Basedau and Stroh, 2012, Franck and Rainer, 2012].

In Benin, however, there is some evidence of politics having an ethnic component and clientelism [Battle and Seely, 2010, Fujiwara and Wantchekon, 2013, Wantchekon, 2003] and some correlational evidence that this may lead to less road building in politically marginalised locations Blimpo et al. [2013]. To investigate whether these forces are driving my estimated effects, I construct a dummy variable equal to 1 if the ethnic majority in a location is equal to that of the leader of the time in Benin. Using the Geo-referencing of ethnic groups (GREG Weidmann et al. [2010]) database, I assign each locality in Benin to one of the four major Beninese ethnic groups.[2] Over my sample period, Benin has had three political leaders: in 1992, Nicéphore Soglo (Fon/ Ewe) was in power, in 2002, Mathiew Kérékou (Somba) was in power, and in 2013, Thomas Boni Yayi (Yoruba) was in power. In table 3 I show the results from my baseline movers-design analysis in the first column and, in the second column, replicate this result, additionally controlling for the variable described above. The coefficients on the market access term is stable, and the coefficient on the same ethnicity variable is a precisely estimated 0 effect. I take this as evidence to suggest that threats to identification of this nature are minimal.

[2]Broadly defined as: Ewe, Yoruba, Somba or Barba.

Table 3 Controlling for ethnicity by leader

	(1)	(2)
Exposure to MA before 14	0.00649***	0.00649***
	(0.00195)	(0.00195)
Exposure to MA over 13	-0.00213	-0.00215
	(0.00272)	(0.00273)
Current location has the same predominant ethnicity as the leader		0.0112
		(0.0566)
Observations	109207	109207
R^2	0.353	0.353

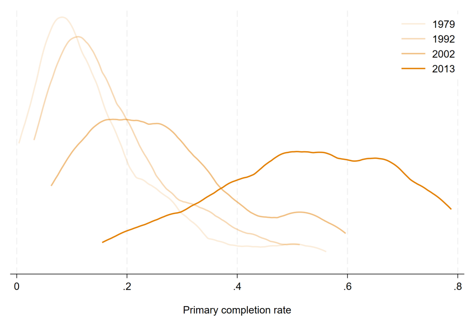
Notes: This table compares the baseline results to those including controls for clientelism. Column one replicates the baseline results using a movers design. Column two includes a dummy variable equal to one if the majority ethnicity of the locality is equal to that of the leader in Benin and zero otherwise.

9 Top-coding

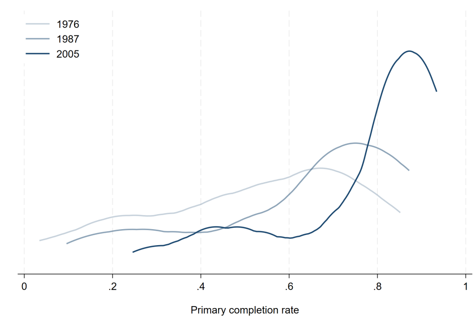
Figure 10 shows how the distribution over localities of primary completion rates for those aged between 15 and 20 has changed in each country over the sample period. These figures show considerable rightward shifts in the distribution, but don't display bunching around 100% — that is they show evidence that top-coding at the upper limit of 100% primary completion is not present.

Figure 10 Changes in the distribution of primary completion rates

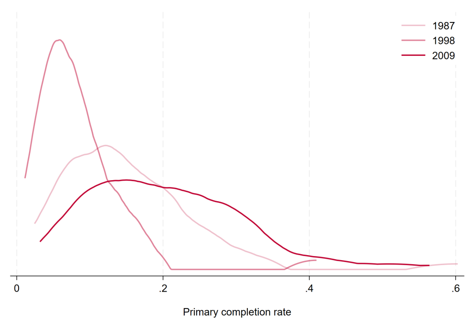
(a) Benin



(b) Cameroon



(c) Mali

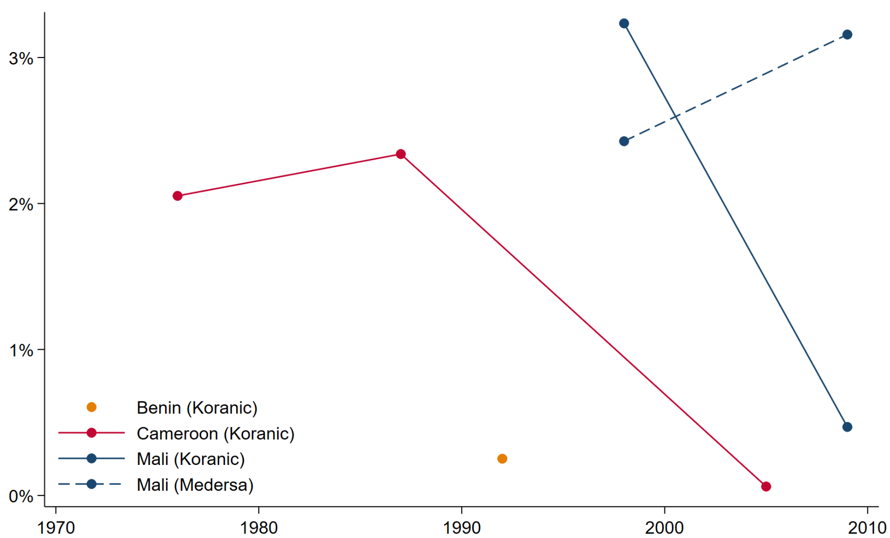


Notes: This figure shows the distribution of primary completion rates in each locality in each country in each census year of those between the ages of 15 and 20. Although primary schooling officially ends at 12 in each country over the time period I study, many children only complete in the years following, and thus I take 15 to be when most who will complete have done so. I cap at 20 in an attempt to capture more recent dynamics, and to remove mechanical correlation across censuses by resampling the same individuals.

10 Koranic schools and Medersas

Koranic schools are a traditional method of education which involves memorising and reciting the Koran. They remain popular in many Muslim countries, and often offer a cheaper or more local method of schooling. In this paper I don't count those who have solely had a Koranic education as having completed primary school, inline with the classification used by IPUMSi. Although these schools primarily concern themselves with memorizing and reciting the Koran, it maybe that they provide some opportunities to those who complete a course at them, and therefore this may be an important dimension I am missing from this analysis — or in the case where students switch from Koranic to state-sponsored schooling, I may be overstating the impact. Fortunately, in some of the censuses used I can distinguish between those at a Koranic school, from those at a secular school. Figure 11 plots the proportion of 6 to 14 year old's in each census where data is available, who are at a Koranic school. It's clear from figure 11 that Koranic education is in the vast minority (maximum 3% of children) and appears to be declining further. It's likely that many more students attend Koranic schools in the evening or on weekends in addition to attending state school — but this dimension is not covered in the data and is less consequential. Figure 11 also shows the proportion of students in Mali at a Medersas [Boyle, 2014] which is a religious school in Mali that follows the national curriculum, toeing the line between Koranic schools and state schools. These schools are on the rise, but still constitute a very small proportion of the overall education.

Figure 11 Proportion of children enrolled in Koranic schools or Medersas



Notes: This figure shows the proportion of primary school aged children (6 to 14) who report attending a Koranic school or a Medersa in the Census.

References

Matthias Basedau and Alexander Stroh. How ethnic are african parties really? evidence from four francophone countries. *International Political Science Review*, 33 (1):5–24, 2012.

Matthias Basedau, Gero Erdmann, Jann Lay, and Alexander Stroh. Ethnicity and party preference in sub-saharan africa. *Democratization*, 18(2):462–489, 2011.

Martin Battle and Jennifer C Seely. It's all relative: Modeling candidate support in benin. *Nationalism and Ethnic Politics*, 16(1):42–66, 2010.

Moussa P Blimpo, Robin Harding, and Leonard Wantchekon. Public investment in rural infrastructure: Some political economy considerations. *Journal of African Economies*, 22(suppl_2):ii57–ii83, 2013.

Helen N Boyle. Between secular public schools and qur'anic private schools: The growing educational presence of malian medersas. *The Review of Faith & International Affairs*, 12(2):16–26, 2014.

Robin Burgess, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. The value of democracy: evidence from road building in kenya. *American Economic Review*, 105(6):1817–51, 2015.

David Canning and Peter Pedroni. Infrastructure, long-run economic growth and causality tests for cointegrated panels. *The Manchester School*, 76(5):504–527, 2008.

Thad Dunning and Lauren Harrison. Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in mali. *American Political Science Review*, 104 (1):21–39, 2010.

Raphael Franck and Ilia Rainer. Does the leader's ethnicity matter? ethnic favoritism, education, and health in sub-saharan africa. *American Political Science Review*, 106(2):294–325, 2012.

Thomas Fujiwara and Leonard Wantchekon. Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics*, 5(4):241–55, 2013.

Leonard Wantchekon. Clientelism and voting behavior: Evidence from a field experiment in benin. *World politics*, 55(3):399–422, 2003.

Nils B Weidmann, Jan Ketil Rød, and Lars-Erik Cederman. Representing ethnic groups in space: A new dataset. *Journal of Peace Research*, 47(4):491–499, 2010.

Alwyn Young. The african growth miracle. *Journal of Political Economy*, 120(4): 696–739, 2012.