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

Luke Heath Milsom*

First version November 2022.
This version September 2025.

Abstract

Where individuals grow up influences their life outcomes — but what characteristics of place matter? In this paper, I consider the role of local connectivity and market integration in determining the causal effect of place on primary school completion in Benin, Cameroon, and Mali. I embed a quantitative spatial economics model within the canonical place-effects framework and use a dual identification strategy combining a movers design with an instrumental variables approach. Growing up in a one standard deviation higher market access location increases your probability of completing primary school by 7 percentage points (12%). I then leverage the full structure of the model to find that the aggregate impact of road building since 1970 in spatial equilibrium is to have increased the causal effect of place by, on average, 33%.

JEL: R12, R42, O18, H54
Keywords: Inequality, Opportunity, Roads, Spatial, Market Access

*KU Leuven and the IFS, luke.heathmilsom@kuleuven.be. I thank Johannes Abeler, Abigail Adams, Sam Altmann, Sarah Clifford, Ian Crawford, Binta Zahra Diop, Dave Donaldson, Christian Dustmann, Thibault Fally, Cecile Gaubert, Douglas Gollin, Marco Gonzalez-Navarro, Nina Guyon, Vernon Henderson, Shihang Hou, Allan Hsiao, Vatsal Khandelwal, Hamish Low, Niclas Moneke, Sanghamitra Mukherjee, Jessica Pan, Barbara Petrongolo, Simon Quinn, Itzhak Rasooly, Ferdinand Rauch, Isabelle Roland, Elizabeth Sadoulet, Giulio Schinaia, Evan Soltas, Adam Storeygard, Gabriel Ulyssea, Tony Venables, Verena Wiedemann, and Hannah Zillessen for useful comments and conversations. I would also like to thank seminar participants at LEAP Bocconi, NEUDC Yale, UEA Conference Washington DC, OxDev Oxford, CEP Junior Trade Workshop LSE, the CSAE conference, applied micro and development economics workshops at Oxford, development workshop at UC Berkeley ARE, European Urban Economics Association conference LSE, Warwick PhD conference, African Meeting of the Econometric Society, Royal Economics Society Young Economist Symposium, Newcastle PhD conference, Infra4Dev World bank, UEA Urban Economics Workshop, National University of Singapore, and Tufts University.

Inequality of opportunity is pervasive. A salient dimension of such inequality is that over space: Where you grow up is an important determinant of later life outcomes. These differences are perhaps most stark in the context of low and middle-income countries. They are also potentially more consequential, stoking regional dissatisfaction, reducing allocative efficiency, and causing misery for those in "left behind" areas. One potential policy solution is to move people to areas of higher opportunity, but this is not scalable. Instead, in this paper, I consider the possibility of moving opportunity to people. To do this, policymakers need to know what characteristics of a location cause it to be a place of high or low opportunity.

I consider how changes in local connectivity through road building shape the geography of opportunity in Benin, Cameroon, and Mali. I define local opportunity as the causal effect of growing up in a given location on your probability to complete primary education, following [Alesina et al., 2021, Chetty and Hendren, 2018b, Milsom, 2023].[1] Individuals growing up in a location that becomes better connected may become exposed to greater opportunities because it may become easier to move from a given origin location to existing areas of higher opportunity [Hsiao, 2024] or, because greater connectivity may cause the spatial distribution of local opportunities to move. Connectivity, however, need not always have a positive impact; it could cause opportunities to leave an area, especially if rival locations become better connected.

Recovering the causal effect of changes in connectivity through road building on local opportunity is challenging for four key reasons. First, measuring changing connectivity through road building is complex because the impact of any given road will depend on the entire pre-existing road network and distribution of economic activity [Donaldson, 2022]. Second, as a result, the impact of any given change in the network

[1] I focus on primary education completion for three main reasons. First, measures of income commonly used in high-income countries are less appropriate in this setting, where the majority of workers are not waged and subsistence agriculture is common. Second, primary completion is the most salient margin of education in my setting, with only 7% of my overall sample going on to complete secondary school. Third, primary completion is widely available in data sources with sufficient geographic granularity and size to be amenable to my analysis.

will spill over in some capacity to all other locations, prohibiting the possibility of any pure control locations. Third, road building is far from random; policymakers may build to service growing locations or galvanise flagging ones, and any derived measure of connectivity will inherit this endogeneity. Finally, changes in local connectivity may change the selection of individuals who move across space, potentially causing primary completion to rise (or fall) in a location simply due to a change in the characteristics of the local population. In this paper, I am interested in the causal effect of place on primary education completion and not in this selection effect, and so require a strategy to separately identify these channels.

To overcome these challenges, I first set up a canonical place-effects model and embed this within a quantitative spatial economics model with education and location choice, expanding upon Hsiao [2024] by endogenising local real wages in spatial equilibrium. This general framework delivers a parsimonious sufficient statistic result: The impact of roads on local opportunity can be summarised by a location's market access. To measure market access, I combine survey data with comprehensive historical road maps that I digitise. To overcome the challenges of endogeneity and selection into location, I combine two identification strategies. First, I use a movers design following Chetty and Hendren [2018a,b] and others to isolate place effects by using variation in exposure to locations due to differences in the age children move across locations. Second, to overcome the endogeneity of road building, I use an instrumental variables strategy leveraging variation in market access due to far-away road building Faber [2014] and develop a novel not-on-least-cost-path identification strategy.

Combining these approaches with the sufficient statistic result, I estimate the reduced form effect of growing up in a higher market access location on the causal effect of place on primary education completion. I find that growing up in a one standard deviation higher market access location increases the probability of completing primary school by 7 percentage points (12%). This result is robust to reasonable variations in specification, only using within-family variation in exposure to locations, and controlling for endogenous exposure to exogenous shocks [Borusyak and Hull, 2023].

To quantify the impact of road building since 1970 on the spatial distribution of opportunity, I then solve the model using exact-hat algebra and identify the remaining parameters from the reduced form results, and using plausibly exogenous variation following Bryan and Morten [2019]. I find that road building since 1970, on average, increases local opportunity by 33% — but that this average hides considerable heterogeneity, with some locations barely changing and others doubling. These relatively large effects suggest that remoteness and location-specific low returns to education are key factors suppressing education completion and causing inequality of opportunity across space in Benin, Cameroon, and Mali.

This paper contributes to the literature in three main ways. First, it contributes to the literature studying spatial inequality and place effects [Chetty and Hendren, 2018a,b, Deutscher, 2020, Laliberté, 2021, van Maarseveen, 2021, Alesina et al., 2021, Rojas Ampuero, 2022], which previously has focused on estimating the causal effect of place, I add to this by going a step further and asking how policy can alter the distribution of causal place effects. A recent strand of this literature has considered place effects within a general equilibrium setting [Chyn and Daruich, 2022, Eckert and Kleineberg, 2024]. I build on this literature by considering changes in connectivity in a low and middle-income country setting, and developing a framework that allows greater location-heterogeneity (relative to Chyn and Daruich [2022]) and takes a market access approach with costly trade as well as migration. Previously, this literature has focused on exploiting the possibilities represented by spatial variation within country borders by moving people to areas of opportunity (see Bryan et al. [2014] and Chetty et al. [2016] for example). In this paper, I instead consider how policymakers could move opportunity to people via road building.

Second, I contribute to the literature on quantitative spatial economic modelling by combining the canonical place-effects framework with a quantitative spatial economics model with education choice [Redding and Rossi-Hansberg, 2017, Allen et al., 2020, Donaldson and Hornbeck, 2016, Donaldson, 2018, Allen and Arkolakis, 2023]. Empirically I consider the effect of roads following (among others) Kebede [2024], Sotelo [2020], Adamopoulos [2025], Castaing Gachassin [2013], and Morten and Oliveira [2024]. A smaller literature considers the interaction between observed educational attainment and trade [Fujimoto et al., 2023, Khanna, 2022, Hsiao, 2024]. Edmonds et al. [2010] studies the impact of the Indian tariff reform of the 1990s and finds that the most impacted areas saw the smallest increases in schooling. Atkin [2016] looks at the impact of growth in export manufacturing in Mexico and similarly finds that more affected areas saw greater declines in schooling. Most related to this work, Adukia et al. [2020] and Asher and Novosad [2020], consider the impact of connecting villages in India to the main road network on educational and economic outcomes. They find evidence of higher attainment in connected villages, with enrollment increasing more in locations where the returns to education are the highest. Relatedly, Hsiao [2024] considers educational investments in spatial equilibrium in the setting of Indonesia. I build on these papers by considering the impact of large-scale inter-city road-building in a different empirical setting, focusing on local opportunity rather than observed primary completion, and allowing for education demand to respond endogenously in spatial equilibrium.

Lastly, I contribute to the literature on identifying the impact of changes in connectivity by developing a novel identification strategy using an iterative not-on-least-cost-path approach. An established literature looks at the causal impacts of colonial railways in Sub-Saharan Africa, such as Jedwab and Moradi [2016] and Jedwab, Kerby, and Moradi [2017]. This has more recently been supplemented by work looking at roads (Moneke [2020], Jedwab and Storeygard [2021], Banerjee, Duflo, and Qian [2020], Faber [2014]), and bridges (Brooks and Donovan [2020], Zant [2022]). I contribute to this literature by expanding on the existing *far-away* variation strategy and developing an alternative to the straight-line instrument, or incidental-middle approach Redding and Turner [2015], Michaels [2008] — one which can be applied in all settings that result in a market access relationship.[2]

The rest of this paper proceeds as follows: section 1 describes the setting and data, section 2 elucidates the model, section 3 explains the identification approach and presents the reduced form regression results, section 4 quantifies the impact of road building since 1970, and finally section 5 concludes.

1 The setting and data: Benin, Cameroon, and Mali since 1970

Benin, Cameroon, and Mali together provide an excellent setting to study how changes in connectivity affect spatial inequality, as well as each country being an important setting in which to study these questions in and of themselves. This is because each country displays considerable past variation in local connectivity due to road building, which I will leverage when estimating the effects of changes in connectivity on spatial inequality. Additionally, due to low preexisting levels of paved road coverage [Gwilliam, 2011, Foster and Briceño-Garmendia, 2009] and high anticipated urbanisation and population growth [UN, 2018], considerable investment in road infrastructure is expected in the near future, making this a particularly direct and policy-relevant setting to study the impacts of road building. In this section, I

[2] I also speak to the related literature that has considered the more reduced-form impacts of transport infrastructure on local outcomes. The conclusion from this literature is somewhat ambiguous. Faber [2014] finds that incidentally connected periphery areas in China have lower GDP relative to similar not-connected areas. On the other hand, Baum-Snow et al. [2017] and Banerjee et al. [2020] find positive impacts of being connected in a similar setting, and Jedwab and Storeygard [2021] similarly in Sub-Saharan Africa. Baum-Snow et al. [2020] attempts to reconcile this, again considering roads in China, by showing that impacts depend on the ex-ante urban hierarchy: Core cities benefit at the expense of periphery ones. This paper can be seen as taking the implicit heterogeneity seriously by allowing the impact of every road on every location to differ, and for this to depend on the entire pre-existing road network and distribution of economic activity. This highlights the importance of not generalising from one road to another.

will outline the two main data sources used in my analysis.

1.1 Census data

I use rich and comprehensive census data to deduce variation in local opportunity. I use data on 8 million observations from 1976 to 2013, across 164 localities from every available census from Benin, Cameroon, and Mali.[3] To estimate causal place effects using a movers design, it's necessary to observe where an individual is when the census was taken, their previous location, their birth location, and how long they have resided in their current location.[4] In addition to migration information, I require data on primary education completion,[5] and age, at a granular and consistent geographic level over multiple cross-sectional waves.

Only three countries in Sub-Saharan Africa — Benin, Cameroon, and Mali fit the stringent data requirements. In Benin, I use censuses from 1992, 2002, and 2013, encompassing over 2.2 million individual-level observations in 77 consistent localities (Communes). In Cameroon, I use censuses from 1976, 1987, and 2005, covering over 3.4 million individual-level observations in 39 consistent localities (Departments). Finally, in Mali, I use censuses from 1998, and 2009, giving over 2.4 million individual-level observations in 48 consistent localities (Circles). In total, this gives me a sample

[3]Data is accessed from IPUMS [2020] and consists of 10% samples, with thanks to the National Institute of Statistics and Economic Analysis in Benin, the Central Bureau of Census and Population Studies in Cameroon and the National Directorate of Statistics and Informatics in Mali, who provided the underlying data.

[4]In the censuses described, I can geo-locate individuals at the second administrative unit level. These localities have a median population of 267,000 across all samples. Benin's Communes have a median population of 103,000 with an inter-quartile range of 71,000 to 173,000. Mali's circles have a median population of 308,000 with an inter-quartile range of 197,000 to 520,000. Cameroon's departments have a median population of 456,000 with an inter-quartile range of 225,000 to 907,000. A broadly comparable geographical unit in the US would be commuter zones. To uncover causal place effects, previous research [Chetty and Hendren, 2018a, Laliberté, 2021, Deutscher, 2020] has mainly relied on administrative data that is not typically publicly available. This data is necessary for these studies to observe full migration histories and to match child and parent outcomes over long time frames. Such rich data is not needed to estimate place effects in the censuses I use; it's possible to discern migration histories from cross-sectional evidence.

[5]In Benin, Cameroon, and Mali, education is compulsory for the first six years of schooling between the ages of 6 and 11/12, which covers primary school.

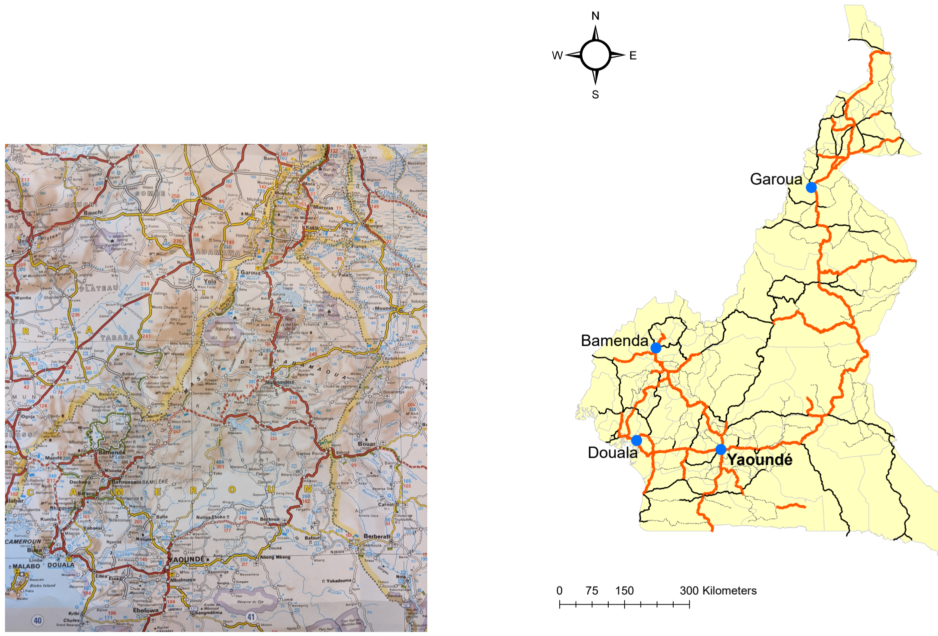
of over 8 million individual-level observations across 164 localities and 444 locality-year cells. A benefit of using this data is that it is freely and publicly available from IPUMS International.

I use the causal effect of growing up in a given location on the probability of completing primary school as my main measure of local opportunity for three main reasons. First, due to the large informal sector, it is unclear whether later life income, as used in other contexts, is appropriate here. Additionally, data on primary completion rates are available at a fine geographic level over time and are less likely to suffer from significant measurement error. Second, primary completion rates are correlated with opportunity more broadly defined in later life. In this setting, individuals who have completed primary school are less likely to work in agriculture, have better housing quality, and greater returns to education [Psacharopoulos and Patrinos, 2018]. Finally, primary schooling is the most salient margin of education. In my sample, about a third of individuals have completed primary school, but only 7% have completed secondary school. Thus, although defensible as a relevant and general measure of local opportunity, primary education completion cannot speak to all potential dimensions, and so opportunity in this paper should be taken to mean local *educational* opportunity.

1.2 The changing geography of connectivity: Digitising historical Michelin road maps

Road data comes from historical Michelin maps, which I have digitised from the following years: 2019, 2012, 2003, 1986, 1976, and 1969. In these maps, it's possible to consistently classify roads into highways, paved roads, improved roads (laterite or gravel), and dirt roads. This classification provides a full description of the (main inter-city) roads over time since 1969 in each country. The ability to distinguish road type is of particular importance, as much of the variation in connectivity, especially in later years, comes from upgrading roads rather than building new ones. Figure 1 gives an example of this process. Panel 1a shows an image of the raw Michelin map of Cameroon in 2019, and panel 1b shows the digitised version.

Figure 1 Digitizing Michelin road maps — Cameroon 2019



(a) Michelin map (b) Digitized map

Notes: This figure shows in panel 1a the original Michelin road map of Cameroon in 2019 and in panel 1b the digitised version. In panel 1b, thick red lines are paved roads, dark black lines are improved roads, and grey lines are dirt tracks. Note that in panel 1a the colour of roads denotes their importance/ frequency of use for long-range trucking and does not necessarily reflect their size or quality, which is denoted instead by the thickness of outlines.

Michelin maps are themselves constructed using four main sources (Jedwab and Storeygard [2021]): the previous Michelin map, government road maps, local information from Michelin tyre stores across Africa, and finally direct correspondence from users. It is generally thought that this process leads to consistent and accurate road mapping over time. Indeed, the success of Michelin maps relied on them being a trustworthy source of information, and so Michelin had a vested interest in providing accurate maps. However, it's still very possible that not every change is noted, and even if a change is noted, it could only be included with some lag. Additionally, although I can use variation in road upgrading, this can only be observed when a road changes categories. That is, road maintenance or changes that would not count as upgrades across categories (such as pothole filling) are not captured. This may be a particular issue in the more recent years, as it is expected that a greater proportion of road spending reflects this unobserved variation.

2 A model of place effects in spatial equilibrium

The first empirical challenge associated with understanding how changes in road building influence local opportunity is that of measurement. How should one measure the effect a given road would have on local connectivity, both in the locations the road connects and in those it does not? The influence of any given road will depend on its position within the entire road network and the pre-existing distribution of economic activity. We can gain intuition on this potential complexity by considering a simplified case. In a three-location economy, the impact on location A of a road connecting locations A and B will depend on the size of the potential market in B. It will also depend on any induced A to B, or B to A, migration. Finally, location C plays a role, perhaps trade is diverted away from C, causing an exodus of workers to A and B. Alternatively, the road could improve outcomes in B, which in turn increases trade from C, causing a relative decline in A.

To overcome this complexity, I embed the simple place effects framework of Chetty and Hendren [2018a] into a quantitative spatial economics model of education completion, building on Hsiao [2024] and Allen et al. [2024]. We take as our point of departure the place-effects literature, where outcomes, Y_k, for individuals k depends on factors specific to k and those specific to where k grew up, denote by $i(k, a)$, the location where k was at age a. In this paper, we are interested in the outcome Y_k, a dummy variable that indicates whether or not an individual has completed primary school. Following Chetty and Hendren [2018a] and others, I assume that place and individual effects are additively separable. Given this, I can decompose outcomes into that due to place effects and that due to individual effects as in equation 1.

$$Y_k = \sum_a \mu_{i(k,a)} + \theta_k \tag{1}$$

Place effects are denoted by μ_i where i is a given location, and θ_k captures place-invariant individual-specific effects. In this section, we provide a structural interpretation for this expression in the case where the outcome of interest is schooling. To do this I develop a quantitative spatial economics model of education and location choice. This model follows Hsiao [2024], extending his framework by endogenising education demand in spatial equilibrium.

Timing and dynamics.

Individuals are born in a given period t in an origin location i. They will stay and be educated in i before choosing a location j to move to in adulthood.[6] Individuals receive education shocks ξ_k and a vector of location-specific skill shocks $\{\varepsilon_{jk}\}$. They choose education e_k after observing their education shock, but before observing their skill shocks, they then observe skill shocks before choosing their migration location. After migrating each migrant has one child who will be educated in their parents migration location.

This framework displays cross-generation dynamics because the pervious periods migration patterns determine the next periods young who become educated and themselves choose where to migrate. However, within an individual, all choices happen within the same period, avoiding the need to consider dynamic choices. I will also make the simplifying assumption that individuals do not consider the utility of their unborn child when considering education choices or migration decisions. The dynamics in this framework are therefore those of spatial equilibrium and will be captured entirely by population movements.

[6] Therefore, in this simple version of the theory, we restrict equation 1 to the single-location setting.

Education.

Individuals choose e_k to maximise expected future utility during work, subject to education costs.

$$e_k^* = \arg\max_e \{\bar{v}_{it}(e) - c_{it}(e, \xi_k)\} \tag{2}$$

Where $\bar{v}_{it}(e)$ is the future expected utility, and $c_{it}(e, \xi_k)$ is the cost of education in location i in period t and is subject to the education shock ξ_k. I parameterise costs as a linear function of education $c_{it}(e, \xi_k) = e \cdot \tau_{it}^{educ} \cdot \xi_k$. Individuals move in adulthood to the location that maximises their utility subject to location-specific shocks ε_{jk}, therefore: $\bar{v}_{it}(e) = \mathbb{E}[\max_j v_{ijt}(e, \varepsilon_{jk})|e]$.

Utility.

Individuals with education e gain utility from working and living in a given location j in period t, which comprises of three components. First, individuals gain utility from local amenities a_{jt}, second from real wage income w_{jt}/P_{jt}, and thirdly are subject to iceberg migration costs $(\tau_{ijt}^m)^{-1}$. Real wage income is given by $w_{jt}/P_{jt} = \frac{r_{jt}}{P_{jt}} h_{jt}$ where r_{jt}/P_{jt} is the real wage per unit of human capital and h_{jt} is the quantity of human capital acquired. This is a non-linear function of education completed $h_{jt} = e^\eta \varepsilon_j$, where ε_j are the location-specific skills shocks that will be resolved before choosing location and are distributed as a type two extreme value distribution with parameter θ. We therefore write utility as follows.

$$v_{ijt} = a_{jt} \cdot \left(\frac{r_{jt}}{P_{jt}} e^\eta \varepsilon_j \right) \cdot \left(\tau_{ijt}^m \right)^{-1} \tag{3}$$

Using this equation and leveraging properties of type two extreme value distributions we can therefore write an expression for expected utility.

$$\bar{v}_{it}(e) = \mathbb{E}[\max_j v_{ijt}(e, \varepsilon_j)|e] = e^\eta \sum_j \pi_{ijt} \frac{a_{jt} r_{jt}}{P_{jt} \tau_{ijt}^m} \mathbb{E}[\varepsilon_j | j] \tag{4}$$

Where π_{ijt} is the probability that an individual moves to j from i in period t. By considering the migration choice problem we can characterise this object.

Migration.

Individuals choose location j to migrate to, taking education e as given. As ε_j has a type two extreme value, the utility representation given in 3 results in the familiar gravity formulation for migration probability π_{ijt}.

$$\pi_{ijt} = \frac{\left(a_{jt}\frac{r_{jt}}{P_{jt}}(\tau^m_{ijt})^{-1}\right)^{\theta}}{\sum_j \left(a_{jt}\frac{r_{jt}}{P_{jt}}(\tau^m_{ijt})^{-1}\right)^{\theta}} \tag{5}$$

Note that using properties of the type two extreme value distribution, we can also write $\mathbb{E}[\varepsilon_j|j] = \gamma\pi_{ijt}^{-1/\theta}$, where $\gamma = \Gamma(1-\frac{1}{\theta})$ and $\Gamma(\cdot)$ is the Gamma function. Putting this together with equation 5 we can simplify equation 4 to the following.

$$\bar{v}_{it}(e) = \gamma e^{\eta}\left(\sum_j \left(a_{jt}\frac{r_{jt}}{P_{jt}}(\tau^m_{ijt})^{-1}\right)^{\theta}\right)^{\frac{1}{\theta}} = \gamma e^{\eta} MA_{it}^{\frac{1}{\theta}} \tag{6}$$

Where we define market access as the local availability of high utility (high amenity and or real wage) locations $\mathrm{MA}_{it} = \sum_j \left(a_{jt}\frac{r_{jt}}{P_{jt}}(\tau^m_{ijt})^{-1}\right)^{\theta}$. The benefits of education are therefore tied to the local availability of high returns to education. Local changes to connectivity will make it easier for individuals to move to high-return locations, and this is well captured within this framework. However, changing connectivity may also alter the distributions of local returns itself. To capture this force, of opportunity moving, we need to endogenise local wages and prices.

Endogenous local real wages.

Firms produce location-specific differentiated goods under perfect competition but are subject to iceberg trade costs denoted by τ^t_{ijt}. Individuals have CES preferences over local varieties with an elasticity of substitution σ. Human capital is the only factor of production. Perfect competition implies that price equals marginal cost, $p_{jt} = r_{jt}/B_{jt}$ where B_{jt} is j,t-specific productivity. CES preferences and iceberg trade costs imply a familiar gravity equation in trade, which in turn gives the usual price aggregator $P_j^{1-\sigma} = \sum_{it}(\tau_{ijt}^{t} r_{it}/B_{it})^{1-\sigma} = TMA_{jt}$. Similarly, assuming symmetric trade costs, given human capital, and assuming goods markets clear, we have that $Y_{jt} = r_{jt}H_{jt} = p_{jt}^{1-\sigma}TMA_{jt}$. Solving the implied system we find $r_{jt} = (H_{jt}^{-1}TMA_{jt})^{1/\sigma}B_{jt}^{(\sigma-1)/\sigma}$. Where H_j is the total local stock of human capital in a given location.

Solving education choice.

Using the above derivations, we can solve an individual, k's, education choice problem to find in logs.

$$\ln(e_k) = \alpha + \frac{1}{\theta(1-\eta)}\ln(\mathrm{MA}_{it}) - \frac{1}{1-\eta}\ln(\tau_{it}^{educ}) - \frac{1}{1-\eta}\ln(\xi_k) \tag{7}$$

Where $\alpha = \frac{1}{1-\eta}\ln(\gamma\eta)$. This equation encapsulates the place effects model. We can rewrite it in the form given by 7 by noting that place-effects are captured by the first two terms, $\mu_{it} = \frac{1}{\theta(1-\eta)}\ln(\mathrm{MA}_{it}) - \frac{1}{1-\eta}\ln(\tau_{it}^{educ})$ and individual-specific effects are given by the final term, $\theta_k = -\frac{1}{1-\eta}\ln(\xi_k)$. Those with higher education skill shocks (lower ξ_k) will complete more education irrespective of local conditions. Those who grow up in locations with higher market access or lower costs of education will complete more education, irrespective of their skill shock.

The sensitivity of educational outcomes to changes in market access depends on two parameters θ and η. The parameter $\theta > 0$ captures the dispersion of location-specific skill shocks. Larger θ implies less dispersed location-specific skill, and so higher migration sensitivity to changes in local (real) wages. Therefore, a higher θ implies a lower sensitivity of education choices to market access as the impact of any location-specific changes is smoothed out to a greater degree through increased migration. The parameter $\eta \in (0,1)$ captures the concavity of the human capital production function. An η closer to 1 implies slower decreasing returns to education and thus greater sensitivity of education completion to incentives to educate, including changes to market access. With only variation in market access, we can not separately identify these two channels, and instead can only identify the bundle: $\beta = (\theta(1-\eta))^{-1}$.

Note that the model-derived market access terms capture how roads affect the ability of individuals to move to areas of higher opportunity, as well as how roads change the underlying spatial distribution of opportunity, and the interaction of these two effects. To see this, we can decompose changes in market access, denoted as $\widehat{\mathrm{MA}}_i$, into these three channels as shown in the equation below.

$$\begin{aligned}
\widehat{\mathrm{MA}}_i - 1 &= \sum_j \pi_{ij} \left(\frac{\hat{r}_j}{\hat{P}_j}\right)^{\theta} \left(\hat{\tau}_{ij}^{m}\right)^{-\theta} \\
&= \sum_j \pi_{ij} \left(\left(\frac{\hat{r}_j}{\hat{P}_j}\right)^{\theta} - 1\right) \left(\left(\hat{\tau}_{ij}^{m}\right)^{-\theta} - 1\right) + \sum_j \pi_{ij} \left(\left(\frac{\hat{r}_j}{\hat{P}_j}\right)^{\theta} - 1\right) + \sum_j \pi_{ij} \left(\left(\hat{\tau}_{ij}^{m}\right)^{-\theta} - 1\right) \\
&= \mathrm{Interaction}_i + \mathrm{MovingOpportunity}_i + \mathrm{Moving2Opportunity}_i
\end{aligned}$$

3 Reduced form: The impact of market access on local opportunity

There are two key challenges associated with taking equation 7 to the data. First, measurement, I need to be able to measure a location's market access and how this changes over time. Second, identification. As τ_{it}^{educ} and ξ_k are not observed, but likely correlated with market access, we need to develop a strategy or pair of strategies to allow identification of β. Indeed, I can write equation 7 in a regression form as:

$$Y_k = \alpha + \beta \cdot \ln\left(\mathrm{MA}_{it}\right) + v_k \tag{8}$$

Where the error term is given by $v_k = \frac{1}{1-\eta}\ln(\tau_{it}^{educ}) + \frac{1}{1-\eta}\ln(\xi_k)$. This formalisation allows a clear elucidation of the two key empirical challenges. First, individuals may select into better locations, causing a correlation between $\ln(\text{MA}_{it})$ and $\ln(\xi_k)$. Second, locations with higher market access may also be conducive to education for endogenous reasons, causing a correlation between $\ln(\text{MA}_{it})$ and $\ln(\tau_{it}^{educ})$, where note τ_{it}^{educ} can be interpreted as all unobserved factors relating to the cost (or benefit) of educating in i in period t.

3.1 Measuring market access

The market access of a given location increases in the distance weighted power-sum of opportunities available in said location $\text{MA}_{it} = \sum_j \left(a_{jt}\frac{r_{jt}}{P_{jt}}(\tau_{ijt}^m)^{-1}\right)^{\theta}$. However none of a_{jt}, r_{jt}, P_{jt} or τ_{ijt} are observed in the data. To overcome this, we can write a locations market access in a more convenient way. Note that $L_{jt} = \sum_i \pi_{ijt}L_{it} = \left(a_{jt}\frac{r_{jt}}{P_{jt}}\right)^{\theta}\sum_i(\tau_{ijt}^m)^{-\theta}L_{it}\text{MA}_{it}^{-1}$. Denote $\widetilde{\text{MA}}_{it} = \sum_i(\tau_{ijt}^m)^{-\theta}L_i\text{MA}_{it}^{-1}$. As shown by Donaldson and Hornbeck [2016] the only solution for this series of equations is that $\text{MA}_{it} = a\widetilde{\text{MA}}_{it}$ for some constant a. Thus we have, up to scale, that $\text{MA}_{it} = \sum_j(\tau_{ijt}^m)^{-\theta}L_{jt}\text{MA}_{jt}^{-1}$. We can take this expression to the data as L_{jt} is observed. To calculate market access we therefore only need to estimate $T_{ijt} = (\tau_{ijt}^m)^{-\theta}$.

3.1.1 Calculating time-varying location-to-location travel times

I parameterise the iceberg travel costs as depending on i to j travel times over the road network prevailing at time t, denoted as t_{ijt}. These times are estimated using the prevailing road network at time t using Dijkstra's algorithm. Data on the road network is generated from historical road maps that I digitise, travel times over the four road categories are taken from Jedwab and Storeygard [2021]. The online appendix section 3 details how t_{ijt} is calculated. I then parameterise these costs in a log-linear fashion such that $\ln(T_{ijt}) = -\theta \cdot \ln(\tau_{ijt}^m) = -\tilde{\theta} \cdot \ln(t_{ijt})$, where $\tilde{\theta} = b \cdot \theta$. Therefore, to calculate market access, it only remains to estimate $\tilde{\theta}$.

3.1.2 Estimating $\tilde{\theta}$

We can estimate $\tilde{\theta}$ using a standard gravity migration equation coupled with data from the censuses on locality-to-locality moves in each period. From the theory the number of individuals moving from i to j in period t can be written as $M_{ijt} = \pi_{ijt} L_{it}$ given the previously derived expression for π_{ijt} and including fixed effects results in the following framework given in equation 9. Where α_{it} and τ_{jt} denote destination-time and origin-time fixed effects respectively.

$$\ln(M_{ijt}) = -\tilde{\theta} \cdot \ln(t_{ijt}) + \alpha_{it} + \tau_{jt} + \varepsilon_{ijt} \tag{9}$$

This theory-consistent gravity equation can be estimated using pseudo Poisson maximum likelihood [Silva and Tenreyro, 2006, Yotov et al., 2016] to find $\hat{\tilde{\theta}} = \hat{\theta} \cdot \hat{b}$. By including this rich set of fixed effects this specification only uses variation in travel times stemming from bilateral travel costs. For example, a common threat to identification would be that as a location becomes more attractive, planners are more likely to build better connections to this location. Oppositely, planners might build connections to a previously flagging location in order to galvanise it. In either case, such behaviour would be captured by the fixed effects in this specification.

Any remaining threats to identification must operate at the bilateral level. One first-order concern of this type is that bilateral cultural ties between two locations may both influence migration and the strength of the road connection between these places. In this case cultural ties, denoted by ct_{ijt}, would be an omitted variable potentially biasing $\hat{\tilde{\theta}}$. Using data from Weidmann et al. [2010], I can control directly for the group similarity between locations.

More generally, one may be concerned that a planner builds better connections between two locations in order to facilitate existing or encourage new migration between these two places. If this were the case changes in travel time due to road building and upgrading maybe for reasons endogenous to migration rates. Moreover, preexisting strong connections maybe due to great social or economic ties, which may also encourage greater bilateral migration. To overcome these concerns I first consider estimating the gravity equation in the cross-section only, and second, consider using straight-line "as the crow flies" bilateral centroid distances as opposed to the estimated travel costs, which use the actual road network.

Table 1 shows the results from estimating equation 9 using the approaches described above. These estimates show remarkable stability across specifications. Due to this consistency the first column is taken as the preferred specification, leading to an estimate of $-\tilde{\theta} = -1.25$. Few estimates of this parameter are available in the literature from a developing country. Perhaps most informative Morten and Oliveira [2024] estimates an elasticity of -1.698 using meso-to-meso migration over a five year period in Brazil. The estimating sample and context are very different between these two settings, but it is nevertheless reassuring that both estimates are similar.

Table 1 Estimating the Migration Elasticity

	(1) PPML	(2) PPML	(3) OLS	(4) OLS	(5) PPML	(6) PPML	(7) PPML
Log(Travel time)	-1.249^{***}	-1.231^{***}	-1.227^{***}	-1.226^{***}	-1.204^{***}	-1.250^{***}	
	(0.0317)	(0.0327)	(0.0174)	(0.0174)	(0.0321)	(0.0374)	
Group Similarity					0.174	0.225	0.0140
					(0.112)	(0.118)	(0.128)
Log(Dist. Crow)							-1.244^{***}
							(0.0411)
Destination-time FE	X	X	X	X	X	X	X
Origin-time FE	X	X	X	X	X	X	X
Sample	Panel	Panel	Panel	Panel	Panel	X-section	X-section
Winsorised		1 & 99		1 & 99			
N	26234	26234	13616	13616	26234	9450	9450

Notes: This table shows the results from estimating equation 9 using various specifications. All specifications use destination-time and origin-time fixed effects with standard errors clustered at the origin-destination pair level. Column one uses PPML, column two also uses PPML winsorising migration flows at the 1st and 99th percentile. Columns three and four instead use a log-linear OLS specification. Column five again uses PPML controlling for group similarity. Columns six and seven move to the cross-section, taking the first available year of data; in both of these columns, group similarity is controlled for. In column seven, instead of using travel times estimated from the actual road network, I use as-the-crow-flies centroid-to-centroid distances.

With estimates of $\tilde{\theta}$, in hand using bilateral travel times I can then calculate each locations market access: $\text{MA}_{it} = \sum_j t_{ijt}^{-1.25} L_{jt} \text{MA}_{it}^{-1}$. This is a series of non-linear equations across locations, which can be efficiently solved using an iterative procedure and has a unique up-to-scale solution [Donaldson and Hornbeck, 2016]. Table 1 in the online appendix shows that estimates of θ are stable across countries and years.

3.2 Identifying the reduced form effect of market access on local opportunity

To identify the causal effect of market access on local opportunity, I combine a movers design with a market access instrumental variables strategy that leverages plausibly exogenous sources of variation stemming from distant road changes. The movers' design uses variation in exposure to different levels of market access by comparing individuals who moved to high market access locations earlier in their childhood to those who moved later and so were less exposed. This allows selection into high-quality (high market access) locations, but requires that selection does not vary systematically with the age at which the child moves. This is the standard assumption in this literature, I perform various robustness and validation exercises probing this assumption in section .

To operationalise the movers' design, I restrict my sample to one time movers between the ages of 14 and 18. In my sample, 81% of individuals in this age range have not moved, 13% have moved once, and 6% have moved more than once. I then make the simplifying assumption of a linear dose effect. Under this assumption what matters for an individual is their exposure to market access over their childhood. I compare children in the same cohort who move from the same origin location to destination locations with different levels of market access at different ages. Leveraging only variation in the age-at-move I can then estimate the causal effect of spending an additional year of childhood in a one percent higher market access location on the probability of completing primary school.

We can write equation 8 in a sample of one time movers between 14 and 18 as $Y_k = \frac{1}{18}\left(m_k * \mu_{o(k),c(k)} + (18 - m_k) * \mu_{d(k),c(k)}\right) + \theta_k$. Where $o(k)$ is individual k's origin location, $d(k)$ is their destination location, m_k is the age at move, and $c(k)$ is k's cohort. From now on, I will suppress implicit dependence on k. I can include origin by cohort fixed effects and write place effects in terms of differences to find $Y_k = \left(1 - \frac{m_k}{18}\right) \cdot \Delta\mu_{odc} + \alpha_{oc} + \theta_k$. Following the theory discussion, I then write the location effect for a given location l as $\mu_{lc} = \beta \cdot \text{MA}_{lc} + v_{lc}$, where v_{lc} captures unobserved aspects of place effects. Substituting this into the above, we find $Y_k = \left(1 - \frac{m_k}{18}\right) \Delta\text{MA}_{odc} + \alpha_{oc} + \tau_m + \varepsilon_k$, where $\varepsilon_k = \left(1 - \frac{m_k}{18}\right) \Delta v_{odc} + \theta_k$ is unobserved. The conditioning on ΔMA_{odc} in levels, including age-at-move fixed effects, and finally allowing effects to differ before vs after moving at age 13 gives me the specification in 10. Following Milsom [2023], and Chetty and Hendren [2018a], I allow differing intercept and slope coefficients for those who move before age 14 and those who move after (inclusive). Those who move after 14 have already completed primary school, and so we do not expect market access to affect their outcomes. The coefficient of interest is β_1, which can be interpreted as the impact of spending an additional year of childhood in a one-unit (one percent) higher market access location on the probability of completing primary school.

$$\begin{aligned} Y_k = \mathbb{1}_{[m_k \leq 13]} \cdot (\alpha_1 + \beta_1 \cdot (18 - m_k)) \cdot \Delta\text{MA}_{odc} + \\ \mathbb{1}_{[m_k > 13]} \cdot (\alpha_2 + \beta_2 \cdot (18 - m_k)) \cdot \Delta\text{MA}_{odc} + \\ \tau_m + \alpha_{oc} + \varepsilon_k \end{aligned} \tag{10}$$

By specifying the error term ($\varepsilon_k = \left(1 - \frac{m_k}{18}\right) \Delta v_{odc} + \theta_k$), we can clearly see the two potential sources of bias. First, correlation between $m_k \Delta\text{MA}_{odc}$ and θ_k will cause bias. This occurs exactly when systematically "better" or "worse" individuals sort into higher or lower market access locations systematically earlier or later. That is, as with normal mover designs, we can allow sorting into better locations, but not that this happens systematically differently at different ages-at-moves. The assumption maintained here is weaker than the usual, as here we only need no such selection into locations on one characteristic — market access. The second potential source of bias is correlation between $\Delta \mathrm{MA}_{odc}$ and Δv_{odc}. This will occur if higher market access locations are better on some other unobserved dimension, i.e. if there is a correlation between MA_{lc} and v_{lc}. This seems possible in the case of road building, as roads are not exogenously placed. The planner could build roads to service booming locations or to galvanise flagging ones, and market access terms will inherit this endogeneity. To overcome this, I require an additional empirical strategy and leverage the construction of market access terms to consider only far-away or not-on-least-cost-path variation.

Previous identification strategies designed to overcome the endogeneity of road placement include using placebo lines from planned but unbuilt routes Donaldson [2018], Okoye, Pongou, and Yokossi [2019], using straight line or least-cost path spanning tree instruments[7] Moneke [2020], Michaels [2008], Ghani, Goswami, and Kerr [2016], Faber [2014], or leveraging *far-away* variation in road changes Donaldson and Hornbeck [2016], Jedwab and Storeygard [2021]. In my setting, it is difficult to see how the first two approaches can be implemented. First, I don't have data on unbuilt but planned placebo lines. Second, there is no clear set of locations that are being connected, and in addition, a significant proportion of the variation in travel times comes from road upgrading rather than the building of entirely new roads, in this setting, it's unclear how localities are "incidentally" connected.

The third strategy, leveraging far-away variation in roads, is appealing but suffers from a number of known drawbacks. First, it's unclear how far "far-away" should be; and although researchers can present many distances, it is ultimately an ad-hoc choice. Second, and more fundamentally, variation due to large projects which may be far away but for endogenous reasons, or relatively far away connections that are built to ease transport to, or encourage trade to a given location, remain threats to

[7]Locations that just happen to lie between two cities that are being connected by a road may be plausibly described as exogenous. This is also known as the *inconsequential units* IV.

identification.

In this paper, I propose a novel identification strategy that builds upon the far-away variation approach by considering not-on-least-cost-path variation. The not-on-least-cost-path variation approach only uses changes in a locality i's market access that stems from indirect changes to all other locations' market access freezing the least cost path from i to all other locations.[8] This approach has two intuitive explanations. First, one can consider it as the same as using far-away variation, but whereas far-away variation defines distance over Euclidean space, not-on-least-cost-path defines distance over network space, where a "distance" of one refers to one-degree removed indirect variation. Under this interpretation, we take the network structure seriously and resolve the ad-hoc nature of what "far-away" may mean by appealing to the theory. Secondly, one can think of not-on-least-cost-path variation as approximating the decision-making process of the policymaker building roads, and using the residual variation. If a central planner builds roads to/ from i in order to directly improve its connectivity, we don't use that variation and instead consider the residual variation in market access.

To formalise the above discussion, consider a generic market access variable $\text{MA}_{it} = \sum_j (\tau_{ijt}^m)^{-\theta} L_{jt} (\text{MA}_{jt})^{-1}$. Following Donaldson [2018], I first remove own-location market access (therefore sum over other locations $j \neq i$) and freeze the market size variable at the initial level (L_{j0}) as these objects are co-determined with the outcome variable. Then I can use the not-on-least-cost-path variation (freeze the least cost path to the initial value τ_{ij0}^m) of degree n by constructing the following instrument.

$$\text{MA}_{it}^{IV} = \sum_{j \neq i} (\tau_{ij0}^m)^{-\theta} L_{j0} \left(\text{MA}_{jt}\right)^{-1}$$

[8] This approach is not to be confused with one variation of the incidental connection approach which uses an estimated least cost route over an estimated cost surface to instrument for actual connections. I don't take this approach in this paper for the same reasons that incidental connection approaches, in general, are unsuitable, and in addition, because a significant proportion of the variation comes from within road quality, and it is unclear how to leverage this intensive margin dimension using this approach.

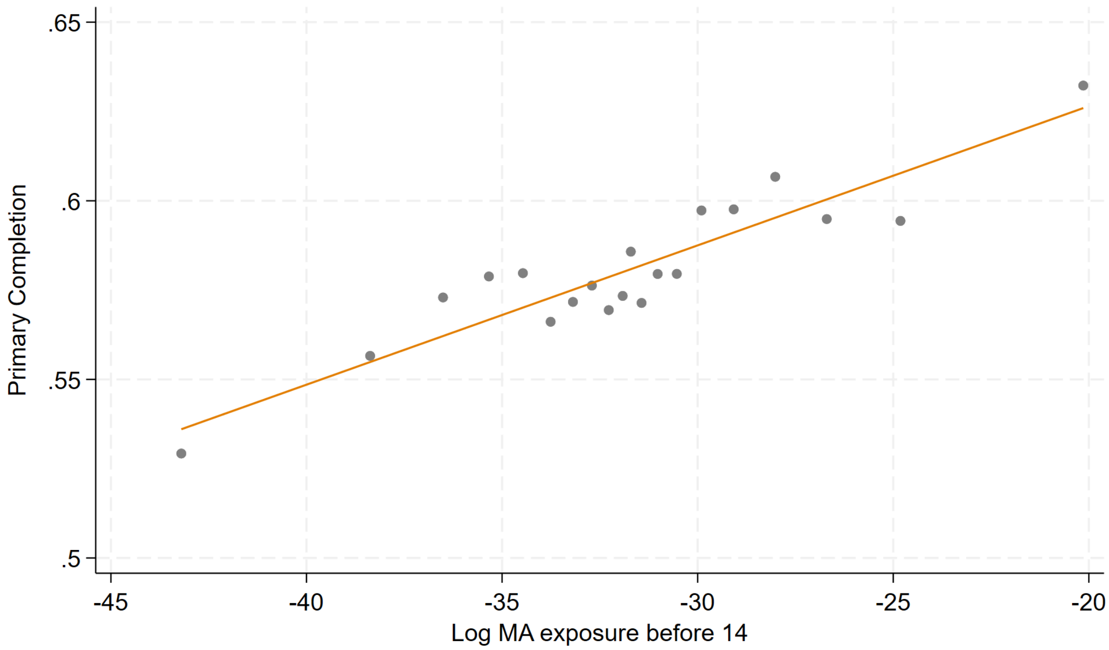
Where $\mathrm{MA}_{it} = \sum_{j \neq i} (\tau^m_{ijt})^{-\theta} L_{j0} \left(\mathrm{MA}_{jt}\right)^{-1}$ gives actual market access terms.

This approach does, however, have some limitations. This strategy will not be able to overcome endogeneity that occurs at the level of large geographies, for example, a program to build more roads in the south of the country to stimulate growth there. However I show in online appendix 8 that clientelism is not of first-order concern in this setting, and include region fixed effects. In addition, this approach, much like any that relies on market-access type measures, will suffer from the Borusyak and Hull [2023] critique of endogenous exposure to exogenous shocks. However, this is relatively easily overcome by permuting over possible roads, a procedure which is described in more detail below.

3.3 Results

Figure 2 shows the binscatter relationship between primary completion rates and log market access exposure before age 14, controlling for origin by cohort and age at move fixed effects, but without following a movers design or tackling the endogeneity of market access. Nevertheless, this figure shows a strong positive and approximately linear association, individuals who were more exposed to higher market access during the first 13 years of their childhood were more likely to complete primary education.

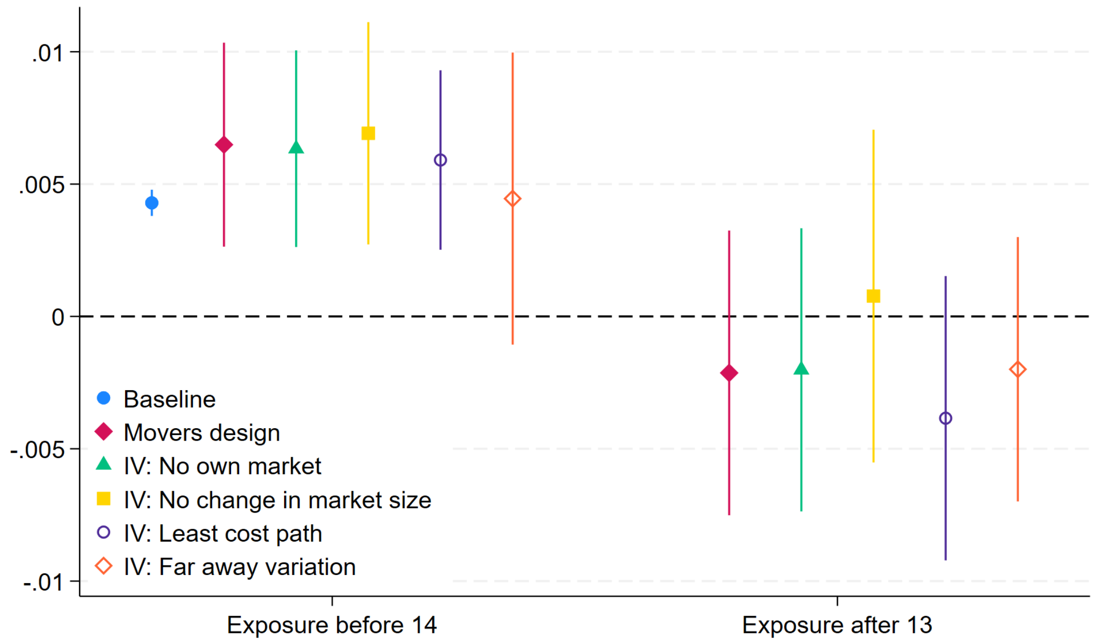
Figure 2 Binscatter relationship without instrumenting



Notes: This figure shows the non-parametric relationship between market access exposure before 14 and primary completion in a bin-scatter figure, conditional on origin by cohort and age at move fixed effects.

Figure 3 shows the results from estimating equation 10 on a sample of one-time movers. It displays coefficients $\widehat{\beta}_1$ and $\widehat{\beta}_2$ as well as their 95% confidence intervals. Each colour and marker shape corresponds to a different empirical strategy. Across all specifications, Figure 3 shows a clear positive impact of exposure to high market access on primary completion rates before 14, and a reassuringly null impact after.

Figure 3 Reduced form results



Notes: This figure plots the estimated β_1 and β_2 from estimating the following equation on a sample of one-time movers between the ages of 14 and 18: $\text{PrimaryCompletion}_k = \mathbb{1}_{[m_k \leq 13]} \cdot (\alpha_1 + \beta_1 \cdot (18 - m_k)) \cdot \Delta\text{MA}_{odc} + \mathbb{1}_{[m_k > 13]} \cdot (\alpha_2 + \beta_2 \cdot (18 - m_k)) \cdot \Delta\text{MA}_{odc} + \tau_m + \alpha_{oc} + \varepsilon_k$. Coefficients can be interpreted as the percentage point impact of spending an additional year in a one percent higher market access location. Each colour and marker shape represents a separate regression specification as indicated in the legend. Spikes represent 95% confidence intervals where standard errors are clustered at the origin and destination level. The first-stage Kleibergen-Paap rk Wald F statistics for the instrumental variable specifications are 5375, 106.6, 45.6, and 9.5, respectively.

Coefficients can be interpreted as the percentage point impact of spending an additional year in a one percent higher market access location. Therefore, the overall impact for non-movers in a given location of a one percent increase in market access is a $0.005 \times 13 \times 100 = 7$ percentage point increase in the probability of completing primary education. The standard deviation of log exposure is about 1, so moving to a 1 standard deviation better location (in logs) is roughly equivalent to moving to a 1 percentage point higher market access location. In my sample, on average 58% have completed primary school; therefore, moving to a one SD higher market access location increases the probability of completing primary school by 12%. This is a large and positive impact — high market access causes greater local opportunity.

3.4 Robustness and validity checks

In this section, I discuss four robustness and validity checks of the above results. Details of these tests and results can be found in the online appendix section 7.

First, the movers' design relies on the assumption that selection effects do not vary with the age at move. This cannot be tested in general; however, I can test whether selection on some observable characteristics varies with the age at move. To do this, I compare the "quality of move" measured as the change in location market access for those whose mothers have, and do not have, primary education, at different ages. If individuals with educated mothers systematically moved earlier (or later) to higher market access locations relative to those with non-educated mothers, this would be in violation of the identifying assumption. Figure 7 in the online appendix section 7 plots these results and finds no difference between those with and those without educated mothers.

Second, I include household fixed effects in equation 10, therefore only focusing on variation in age-at-move within a household across siblings. This specification obviates concerns around time-invariant household characteristics driving the observed relationship. Figure 8 in online appendix 7 shows the results. When including household fixed effects, I find that $\widehat{\beta}_1$ is slightly attenuated, but is far from being significantly so in either an economic or statistical sense.

Third, I show the robustness of the main result to varying specifications. Figure 9 shows that similar results are found if (i) I don't include origin by year born fixed effects but only year born fixed effects, (ii) I replace origin by year born fixed effects with origin fixed effects only, (iii) I remove age at move fixed effects, and (iv) I perform a PPML regression in levels.

Finally, I control for the possible non-random exposure to exogenous shocks baked into a market access type design Borusyak and Hull [2023]. Intuitively, this problem arises when, under random road placement, some locations will still expect to have higher or lower market access due to, for example, their initial location in the network (central locations would mechanically see higher gains in market access). To overcome this, I follow Borusyak and Hull [2023] and specify a data-generating process for market access. I construct 250 random road networks starting from the baseline in 1970 and randomly upgrading roads until the actual country-year change in travel time has been reached. I then calculate each location's market access over each random network in each period and use the average of these terms, a location's *expected* market access. I then control for this expected market access and repeat the analysis shown in Figure 3. The online appendix section 1 shows the results from this exercise in figure 1 — coefficients do not qualitatively or quantitatively change.

In addition to the above, one may have setting-specific empirical concerns. In the online appendix section 8, I show that in this setting, clientelism is not a threat to identification Burgess et al. [2015]. In the online appendix section 9, I show that top-coding due to primary completion rates nearing 100% is not a concern. Finally, in the online appendix section 10, I show that the prevalence of Koranic schools or Medersas remains too low to be of empirical relevance in my setting.

4 Quantifying the effect of roads built since 1970

Section 3 recovered the reduced form impact of changes in connectivity on local opportunity. However, to recover the aggregate impact of road building since 1970, we need to know the counterfactual market access under the scenario where no roads had been built. Denote by a prime counterfactual variables, then our object of interest is: $\mathrm{MA}'_i = \sum_j \left(\tau_{ij}^{m'}\right)^{-\theta} L'_j(\mathrm{MA}'_j)^{-1}$. To calculate this quantity, we need to know the counterfactual population distribution L'_i, which in turn depends on the full structure of the model.

To recover this object, I will proceed in four steps. First, I derive a parsimonious non-linear system of equations in each location that determines all endogenous variables. Second, to avoid backing out location fundamentals, I solve this system in differences using exact hat algebra. Third, I identify the remaining model parameters. Fourth, I check for the existence and uniqueness of equilibrium and solve the system for the given counterfactual road network.

4.1 Solving the system

We can solve this system into one of three endogenous variables in three equations in each location in each period. These three endogenous variables are (1) the stock of human capital in a given location H_{jt}, (2) market access MA_{jt} and (3) goods market access TMA_{jt}.

The stock of human capital is given by $H_j = \sum_i L_{it-1}\pi_{ijt}\mathbb{E}[e_i^\eta]$. Note that as $e_i = \left(\gamma \mathrm{MA}_i^{1/\theta}(\tau_i^{educ})^{-1}\right)^{1/(1-\eta)}\xi^{-1/(1-\eta)}$ we have that $\mathbb{E}[e_i^\eta] = \phi_\xi\left((\tau_i^{educ})^{-1}\mathrm{MA}_{it}^{1/\theta}\right)^{\eta/(1-\eta)}$ where $\phi_\xi = \gamma\mathbb{E}[\xi^{-\eta/(1-\eta)}]$. Into this expression we can substitute our expression for migration probabilities π_{ijt} and, simplifying, find the following.

$$H_{jt} = \phi_\xi a_{jt}^\theta \left(\frac{r_{jt}}{P_{jt}}\right)^\theta \sum_i L_{it-1}(\tau_{ijt}^m)^{-\theta}(\tau_{it}^{educ})^{\frac{-\eta}{1-\eta}}\mathrm{MA}_{it}^{\frac{\eta+\theta(1-\eta)}{\theta(1-\eta)}} \tag{11}$$

Into this expression we can substitute our known expressions for real wages and prices $r_{jt} = \left(\frac{\mathrm{TMA}_{jt}}{H_{jt}}\right)^{1/\sigma} B_{jt}^{\frac{\sigma-1}{\sigma}}$, and $P_{jt}^{1-\sigma} = \mathrm{TMA}_{jt}$ to find the following.

$$H_{jt}^{1+\frac{\theta}{\sigma}} = \phi_\xi a_{jt}^\theta B_{jt}^{\frac{\theta(\sigma-1)}{\sigma}}\mathrm{TMA}_{jt}^{\frac{\theta(1-2\sigma)}{\sigma(1-\sigma)}} \sum_i L_{it-1}(\tau_{ijt}^m)^{-\theta}(\tau_{it}^{educ})^{\frac{-\eta}{1-\eta}}\mathrm{MA}_{it}^{\frac{\eta+\theta(1-\eta)}{\theta(1-\eta)}} \tag{12}$$

Finally, subbing in our expressions for real wages into market access and goods market access terms we can write.

$$\mathrm{MA}_{jt} = \sum_i \left(a_{it}B_{it}^{\frac{\sigma-1}{\sigma}}\mathrm{TMA}_{it}^{\frac{1-2\sigma}{\sigma(1-\sigma)}}H_{jt}^{-1/\sigma}(\tau_{ijt}^m)^{-1}\right)^\theta \tag{13}$$

$$\mathrm{TMA}_{jt} = \sum_i \left(\tau_{ijt}^t\left(\frac{\mathrm{TMA}_{it}}{H_{it}B_{it}}\right)^{1/\sigma}\right)^{1-\sigma} \tag{14}$$

Equations 12, 13, and 14 define three equations in three unknowns for each location and can jointly be solved to determine all endogenous variables. We can write the resulting system in terms of changes using the exact hat algebra approach.

$$\widehat{H}_{jt}^{1+\frac{\theta}{\sigma}} = \widehat{\mathrm{TMA}}_{jt}^{\delta} \cdot \sum_i \omega_{ijt} \cdot \left(\widehat{\tau}_{ijt}^{m}\right)^{-\theta} \cdot \widehat{\mathrm{MA}}_{it}^{\beta}$$

$$\widehat{\mathrm{TMA}}_{jt} = \sum_i \rho_{ijt} \cdot \left(\widehat{\tau}_{ijt}^{t}\right)^{1-\sigma} \cdot \widehat{\mathrm{TMA}}_{it}^{\frac{1-\sigma}{\sigma}} \cdot \widehat{H}_{it}^{\frac{\sigma-1}{\sigma}}$$

$$\widehat{\mathrm{MA}}_{jt} = \sum_i \pi_{ijt} \cdot \left(\widehat{\tau}_{ijt}^{m}\right)^{-\theta} \cdot \widehat{\mathrm{TMA}}_{it}^{\delta} \cdot \widehat{H}_{it}^{-\frac{\theta}{\sigma}}$$

Where $\beta = \frac{\eta+\theta(1-\eta)}{\theta(1-\eta)}$ and $\delta = \frac{\theta(2\sigma-1)}{\sigma(\sigma-1)}$. Shares are defined as normal.

$$\text{Human capital share:} \quad \omega_{ijt} = \frac{H_{ijt}}{H_{jt}} = \frac{L_{it-1}(\tau_{ijt}^{m})^{-\theta}\mathrm{MA}_{it}^{\beta}}{\sum_q L_{qt-1}(\tau_{qjt}^{m})^{-\theta}\mathrm{MA}_{qt}^{\beta}}$$

$$\text{Trade share:} \quad \rho_{ijt} = \frac{X_{ijt}}{X_{jt}} = \frac{(\tau_{ijt}^{t})^{1-\sigma}\mathrm{TMA}_{it}^{(1-\sigma)/\sigma} H_{it}^{(\sigma-1)/\sigma}}{\sum_q (\tau_{qjt}^{t})^{1-\sigma}\mathrm{TMA}_{qt}^{(1-\sigma)/\sigma} H_{qt}^{(\sigma-1)/\sigma}}$$

$$\text{Migration share:} \quad \pi_{ijt} = \frac{L_{ijt}}{L_{jt}} = \frac{(\tau_{ijt}^{m})^{-\theta}\mathrm{TMA}_{jt}^{\theta\delta_\sigma} H_{jt}^{-\theta/\sigma}}{\sum_q (\tau_{iqt}^{m})^{-\theta}\mathrm{TMA}_{qt}^{\theta\delta_\sigma} H_{qt}^{-\theta/\sigma}}$$

We observed L_{ijt} and so π_{ijt} directly in the data. Using data on L_{it}, estimated market access terms and estimated iceberg migration cost, I can calculate ω_{ijt}. Finally, given ω_{ijt}, I can back out H_{it}, combine with estimates of iceberg trade costs, and calculate TMA_{it} to find ρ_{ijt}. To calculate TMA_{it}, I need data on output at the location-year level, Y_{it}. I find this by taking an Engle curve approach following Young [2012]; details can be found in online appendix 5.

This system of equations falls within the category described in Allen et al. [2024] and therefore, existence and uniqueness can be proved by relying on their results and is given when $\rho(|A|) < 1$ where $A = B\Gamma^{-1}$ and

$$\Gamma = \begin{pmatrix} 1 + \frac{\theta}{\sigma} & -\delta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 0 & \beta \\ \frac{\sigma-1}{\sigma} & \frac{1-\sigma}{\sigma} & 0 \\ -\frac{\theta}{\sigma} & \delta & 0 \end{pmatrix}$$

For the estimated parameter values, I find that $\rho(|A|) < 1$. Finally, to pin down the scale, I need to make a normalisation and choose to set the change in local opportunity to 0 for the location least affected by changes in connectivity, and therefore report results relative to this location. The base location is given by $i^* = \arg\min_i \left\{ \sum_j \hat{t}_{ij} \right\}$. Given parameter estimates σ, θ, η and data on the shares $\{\omega_{ijt}, \rho_{ijt}, \kappa_{ijt}\}$ I can then solve for endogenous variables for any given set of shocks to transport costs τ^t_{ijt} and τ^m_{ijt}. Therefore, two steps are remaining: first, to estimate parameters and second, to calculate shocks of interest to transport costs.

4.2 Identification of model parameters

We need to identify parameters η, σ, θ and trade and migration iceberg costs $T^t_{ijt} = (\tau^t_{ijt})^{\sigma-1}, T^m_{ijt} = (\tau^m_{ijt})^{\theta}$. In section 3 we already identified T^m_{ijt} as well as a bundle of η and σ. Therefore, it only remains to identify η, σ, and θ separately as well as T^t_{ijt}.

4.2.1 Estimating trade costs T^t_{ijt}

As with T^m_{ijt} I can parameterise trade costs such that $T^t_{ijt} = (\tau^m_{ijt})^{1-\sigma} = t^{1-\tilde{\sigma}}_{ijt}$ where $\tilde{\sigma} = a \cdot \sigma$. Although data on migration is remarkably rich, covering a large time span and granular geography, allowing estimation of T^m_{ijt} with confidence — no analogous data exists for trade across space in this setting. As a result of this, I am forced to calibrate values of the parameters governing trade in goods across space, ϕ and a, to values commonly found in the literature. Following Morten and Oliveira [2024] who leverage variation in trade across Brazilian states I take $\tilde{\sigma} = 2.9$, and following

Simonovska and Waugh [2014] I take $\sigma = 5$, and therefore $a = 0.58$.

4.2.2 Separately identifying θ and η

Following the strategy employed by Bryan and Morten [2019], I use model-implied regressions to separately identify θ and η. Note that the expected wage individuals from i who move to j attain is given by $\mathbb{E}[w_{jt}|i] = \gamma r_{jt} h_{it} \pi_{ijt}^{-1/\theta}$ an expression that motivates the following regression.

$$\ln(\bar{w}_{ijt}) = \alpha_{it} + \gamma_{jt} - \frac{1}{\theta}\ln(\pi_{ijt}) + v_{ijt} \tag{15}$$

This regression will therefore allow me to find estimates for θ which, coupled with our previous estimates of $\frac{1}{\theta(1-\eta)}$, will allow me to estimate η. However, estimates from equation 15 may be biased. As discussed in Bryan et al. [2014], any shock that affects the wages that individuals from i receive in j will also impact migration from i to j, for example, if demand for skills generated in i increases in j. To overcome this I use the instrumental variable approach suggested by Bryan and Morten [2019] and instrument $\ln(\pi_{ijt})$ with $\ln(\pi_{\neg ijt})$, where $\pi_{\neg ijt}$ is the proportion of people from other origins who migrate to j in period t.

Table 2 shows the results from estimating equation 15 with various specifications. Table 2 also shows the implied values of $\widehat{\theta}$ and $\widehat{\eta}$, combining information with that from estimating equation 10 with corresponding standard errors calculated using the Delta method. Column two is our baseline specification where $\ln(\pi_{ijt})$ is instrumented as discussed above. I find the following parameter estimates $\theta = 22.64\ (2.214), \eta = 0.434\ (0.170), \sigma = 5$. This compares to Hsiao [2024] who finds $\theta = 20, \eta = 0.224$, and Bryan and Morten [2019] who find $\theta = 28$.

Table 2 Identifying θ and η

	(1)	(2)	(3)	(4)
$\ln(\pi)$	-0.0411***	-0.0442***	-0.0447***	-0.0473***
	(0.00356)	(0.00432)	(0.00425)	(0.00432)
θ	24.30	22.64	22.36	21.15
	(2.11)	(2.21)	(2.12)	(1.93)
η	0.47	0.43	0.43	0.39
	(0.16)	(0.17)	(0.17)	(0.18)
IV		X	X	X
F-stat		3247	3379	3349
FE	ot,dt	ot,dt	o,d,t	o,d
Obs	13999	13999	13999	13999

Notes: This table shows the results from estimating equation 15 and the implied values of $\widehat{\theta}$ and $\widehat{\eta}$ combining information with that from estimating equation 10 with corresponding standard errors calculated using the Delta method. Column (1) shows results with no instrumenting, including origin-time and destination-time fixed effects. Columns (2) to (4) show results instrumenting following Bryan and Morten [2019] with various fixed effects specifications. Column (2) includes origin-time and destination-time fixed effects and is our baseline specification, column (3) includes origin, destination, and time fixed effects and finally column (4) only includes origin and destination fixed effects.

4.3 Quantification results

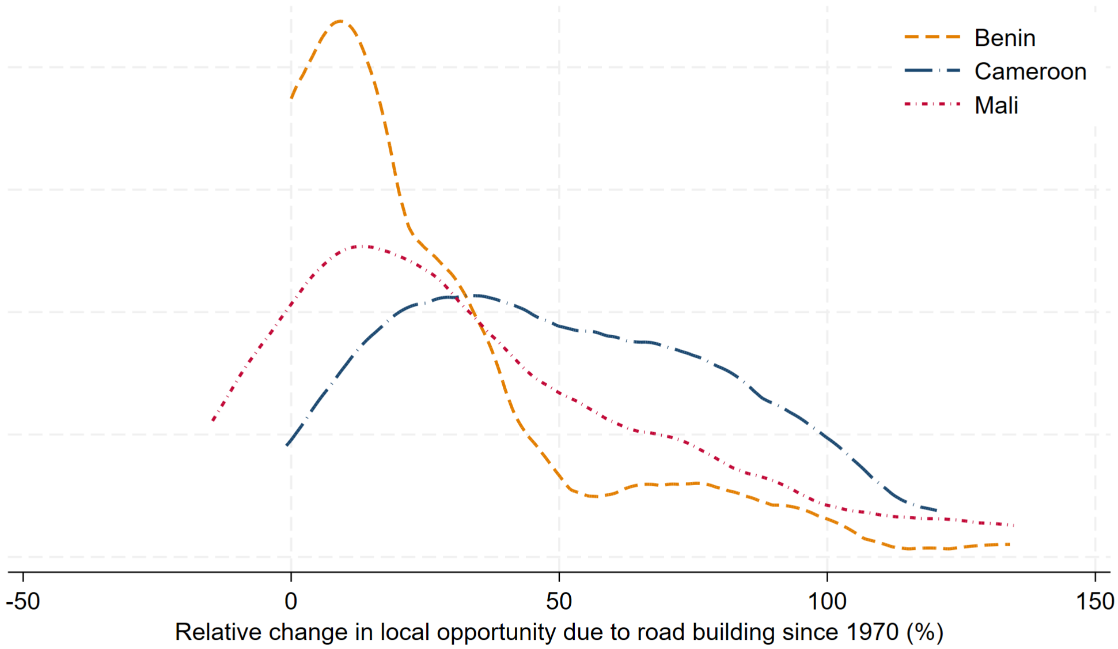
Using the full structure of the model, I can quantify the causal effect of road building between 1970 and 2020 on the distribution of local opportunity. I find that, on average, road building increased local opportunity. Due to road building since 1970, the causal effect of place on primary completion rates increased by, on average, 33%. This corresponds to a 0.53 percentage point higher average annualised growth rate. To understand the magnitude of this effect, note that over the same period across Benin, Cameroon, and Mali, the average annualised growth rate in primary education completion was 1.82 percent.[9]

Figure 4 shows how this average effect varies across locations. In each country, there is significant variation in the impact of road building since 1970 on local oppor-

[9]This average weights each location equally. Data from the World Bank available here.

tunity. Some locations saw their opportunity double, whereas others saw no change or even a negative change.

Figure 4 Aggregate effects of road building since 1970



Notes: This figure shows the distribution over locations of the impact of road building since 1970 on the causal effect of place measured in percentage change on baseline.

These results suggest a sizable (average) effect of changes in connectivity since 1970 on primary completion rates. As discussed in the online appendix section 4, travel times have decreased by on average 33% over this period, a commensurate magnitude. Nevertheless, these results suggest that remoteness and a lack of market integration are key factors suppressing local education completion, and thus causing inequality of opportunity across space in Benin, Cameroon, and Mali. The policy implications are that increasing connectivity, through road building, is an effective way of shaping local opportunity in Benin, Cameroon, and Mali.

5 Conclusion

This paper studies how connectivity of space shapes this geographical distribution of opportunity, and therefore how policymakers can affect spatial inequality of opportunity through road building, in the setting of Benin, Cameroon, and Mali.

To study the impact of road building, I develop an approach to measure the effect of any given road on locations across the entire network. This is challenging because not only does it matter what a given road connects, but roads will impact outcomes in all locations across the network. To overcome these challenges, but remain as general as possible, I turn to theory and develop a sufficient statistic approach that is consistent with a broad class of data-generating processes. This result endogenises skill premia across localities in a many-location setting with costly movement of goods and individuals over space, and education choice. It states that a location's market access captures all the potentially complex effects of roads on local opportunity.

The sufficient statistic result suggests an expression that can be directly taken to the data but requires an identification strategy to overcome the endogeneity of road placement, which market access terms inherit, and the endogenous sorting of individuals across space. Combining a mover's design and a far-away or not-on-least-cost-path identification strategy, I find that a one standard deviation increase in market access on average increases the probability of completing primary education by 7 percentage points (12%). This result is robust across reasonable specifications, to only using within-household variation in movers, and correcting for endogenous exposure to exogenous shocks Borusyak and Hull [2023].

Leveraging the full structure of the model and estimating parameters, I then ask what the aggregate impact of road building since 1970 has been on the spatial distribution of opportunity. I find that road building since 1970, on average, increased local causal place effects by 33% — but that this average hides considerable heterogeneity, with some locations barely changing but others over doubling.

In sum, this paper finds that road building impacts the causal effect growing up in a given location has on a child's probability of completing primary school. However, road building is not a silver bullet — locations are differentially affected, and the impact could even be negative.

References

Tasso Adamopoulos. Spatial integration and agricultural productivity: Quantifying the impact of new roads. *American Economic Journal: Macroeconomics*, 17(1): 343–378, 2025.

Anjali Adukia, Sam Asher, and Paul Novosad. Educational investment responses to economic opportunity: evidence from indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–76, 2020.

Alberto Alesina, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou. Intergenerational mobility in africa. *Econometrica*, 89(1):1–35, 2021.

Treb Allen and Costas Arkolakis. Economic activity across space: a supply and demand approach. *Journal of Economic Perspectives*, 37(2):3–28, 2023.

Treb Allen, Costas Arkolakis, and Yuta Takahashi. Universal gravity. *Journal of Political Economy*, 128(2):393–433, 2020.

Treb Allen, Costas Arkolakis, and Xiangliang Li. On the equilibrium properties of spatial models. *American Economic Review: Insights*, 6(4):472–489, 2024.

Sam Asher and Paul Novosad. Rural roads and local economic development. *American economic review*, 110(3):797–823, 2020.

David Atkin. Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85, 2016.

Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 145:102442, 2020.

Nathaniel Baum-Snow, Loren Brandt, J Vernon Henderson, Matthew A Turner, and

Qinghua Zhang. Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448, 2017.

Nathaniel Baum-Snow, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt. Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115:103124, 2020.

Kirill Borusyak and Peter Hull. Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185, 2023.

Wyatt Brooks and Kevin Donovan. Eliminating uncertainty in market access: The impact of new bridges in rural nicaragua. *Econometrica*, 88(5):1965–1997, 2020.

Gharad Bryan and Melanie Morten. The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, 127(5):2229–2268, 2019.

Gharad Bryan, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748, 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.

Marie Castaing Gachassin. Should i stay or should i go? the role of roads in migration decisions. *Journal of African Economies*, 22(5):796–826, 2013.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162, 2018a.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, 2018b.

Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

Eric Chyn and Diego Daruich. An equilibrium analysis of the effects of neighborhood-based interventions on children. Technical report, National Bureau of Economic Research, 2022.

Nathan Deutscher. Place, peers, and the teenage years: Long-run neighborhood effects in australia. *American Economic Journal: Applied Economics*, 12(2):220–49, 2020.

Dave Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934, 2018.

Dave Donaldson. Blending theory and data: A space odyssey. *Journal of Economic Perspectives*, 36(3):185–210, 2022.

Dave Donaldson and Richard Hornbeck. Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016.

Fabian Eckert and Tatjana Kleineberg. The geography of opportunity: Education, work, and intergenerational mobility across us counties. Technical report, Technical report, Working Paper, 2024.

Eric V Edmonds, Nina Pavcnik, and Petia Topalova. Trade adjustment and human capital investments: Evidence from indian tariff reform. *American Economic Journal: Applied Economics*, 2(4):42–75, 2010.

Benjamin Faber. Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070, 2014.

Vivien Foster and Cecilia M Briceño-Garmendia. Africa's infrastructure: a time for transformation. 2009.

Junichi Fujimoto, David Lagakos, and Mitchell VanVuren. *Aggregate and Distributional Effects of'Free'Secondary Schooling in the Developing World*. National Bureau of Economic Research, 2023.

Ejaz Ghani, Arti Grover Goswami, and William R Kerr. Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal*, 126(591):317–357, 2016.

Ken Gwilliam. *Africa's transport infrastructure: Mainstreaming maintenance and management*. The World Bank, 2011.

Allan Hsiao. Educational investment in spatial equilibrium: Evidence from indonesia. 2024.

IPUMS. Integrated public use microdata series, international: Version 7.2 [dataset]. 2020.

Remi Jedwab and Alexander Moradi. The permanent effects of transportation revolutions in poor countries: evidence from africa. *Review of economics and statistics*, 98(2):268–284, 2016.

Remi Jedwab and Adam Storeygard. The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960-2010. *Journal of the European Economic Association*, 2021.

Remi Jedwab, Edward Kerby, and Alexander Moradi. History, path dependence and development: Evidence from colonial railways, settlers and cities in kenya. *The Economic Journal*, 127(603):1467–1494, 2017.

Hundanol A Kebede. Gains from market integration: Welfare effects of new rural roads in ethiopia. *Journal of Development Economics*, 168:103252, 2024.

Gaurav Khanna. Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *The American Economic Review*, 2022.

Jean-William P Laliberté. Long-term contextual effects in education: Schools and neighborhoods. *American Economic Journal: Economic Policy*, 2021.

Guy Michaels. The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701, 2008.

Luke Heath Milsom. Spatial inequality of opportunity in west africa. *Journal of Public Economics*, 225:104944, 2023.

Niclas Moneke. Can big push infrastructure unlock development? evidence from ethiopia. Technical report, Working paper, 2020.

Melanie Morten and Jaqueline Oliveira. The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, 16(2):389–421, 2024.

Dozie Okoye, Roland Pongou, and Tite Yokossi. New technology, better economy? the heterogeneous impact of colonial railroads in nigeria. *Journal of Development Economics*, 140:320–354, 2019.

George Psacharopoulos and Harry Anthony Patrinos. Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5): 445–458, 2018.

Stephen J Redding and Esteban Rossi-Hansberg. Quantitative spatial economics. *Annual Review of Economics*, 9:21–58, 2017.

Stephen J Redding and Matthew A Turner. Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, 5: 1339–1398, 2015.

Fernanda Catalina Rojas Ampuero. *Sent Away: The Long-Term Effects of Slum Clearance on Children and Families*. PhD thesis, UCLA, 2022.

JMC Santos Silva and Silvana Tenreyro. The log of gravity. *The Review of Economics and statistics*, 88(4):641–658, 2006.

Ina Simonovska and Michael E Waugh. The elasticity of trade: Estimates and evidence. *Journal of international Economics*, 92(1):34–50, 2014.

Sebastian Sotelo. Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738, 2020.

UN. 2018 revision of world urbanization prospects, 2018.

Raoul van Maarseveen. The effect of urban migration on educational attainment: evidence from africa. *Available at SSRN 3836097*, 2021.

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.

Yoto V Yotov, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization Geneva, 2016.

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

Wouter Zant. Measuring trade cost reductions through a new bridge in mozambique: Who benefits from transport infrastructure? *Journal of African Economies*, 31(4): 384–408, 2022.