The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010

Remi Jedwab and Adam Storeygard*

September 25, 2017

Abstract

Previous work on transportation investments has focused on average impacts in high- and middle-income countries. We estimate average and heterogeneous effects in a poor continent, Africa, using roads and cities data spanning 50 years in 39 countries. Using changes in market access due to distant road construction as a source of exogenous variation, we estimate an 30-year elasticity of city population with respect to market access of 0.05-0.20. Our results suggest that this elasticity is stronger for small and remote cities, and weaker in politically favored and agriculturally suitable areas. Access to foreign cities matters little.

JEL Codes: R11; R12; R4; O18; O20; F15; F16

Keywords: Transportation Infrastructure; Paved Roads; Urbanization; Cities; Africa; Market Access; Trade Costs; Highways; Internal Migration; Heterogeneity

---

*Corresponding author: Adam Storeygard: Department of Economics, Tufts University, adam.storeygard@tufts.edu. Remi Jedwab: Department of Economics, George Washington University, jedwab@gwu.edu. We thank Treb Allen, Nathaniel Baum-Snow, Kristian Behrens, Gharad Bryan, Kerem Cosar, Victor Couture, Dave Donaldson, Gilles Duranton, Benjamin Faber, Edward Glaeser, Vernon Henderson, Melanie Morten, Paul Novosad, Elias Papaioannou, Harris Selod, Matthew Turner, Anthony Venables, Leonard Wantchekon and seminar audiences at the African Econometric Society meetings (Kruger), African School of Economics, Berkeley Haas, Boston Fed, CGD, CSAE (Oxford), George Mason, Harvard, IGC (LSE), IMF, Merced, NBER SI, NEUDC (Brown), North Dakota, Sciences Po, Stanford, UB/IEB, UQAM, UVA, World Bank, the World Bank Secondary Towns Conference, and the World Bank-GWU Urbanization Conference for helpful comments, and Yasmin Abisourour, Karen Chen, Rose Choi, Taher Elsheikh, Yury Higuchi, Erin McDevitt, and Emily Ryon for research assistance. We are grateful to Uwe Deichmann and Siobhan Murray for sharing their roads data, François Moriconi-Ebrard for help with data collection, and Durwood Marshall for programming advice. We thank the World Bank's Strategic Research Program on Transport Policies for Sustainable and Inclusive Growth, the Elliott School of International Affairs (SOAR Research Grant) and the Institute for International Economic Policy at GWU, and the Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development, for financial assistance.

We consider the effect of roads upgraded between 1960 and 2010 on city population growth in 39 sub-Saharan African countries during that period, as a result of increased market access to other cities. Using a novel instrumental variables strategy based on road changes faraway to account for potential endogeneity of market access, we find that a 10% increase in market access induces a 0.5–2% increase in city population on average over the course of the 30 years after market access changes. The OLS effect is smaller, suggesting that far from anticipating future growth, roads may be more often built in otherwise lagging regions. This is consistent with a network that is expanding from the largest cities at independence to poorer, more remote places later. Our approach allows us to explore heterogeneous effects across time and space. Effects are roughly constant across the first three decades of road-building, subsequently falling. Our results then suggest that effects are larger for smaller and more isolated cities, and market access changes to domestic rather than foreign cities, and weaker in politically favored and more agriculturally suitable areas.

Sub-Saharan Africa is an important context for studying roads and cities. It is the least urbanized world region, as well as the one with the least developed transport network. Its urbanization rate crossed one third as the global rate crossed one half in the past decade (United Nations, 2015). The region's 3.4 km of roads, 0.7 km of them paved, per 1000 residents, represent less than half and one fifth of the respective global averages (Gwilliam, 2011). The region's transport infrastructure is also limited compared to other developing regions. Road density is less than a third of South Asia's, and only a quarter of the network is paved (World Bank, 2010a), against 60% in India (Government of India, 2016) and two-thirds in China (World Bank, 2015). This combination of low urbanization and poor connectivity means that many people lack access to national and global markets (Limão and Venables, 2001; Atkin and Donaldson, 2015). While road construction was rapid in the 1960s and 1970s post-Independence, it slowed substantially in the subsequent decades along with overall public investment.

African countries have begun to make large transportation investments again.[1] International donors increasingly consider these projects in the context

[1]The Economist. 2015. "African roads and rails: All aboard." Print edition, 28 February.

of a Trans-African Highway (TAH) system, and describe them as having the potential to transform their regions (ADB and UNECA, 2003). For example, the World Bank writes of a project connecting Abidjan and Lagos: "The potential of the corridor to become a catalyst for economic growth and regional integration in the sub-region is well documented" (World Bank, 2010b). It is thus imperative to consider the effect of earlier road construction on the economic geography of the region, with a view to understanding the effect of future projects.

Our work relates primarily to the empirical literature on the effect of market access, and specifically intercity transport costs, on the growth of local areas in developing countries (e.g. Banerjee et al., 2012; Faber, 2014; Storeygard, 2016; Jedwab and Moradi, 2016; Donaldson and Hornbeck, 2016; Donaldson, forthcoming) (for comprehensive overviews of the literature, see Redding and Turner (2015) and Berg et al. (2015)).[2] More generally, a large literature has looked at how market access affects the growth of neighborhoods (Ahlfeldt et al., 2015), cities (Redding and Sturm, 2008), regions (Hanson, 1998), and countries (Feyrer, 2009). Another large literature has looked at the effect of large highway projects on a variety of outcomes (e.g. Rothenberg, 2013; Ghani et al., 2016; Coşar and Demir, 2016; Baum-Snow et al., 2017a,b). Finally, a smaller literature has emphasized the specific role of road quality, which is the main source of variation in this work (e.g. Casaburi et al., 2013; Gertler et al., 2015; Asher and Novosad, 2016). The paper makes several contributions to this literature.

First, we document the development of a 140,000 km continental paved road network from near its beginnings to the present. This data richness allows us to consider the timing of effects in ways that previous work, which is mostly based on two or three cross-sections instead of our six over 50 years, cannot. We also use the universe of paved and improved roads, as opposed to highways alone as considered by many studies, and study an evolution of the road network rather than a revolution of the kind that China has experienced since 1988, building

Transport also accounted for 14% of World Bank lending, and 22% of African Development Bank disbursements 2012–2015 (World Bank, 2016; African Development Bank, 2012–2015).

[2]For studies on developed countries, see Chandra and Thompson (2000); Baum-Snow (2007); Michaels (2008); Duranton and Turner (2012); Duranton et al. (2014); Behrens et al. (2016).

35,000 km of highways (Faber, 2014; Baum-Snow et al., 2017a,b). To the extent that gradual evolution is more likely in the future of developing regions, this is a distinct and instructive context. There are also studies on rural transportation (Casaburi et al., 2013; Bryan et al., 2014; Stanig and Wantchekon, 2015; Asher and Novosad, 2016). However, while rural (earthen) roads programs strongly impact villages, they are much less costly than intercity (non-earthen) road investments.[3] We also hope that our novel dataset will allow researchers and policy makers to better understand the constraints on Africa's development.

Second, building on Donaldson and Hornbeck (2016), we develop a novel identification strategy, relying on the variation in market access induced by roads built far away. With respect to the typology of identification strategies introduced by Redding and Turner (2015), this is not a context in which comprehensive planned or historical networks are available, our scope limits the possibility of randomized experiments and regression discontinuity designs, and the inconsequential places approach is also not appropriate because of the piecemeal nature of much of the road construction. However, our identification strategy has the advantage of being implementable in most contexts, which could facilitate the comparison of effects across countries and over time.[4]

Third, we consider a wide variety of heterogeneous effects, a subject that has received less attention in the literature, and could be especially important given the region's heterogeneity in physical, economic, and political geography. We find suggestive evidence that the effect of market access is stronger for cities that are: (i) small and remote;[5] (ii) surrounded by poor agricultural land;[6] and (iii) less

[3] Collier et al. (2015) estimate that highways, paved roads and improved roads are 31, 8 and 4 times more expensive than earthen roads, implying that the total cost of road upgrades in our sample was 61.5% of 2000 regional GDP. By comparison, the large rural road program studied by Asher and Novosad (2016) cost 1.8% of India's GDP in 2015.

[4] Strategies based on planned/historical networks or accidental connections can explain well the location of road investments, but often not their timing, limiting inference about the timing of effects. Randomized experiments and regression discontinuity designs have only been used to study rural roads, since their implementation is harder for intercity roads.

[5] This suggests that roads contributed to the decentralization of economic activity in our context, in line with some work (Redding and Sturm, 2008; Banerjee et al., 2012; Rothenberg, 2013), but less so with recent papers on China (Faber, 2014; Baum-Snow et al., 2017b).

[6] This is consistent with Ricardian internal trade models, and echoes Asher and Novosad (2016), who find that roads cause outmigration from villages with low agricultural productivity.

likely to be politically favored.[7] Effects are driven primarily by access to domestic cities, and ports, suggesting a role for access to overseas markets.[8] Together, they suggest that the impact of transportation investments varies by context.

An important recent literature has estimated the effects of transport infrastructure investment within a general equilibrium trade model (Allen and Arkolakis, 2014, 2016; Fajgelbaum and Redding, 2014; Alder, 2015; Morten and Oliveira, 2017). This is not feasible in our environment, where data are substantially less available, even compared to the middle-income developing countries previously studied.[9] In particular, no data on within-country variation in trade, migration, production, wages, prices and amenities are available for more than a small subset of our sample over time.

Our work also builds on the literature considering how cities in developing countries grow. Previous work on transport and city growth in Africa has emphasized the earlier railroad revolution (Jedwab and Moradi, 2016; Jedwab et al., 2017a) or variable costs of road transport (Storeygard, 2016), but not road construction, which is likely to have a larger effect on transport costs in the future. Other work on urbanization in Africa is primarily cross-country in nature and does not consider variation across cities within countries (Fay and Opal, 2000; Henderson, Roberts and Storeygard, 2013; Gollin, Jedwab and Vollrath, 2016; Jedwab, Christiaensen and Gindelsky, 2017b; Jedwab and Vollrath, 2017).

1. Data and Background

We focus on mainland sub-Saharan Africa, for which we create a new spatial dataset of roads and cities over fifty years: 199,814 cells of 0.1x0.1 degrees (≈

[7]This is consistent with the literature documenting political motivations in the allocation of roads (Knight, 2004; Burgess et al., 2015; Blimpo et al., 2013). If roads are built in some areas because of their marginal political returns rather than their economic returns, they may be less effective in promoting growth (Tanzi and Davoodi, 1998). Our analysis is then based on a new dataset reporting the place of origin of the 189 heads of state of the 39 countries in 1960-2010. To our knowledge, we present the first such dataset covering virtually all of sub-Saharan Africa.

[8]While the stronger role of access to world markets is in line with Fajgelbaum and Redding (2014) and Baum-Snow et al. (2017a), the differential seems to be smaller in Africa, possibly due to oligopolistic intermediaries (Atkin and Donaldson, 2015).

[9]Among the 39 sub-Saharan African countries in 2015, median per capita GDP was about \$2,000 only (PPP, current international \$), much less than for other developing countries studied in the literature (Brazil: \$16,000; China: \$14,000; India: \$6,000; Indonesia: \$11,000).

11x11 km) for 42 countries every 10 years between 1960 and 2010. In our econometric analysis, we will focus on the 2,789 cells that reached an urban population of at least 10,000 at some point since 1960 in 39 of these countries. Sections A.1-A.4 of the Appendix provide further details on the data.

1.1. Roads, 1960-2010

We combine information from two sets of sources. First, Nelson and Deichmann (2004) provides road locations for all of Africa. These data nominally represent roads existing in 2004, based primarily on the US government's Digital Chart of the World database, with limited information on road type. Second, using these road locations as a baseline, we digitized 64 Michelin road maps produced between 1961 and 2014 to represent contemporary road conditions for three broad regions: Central/South (19 countries), North/West (18) and North/East (5). Appendix Figures A.1 and A.2 show the countries and years, respectively, covered by each region. The average gap between maps across regions is under 2.5 years, and the longest is 7 years. While specific road categories vary somewhat across maps, the distinction between *highways*, other *paved roads*, *improved roads* (laterite or gravel), and *dirt roads* is nearly universal.

The Michelin maps report highways and intercity paved and improved roads comprehensively, but their coverage of earthen roads is less complete, with some changes clearly due to coverage changes as opposed to new roads. Based on the assumption that roads change quality but rarely move or disappear, we thus code each segment from the Nelson and Deichmann (2004) map as paved or improved in each year it is shown as such by Michelin, and assume that the remaining segment-years are earthen. We also code a small number of segments as highways in the eight countries where they appear after 1973.[10]

Michelin uses four sources to create the maps: (i) the previous Michelin map, (ii) government road censuses/maps, (iii) direct information from its tire stores across Africa, and (iv) correspondence from road users including truckers.[11]

[10]See Section A.1 in the Appendix for details on the road data. Appendix Figure A.3 shows the Michelin map for Sierra Leone in 1969, as well as the associated GIS map.
[11]This paragraph is based on our discussions with Michelin employees.

The latter two sources of information are especially important, and new to this literature.[12] Michelin has been producing maps since 1910, with its first map for West Africa appearing in 1938. As one of the largest tire companies in the world since the early 1970s (Rajan et al., 2000), unlike other organizations producing maps, Michelin has long maintained a large network of stores distributing its tires, in addition to its maps. Many truck drivers in Africa use both, and are in regular contact with this network. Because inaccurate characterization of road surface leads to delays or truck damage, truckers complain to the store managers when the information is inaccurate, and the store managers relay this information to Michelin cartographers. Michelin also focuses on road surfaces whereas other maps classify roads as primary/secondary or major/minor, which is less informative about road quality. We are unaware of another source of maps with similarly broad coverage over such a long period.

We believe that this process leads to generally consistent information across countries and time, but this does not mean that the evolution of every road segment is perfectly characterized. This has several implications. First, this revision process means that changing conditions may be reflected in the maps with a lag. The lag is unlikely to be long because: (i) Michelin dealers collect data on ongoing projects and their maps are intended to reflect the year a road will open and (ii) periods between maps are generally short. Second, Michelin's network is more sparse in some countries and periods. Country-year fixed effects should ameliorate the effect of this to some extent. Coverage of the early 1960s is more limited; as we show, results are robust to excluding the decades affected by 1960s roads. Finally, we cannot capture the quality of roads within a surface class, so when a severely potholed paved road is resurfaced, our data do not reflect this. This work may have been especially prevalent since 2000, as we explain below, so we may underestimate recent changes. Results are robust to excluding the 2000s.

1.2. City Location and Population, 1960-2010

We obtained location and population estimates of cities in 33 countries from *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa.*[13] These

[12]Burgess et al. (2015) use these data for Kenya 1964–2002 alone.

[13]http://www.africapolis.org; 15 countries are from part I and 18 from part II.

sources generated estimates using various sources including population censuses, "non-native" population censuses, demographic studies, administrative counts, electoral counts, and statistical abstracts. Based on an initial list of cities with at least 5,000 inhabitants in the most recent census circa 2000, their final database nominally includes all cities that reached a population of at least 10,000 at some point since 1960. They also define agglomerations in circa 2000 using satellite imagery. If two distinct cities in 1970 ultimately merged, in the sense that their urban land cover is contiguous, they are treated as one city in Africapolis throughout. Thus we are not studying reallocation *within* urban areas.

We build on the Africapolis data in three ways. First, we use analogous sources to produce an analogous database for 6 southern African countries not in the Africapolis samples.[14] Second, we add a small number of missing cities in Africapolis countries that achieved a population over 10,000 at some point between 1960 and 2010. Finally, we add missing locations, and corrected locations that appeared to be incorrect, based on Google Earth, GeoNet, and Wikipedia, aggregating multiple administrative cities into one agglomeration using more recent satellite imagery from Google Earth.[15]

Population figures for all cities are exponentially interpolated and extrapolated between raw data years to obtain estimates for each year (1960, 1970, 1980, 1990, 2000 and 2010). The resulting sample includes population estimates for all cities with a population of over 10,000 at some point since 1960, in all sample years in which their population exceeded 10,000, and in 60% of sample years in which they did not reach 10,000. Information on smaller cities is not systematically available for our sample region and period. We are thus studying the intensive margin of the growth of cities over 10,000. We do not consider their entry into the sample, as we do not have consistent information on whether that entry involves growing from 9,990 or 1,000 to 10,000 in the previous decade. Over our sample period 1960–2010, 84% of urban population growth, representing 171

[14]Comparable cities data are not available for South Africa. In calculating measures of market access for the remaining 39 countries, we do however include the 20, 1, and 1 largest (in 2010) cities in South Africa, Lesotho and Swaziland, to minimize bias in measures for cities near them.

[15]Details on the city data can be found in Section A.2 of the Appendix. The sources used for all countries are listed in Appendix Table A.1.

of 203 million new urban residents, was on this intensive margin.

1.3. Other data

We compile several additional datasets and assign them to cells: (i) the names and the location of national and provincial capitals in both 1960 (N = 346) and 2010 (N = 481); (ii) the location of open mines (incl. fields) between 1960 and 2010 (N = 288); (iii) land suitability for food/cash/all crops today (assuming low input levels and no irrigation); (iv) average rainfall in 1900-1960; (v) the locality of origin and ethnicity of the 189 heads of state of the 39 sample countries between 1960 and 2010, and historical spatial boundaries of ethnic groups; (vi) the location of navigable rivers; (vii) the location of railroad lines and when each line was built; (viii) the location of 65 and 44 international ports in 1960 and 2005 respectively; (ix) the location of 466 airports in 2007; (x) the location of 837 customs posts circa 2010; (xi) the location of natural parks covering 26,252 cells circa 2015; and (xii) the mean and standard deviation of altitude.

We also obtained country-level data on: (i) the population, per capita GDP (1990 International Geary-Khamis $) and polity score – a measure of democratization – of the country in each year; and (ii) whether the country was still a colony, experienced an international/civil war, hosted refugees, or suffered a multi-year drought in each decade (see Appendix Section A.4 for details).

1.4. Aggregate Patterns in Road Building and Urban Growth

Figure 2a shows aggregate lengths of highways and paved and improved roads over time, and Figure 2b shows their cumulative shares, assuming a constant stock of total roads as measured circa 2004. In 1960, a length of less than 5% of today's network was paved. Following the independence of most African countries in the early 1960s and into the 1970s, the paved network expanded much more rapidly, fueled by massive public investments (e.g. O'Connor, 1978; Wasike, 2001; Pedersen, 2001). The stock of improved roads also increased in the 1960s, but it decreased in the 1970s as more initially improved roads were paved.

Beginning in the mid-1980s, worsening macroeconomic conditions decreased the pace of road transformation markedly (Konadu-Agyemang and Panford, 2006;

Gwilliam, 2011). Although investment may have increased again since the mid-2000s, this is not reflected in our data. We believe this is because investment may have been directed primarily towards restoring and rebuilding existing paved roads. As explained by World Bank (1988) and Konadu-Agyemang and Panford (2006), roads deteriorated badly in most African countries in the 1980s and after, as road maintenance agencies were systematically underfunded.[16]

Figures 3a and 3b map cities over 10,000 in 1960 and 2010. The sheer number of such cities has increased dramatically, from 418 in 1960 to 2,859 in 2010. In 1960, a large fraction of these cities were trading centers or regional administrative centers established by colonial administrations (Bairoch, 1988; Coquery-Vidrovitch, 2005). The urban population of the 39 countries, here defined as the total population of all cities over 10,000, has increased from less than 25 million in 1960 to almost 250 million in 2010. The analogous urbanization rate increased from only 9% in 1960 to 28% in 2010.

City population is a convenient measure of local economic development, when migration accommodates spatial equilibrium. It is also of interest in its own right (see for example DeLong and Shleifer, 1993 and Acemoglu et al., 2005). No subnational GDP or wage data exist for most countries in the sample. Even total population (and therefore urbanization rate) is often available only for coarse regions and more extrapolated in early periods.[17] Thus, city population is the best available measure of local economic development for Sub-Saharan Africa from 1960 to date. For a subsample of decades, we consider night lights, available from 1992, as a proxy for city income, following Storeygard (2016).[18]

[16]For example, in Kenya, the government has invested a lot of money rebuilding the main paved road from Mombasa to Nairobi to its original 1960s standard (Burgess et al., 2015). See Jedwab and Storeygard (2017) for evidence that our data are consistent with other sources at the country level, even for the most recent period.

[17]Henderson et al. (2017) use information on populations for subnational units of 89 censuses in 29 countries. These data are not consistently available back to the 1960s for most countries.

[18]While Young (2013) uses household asset ownership and child mortality from the Demographic and Health Surveys as measures of economic development, these data do not exist before the late 1980s, have limited geographic information before the late 1990s, are not representative at the local level, and exclude many medium-sized and small cities.

2. Empirical Methods

We study how increased market access to other cities affects city population growth in 2,789 urban cells in 39 sub-Saharan African countries sampled every ten years between 1960 and 2010. We now describe: (i) how we construct market access; (ii) our baseline specification; and (iii) our identification strategies.

2.1. Construction of Market Access to Other Cities

Following Donaldson and Hornbeck (2016), we define origin cell o's *market access* (*MA*) in year t, as $MA_{ot} = \sum_{d \neq o} P_{dt} \tau_{odt}^{-\theta}$, where P is urban population, d indexes destination cells, τ_{odt} is the time required to travel between cells o and d, and $\theta = 3.8$ is an estimate of trade elasticity measuring how trade volumes fall as trade costs increase, from Donaldson (forthcoming). In our analysis, we focus on how changes in the road network affect travel times between cells.

As explained above, our unit of analysis is a 0.1 by 0.1 degree grid square. Using these units simplifies computation compared to the full vector road network, and avoids problems due to missing topological information, concerning which segments connect to each other and which do not, in vector roads datasets. We assign to each grid square in each year a speed of travel for the fastest road segment type falling in the grid square in the year, or a baseline speed if no roads are present. We assume 80, 60, 40, 12, and 6 km/h on highways, paved roads, improved roads, earthen roads, and areas with no roads, respectively. The precise values are illustrative; results are insensitive to a scale factor.[19]

The time required to travel from each cell to all cells containing cities is calculated every ten years from 1960 to 2010 using Dijkstra's algorithm, the road speed assumptions above, and the great circle distances between neighboring cell centroids.[20] When a map is not available for a given year, we interpolate speeds between the closest map years before and after.[21]

[19]We show below that results hold if we use alternative trade elasticities or speeds.

[20]See Appendix Section A.3 for details. Appendix Figure A.4 shows how we obtain market access changes for Sierra Leone between 1970 and 1980. Appendix Figure A.5 shows the change in market access between 1960 and 2010 for the 187,900 cells of the 39 sample countries.

[21]For roads in 1960, we assign roads from the earliest available year (1961 for Central/South, 1965 for North/West, and 1966 for North/East). This assumes no road building between 1960

2.2. Baseline Specification

We are interested in how market access MA affects urban population P, so our baseline regression equation (for cell o in country c in year t) is:

$$\ln P_{ot} = \beta_0 \ln MA_{ot} + \lambda_o + \rho_{ct} + \phi_t \Omega_o + \epsilon_{0ot} \tag{1}$$

which includes cell fixed effects λ_o and country-year fixed effects ρ_{ct} to account for time-invariant city characteristics and flexible national trends, respectively, and a third-order polynomial in longitude and latitude Ω_o interacted with year fixed effects to control for unobservables correlated across space within country-decades. We consider several lags of market access change, suppressed from equations for clarity, to look for changing impacts over time, as we do not expect the effect of road changes on population to be instantaneous.

In first differences (at ten-year intervals, since we have urban data every ten years), cell fixed effects cancel and this becomes:

$$\Delta \ln P_{ot} = \beta_0 \Delta \ln MA_{ot} + \Delta\rho_{ct} + \Delta\phi_t \Omega_o + \Delta\epsilon_{0ot}. \tag{2}$$

We further control for initial log population in the first-difference specification to account for any divergence (convergence) if large cities grow faster (slower) than small cities, due to local increasing returns or mean reversion.

Suppressing fixed effects and controls, stacking across all o, and defining the matrix T_t with off-diagonal elements in row o and column d equal to $\tau_{odt}^{-\theta}$ (and diagonal elements equal to zero), (1) becomes:

$$\ln P_t = \beta_0 \ln T_t P_t + \epsilon_{0t} \tag{3}$$

a log-transformed spatial lag specification. Then, using (3), (2) becomes:

$$\begin{aligned} \Delta \ln P_t &= \beta_0 \Delta \ln T_t P_t + \Delta\epsilon_{0t} \\ &= \beta_0 (\ln T_t P_t - \ln T_{t-10} P_{t-10}) + \Delta\epsilon_{0t} \\ &= \beta_0 (\ln T_t P_t - \ln T_{t-10} P_{t-10} + \ln T_t P_{t-10} - \ln T_t P_{t-10}) + \Delta\epsilon_{0t} \\ &= \beta_0 (\ln T_t P_t - \ln T_t P_{t-10}) + \beta_0 (\ln T_t P_{t-10} - \ln T_{t-10} \ln P_{t-10}) + \Delta\epsilon_{0t} \end{aligned} \tag{4}$$

and the first map, which underestimates road building in the 1960s. We will explain later that results hold when dropping decades plausibly affected by 1960s road-building.

Changes in market access come from either changes in the population of other cities P (weighted by travel times T in t) or changes in travel times T to these other cities (weighted by the population of cities P in t-10). From (3) and (4), it is apparent that market access is mechanically endogenous, since city o's growth affects the growth of other cities d, which in turn affects city o's growth.

2.3. Identification Strategies

Our chief identification concerns are this reverse causality and omitted variables. Market access changes due to both changes in the population of city trading partners and changes in the roads connecting them. Unmeasured factors increasing a city's population could also increase its' neighbors' population, and therefore its market access. Furthermore, roads could be built in anticipation of city growth, or in anticipation of city stagnation in order to prevent it. Misspecified functional form and measurement error may also bias estimates.

Instrument fixing population. We can build an instrument for the change in market access $\Delta \ln MA_o$ that fixes population of the other cities P_d in t-10, and thus only relies on changes in travel times/roads T between t-10 and t (the second component in (4)), limiting the scope for reverse causality. This instrument is:

$$\Delta_R \ln MA_{ot} = \ln(\sum_{d \neq o} P_{d,\mathbf{t\text{-}10}} \tau_{o,d,\mathbf{t}}^{-\theta}) - \ln(\sum_{d \neq o} P_{d,\mathbf{t\text{-}10}} \tau_{o,d,\mathbf{t\text{-}10}}^{-\theta}). \quad (5)$$

Instrument also excluding local road changes. The problem with the previous instrument is that local road changes do not necessarily satisfy the exclusion restriction. Unobserved factors may drive both city o's growth/decline and surface improvement/deterioration of roads to neighboring cities d. One solution to this problem is to restrict attention to changes in *non-local* roads, i.e. road changes taking place sufficiently far away from city o that they are less likely to be driven by local factors that also drive city o's growth.

Defining "far away" as outside an exclusion radius $j \in (5,10,15)$ cells (roughly 55, 111, or 167 km) of city o, we define a class of instruments IVj:

$$\Delta_R^{out,j} \ln MA_{ot} = \ln\left(\sum_{\delta(d,o)\geq j} P_{d,\mathbf{t\text{-}10}} x_{od,\mathbf{t}}^{-\theta} + \sum_{\substack{d\neq o\\ \delta(d,o)<j}} P_{d,\mathbf{t\text{-}10}} x_{od,\mathbf{t\text{-}10}}^{-\theta}\right) - \ln\left(\sum_{d\neq o} P_{d,\mathbf{t\text{-}10}} x_{od,\mathbf{t\text{-}10}}^{-\theta}\right) \quad (6)$$

where δ is the Euclidean distance metric. They exploit the variation in the change in market access $\Delta \ln MA_{ot}$ coming from changes in roads more than j cells away from city o. Figure 4 shows a schematic version of this setup. City o's overall market access at time t is a function of the cost of traveling to cities d_1–d_4 and their population at time t. In calculating the change in market access from $t-10$ to t, the instrument uses population from $t-10$, as well as changes to non-local roads r_2, r_3, r_4, r_5 and r_8 between $t-10$ and t. Any changes to local roads r_1, r_6, or r_7 between $t-10$ and t are excluded from the instrument, because they could be endogenous to city o's growth.

$\Delta_R^{out,j} \ln MA_{ot}$ is a valid instrument as long as changes in non-local roads are excludable from equation (2). Excludability is threatened if there are factors that affect both city o's growth and the construction of these non-local roads. As the exclusion radius j increases from 5 to 15 cells, we exploit less local road changes, and are more likely to satisfy the exclusion restriction. However, faraway road changes are less likely to determine changes in market access, so instruments exploiting road changes far away are weaker. Given this trade-off between excludability and strength of the instruments, we report results for multiple radii.

Excluding selected non-local road changes. Construction of faraway radial roads could proxy for construction of near radial roads, which are due to city o's growth, with both being driven by policymakers wanting to connect city o to elsewhere. We call this phenomenon *co-investment*. For example, in Figure 4, the government may simultaneously upgrade roads r_1, r_2 and r_3 in order to better connect city o and city d_1. In that case, road changes outside the exclusion radius (r_2 and r_3) may not satisfy the exclusion restriction because they are correlated with road changes inside the exclusion radius (r_1).

Alternatively, construction of faraway radial roads could be due to city o's growth inducing demand for a connection between city o and faraway cities, but if roads near city o are already good, they may not be (measurably) improved, leaving measurable improvements to be found only far away. We call this phenomenon *radial extension*. In Figure 4, the government may decide to upgrade roads r_2 and r_3 in order to better connect city d_1 to city o. If r_1 cannot be upgraded further, this will not constitute co-investment, but road changes outside (r_2 and r_3) may not satisfy the exclusion restriction if they are correlated with nearby non-road investments also causing city o's growth.

In order to address these concerns, we harness the idea that this connection between near and far road construction is much more likely if they are both in the same direction from city o. We thus introduce a discrete local radial coordinate system for city o. A road can be built in either the *inner* or *outer* ring ($s \in [1,2]$) with respect to city o, in one of 8 octants ($q \in [1,8]$), subtended by the 8 cardinal and intermediate directions of the compass. Let the stock of (improved, paved and highway) roads in octant q in ring s with respect to city o in year t be R_{otqs}. In this framework, changes in $\sum_{q,s} R_{otqs}$ are what drive road-based changes in market access, and the instrument $\Delta_R^{out,j} \ln MA_{ot}$ above is entirely based on road changes in the outer rings ($s = 2$), $\Delta \sum_q R_{otq2}$.

Using this terminology, co-investment is equivalent to $corr(\Delta R_{otq2}, \Delta R_{otq1}) > 0$ driving $corr(\Delta R_{otq2}, \Delta P_{ot}) > 0$ due to an omitted variable inducing road building toward city o from elsewhere. In this case, road building in the outer ring is proxying for potentially endogenous road building in the inner ring. We address this by excluding city-periods with octants in which there is inner and outer radial road-building, or more formally, dropping city o in years t, $t+10$, and $t+20$ (i.e. all years in which road-building between $t-10$ and t appears on the right hand side, given two lags) if $\exists q : \Delta R_{otq1} > 0 \ \& \ \Delta R_{otq2} > 0$. In Figure 4, this means dropping city o in year t if *in any decade* between $t-30$ and t, r_1 and r_2 (or, e.g., r_6 and r_8), were both upgraded. We do not require the upgraded inner and outer radial roads to be contiguous. We limit consideration to roads that pass through designated bands (in gray in the figure) in the inner and outer rings of the same octant, to ignore non-radial roads such as r_9 and r_{10} in Figure 4.

Radial extension then implies $\Delta R_{otq1} = 0$ but only because quadrant q already has a good radial road in its inner ring ($R_{o,t-10,q1} > 0$). We address it by excluding city-periods where an outer road is built in the same quadrant where a paved or improved inner road already exists. Formally, we drop city o in years t to $t+20$ if $\exists q : R_{o,t-10,q1} > 0 \;\; \& \;\; \Delta R_{otq2} > 0$. In Figure 4, this means dropping city o in year t if in any decade between $t-30$ and t, r_2 (or r_8) was upgraded when r_1 (r_6) was already paved or improved.

As a variant of this, we exclude from consideration changes to roads deemed “transcontinental” in the Michelin maps from the first year available (circa 1960), as they are the most likely to be upgraded due to non-local factors, and therefore be endogenous to city o's growth even if they are far away from it.

Dropping potential growth hubs. The above strategies account for endogenous road building that is nearby, or in the same octant as nearby road building or good roads, or deemed transcontinental. As a complementary, more direct approach, we also drop selected cities with observable characteristics that may cause them to grow and cause roads to be built towards them, even from far away. Specifically, we drop city-decades with a set of known shocks, or local resources most likely to drive such shocks, that might affect city growth and road building: largest cities, mines, cash crop regions, head of state's hometown, ports, airports, customs posts, natural parks, colonial status, wars, refugee camps, droughts. Alternatively, we simultaneously control for many of these factors.

Excluding regional mean reversion. Note that in (6) the instruments are constructed using the population of the other cities d in $t-10$ as weights for the changes in travel times/roads. While we control for the initial population of city o in $t-10$, we cannot control for the initial population of the other cities in $t-10$. However, if city o's past growth (between $t-30$ and $t-10$) is correlated with the past population growth and thus population level of the other cities d, it could be that the weights are also endogenous. In that case, the instruments may not satisfy the exclusion restriction. One solution to this problem is to directly control for the two lags of city o's population growth, i.e. $\Delta \ln P_{o,t-10}$ and $\Delta \ln P_{o,t-20}$. Another solution is to use the initial population of the other cities d

in 1960, as opposed to $t-10$, as weights in the instruments.[22] Alternatively, we use population in 1960 (or t-10) to define not just the instruments but also the main change in market access variable: $MA_{ot} = \sum_{d \neq o} P_{d,1960} \tau_{odt}^{-\theta}$.

3. Results: Average Effects

3.1. OLS Results

Table 1 reports estimates of Equation (2), along with variants adding and removing lags and leads. In this and all subsequent tables, values of the dependent variable are divided by 100, so that coefficient can be interpreted as elasticities multiplied by 100: the percentage change in population associated with a doubling of market access. In Column 1, only the contemporaneous change in market access is included. It has a modest impact on city population, with an elasticity of 1.3%. Columns 2–4 add lagged changes in market access from previous decades. Changes in market access in the decade prior to the population change in question and in the decade prior to that each appear to have broadly similar but somewhat smaller effects. The overall effect of a 100% increase in market access in each decade, across these three decades, is thus over 3%. In column 4, the prior decade, 30 years before the measured population change, has a smaller effect that is imprecisely measured.[23]

In column 5, we investigate reverse causality by adding a lead to the column 3 specification; it is insignificant. The last row of coefficients in Table 1 reports the sum of the contemporaneous coefficients and all included lags. Once the second lag is included, the overall 30-year effect is quite stable, regardless of the presence of the lead, with an elasticity of about 3.5 to 4.5%. We thus include two lags for the rest of the paper, so the sample contains the three decades 1980–2010.[24]

3.2. Instrumental Variables (IV) Results

Table 2 reports the results of the IV specifications intended to disentangle the causal effect of market access due to roads on city growth. Column 1 repeats the

[22] The second lag of the change in market access for the period 1980–1990 already used 1960 in (6). We thus run this test on a sample dropping the 1980s as well.
[23] The 2-lag specification has lower Aikike and Bayesian information criteria than the 3-lag one.
[24] See Appendix Table A.3 for the descriptive statistics of the main sample (4,725 observations).

baseline result from Table 1. Columns 2–4 instrument with changes in market access due only to roads built far away, thus excluding changes due to road built nearby as well as recent city growth everywhere. Effects are larger than in the OLS specification, with 30-year elasticities between 8.8% and 17.7%, spread roughly evenly across the three decades and increasing with the radius. Alternatively, these results imply that a one standard deviation in market access growth is associated with a 0.46-0.88 standard deviation in city population growth.

As expected, the instrument is stronger at lower radii, because it includes road changes closer to the city. As shown in Appendix Table A.4, instruments based on wider radii (20 cells instead of 5–15 cells), and, alternatively, based on exclusion of roads within the same country, or within neighboring countries, give somewhat similar results but are weaker (see Appendix Section A.6 for details).

The fact that the IV estimates are larger than the OLS is consistent with the literature (Redding and Turner, 2015). While the initial identification concern in this literature was that more roads are built to cities expected to grow faster, in practice, roads appear to be more likely to have been built toward lagging cities. Alternatively, this downward bias may be the result of measurement error in the market access measure. Finally, the higher IV could reflect heterogeneity in the overall effect, and we explore this possibility further below. The cities most likely to be impacted by road changes far away (i.e. the instrument) are those for which nearby destinations matter little. This is likely to be true of relatively remote cities. The different IVs may thus reflect different local average treatment effects.

The magnitude of the effects we find is smaller than the 0.25 to 0.3 reported for total population in U.S. counties by Donaldson and Hornbeck (2016), the most similar specification to ours in the literature. There are several possible reasons for this. First, there are likely to be higher costs of trade and migration in this context, especially between countries and perhaps across ethnic territories, in part because of limited land markets. In that sense our context may be closer to China with its restrictive Hukou system. Second, there was much lower economic growth overall in our context. Donaldson and Hornbeck (2016) study the period 1870–1890, when the U.S. was experiencing its Second Industrial

Revolution and receiving massive inflows of immigrants. They also report estimated discrete effects of rail construction on agricultural land prices, so that it is a cross-walk to the rest of the literature. As noted by Redding and Turner (2015), these are substantially larger than the effects of roads and railroads on land prices and wages elsewhere in the literature, by a factor of two or more in some cases. This suggests that our results are broadly similar to other contexts.

3.3. Robustness checks

As discussed in Section 2.3., there are several reasons why faraway road changes may not satisfy the exclusion restriction. In Table 3, we investigate whether results hold if we account for: (i) co-investment; (ii) radial extension; (iii) growth hubs; and (iv) regional mean reversion. Rows are structured like Table 2 but only report overall 30-year effects. Row 1 shows the baseline results.

Excluding selected non-local road changes. In rows 2–4 cities with any *co-investment* (road-building in the same decade in the inner and outer rings of the same octant) are dropped. Rows 2, 3 and 4 define the inner ring between 2 and 3 cells from the city, and the outer ring 5–6, 10–11, and 15–16 cells from the city, respectively.[25] The sample is reduced by more than 50%, but results are generally consistent with the baseline. The instrument excluding up to a radius of 15 cells is weak. The row 2 sample drops the most cities, because the 2–3 cell region and the 5–6 cell region are so close to each other. In rows 5–7, cities with any *radial extension* are dropped (using the 2–3 cells for the inner ring and the 5–6, 10–11, and 15–16 cells for the outer rings). Sample sizes again fall by over 50%, but results remain similar. Row 8 exclude roads deemed transcontinental in early 1960s maps from the instrument (i.e. in constructing the instrument, they are assumed to remain with their $t-10$ speed in t), with little effect.[26]

Dropping potential growth hubs. Row 9 drops from the sample each country's 5 largest cities and national and regional capitals from 1960. This is akin to

[25]By construction, IV10 and IV15 (IV15) already exclude co-investment in an outer ring of 5–6 (10–11) cells, so these combinations are not reported.

[26]Results somewhat hold (see Appendix Section A.7 and Appendix Table A.5) if we consider the 1–2 cells for the inner ring, i.e. cells closer to the city, or quadrants instead of octants when the 2–3 cells are used for the inner ring (but we then lose about 2/3 of the sample).

the identification strategies of Michaels (2008) and Faber (2014), in that they do not rely on large cities, whose growth has driven the placement of road construction. Instead, they rely on small cities, which were more likely to be connected incidentally.[27] Rows 10–12 drop (i) cities within 100 km of a mine open at any time between 1960 and 2010 (10); (ii) cities within 100 km of a cell whose land suitability for cash crops is above 90% (11); and (iii) cities within 100 km of the hometown of any of the country's head of states between 1960 and 2010 (12). Results are similar or if anything slightly larger in magnitude. Results also hold when dropping cities within 100 km of: (i) a "top" city (capital, largest, and 2nd largest) in 1960 or 2010; (ii) a port in 1960 or 2005; (iii) an airport in 2007; (iv) a customs post in 2010; or (v) a natural park in 2015 (see Appendix Section A.7 and Appendix Table A.5). Likewise, results hold if we add many controls proxying for physical, economic and political geography (row 13).[28]

Results also hold if we drop country-decades in which the country: (i) was still a colony; (ii) experienced a war; (iii) received refugees; or (iv) suffered a multi-year drought (see Appendix Section A.7 and Appendix Table A.6).

Excluding regional mean reversion. Row 14 includes two lags of the dependent variable. In rows 15–18 population is fixed at its 1960 level in constructing the instruments, and in rows 17–18, in the instrumented market access (*MA*) as well. In rows 16 and 18, the 1980s are dropped because they are the only decade in which an included lag of ΔMA uses the population of the other cities in 1960. In row 19, population is fixed at its $t-10$ level in MA as well as its instrument. In each case, results differ little from the baseline. IV15 instruments are weaker in rows 15–16, and IV15 estimates are larger in rows 17–19.

[27]Appendix Table A.5 shows coefficients are lower and less significant when also dropping regional capitals in 2010. However, if roads promote city growth, and larger cities are more likely to become regional capitals, we may under-estimate the effects when dropping the new ones.

[28]The controls include dummies if the cell contains the capital / largest / second largest city or a regional capital in 1960 or 2010, and the log of the Euclidean distances to these cities, dummies if the cell is within 100 km from a top city in 1960 or 2010, a mine, a cash crop region, a president's hometown, a port in 1960 or 2005, an airport in 2007, a border crossing in 2010, or a natural park in 2015, and the log of the Euclidean distances to these locations, dummies if the cell is on the coast or crossed by a river, and the log of the Euclidean distances to the coast/a river, the mean and standard deviation of altitude (to control for ruggedness), and average rainfall in 1900–1960.

Other specification and sample checks. In Appendix Section A.8 and Appendix Table A.7, we show the effects are robust to changing specifications and samples. More precisely, results hold if we: (i) replace the country-decade fixed effects with decade fixed effects; (ii) cluster standard errors at the country level; (iii) replace the 3 main variables of interest covering the periods (t-30,t-20), (t-20,t-10), and (t-10,t) with one for (t-30,t); (iv) use alternative speeds; (v) allow railroad travel in calculating market access; (vi) use alternative trade elasticities; (vii) add uniform costs of crossing borders; (viii) use iceberg costs; (ix) exclude countries bordering South Africa, North Africa or the Arabian Peninsula, as their market access may be underestimated; (x) drop the 1960s, for which the road data is incomplete, or the 2000s, for which both the road and city data may be incomplete; (xi) use additional population estimates for cell-years under 10,000 to increase the sample's balance; and (xii) restrict the sample to country-decades with population estimates that are the most likely to be reliable.

3.4. Effects on Night Lights/Income

We expect better market access in a city to increase population in the context of a wide class of models allowing for spatial equilibrium. Our results, implying that populations may take up to 30 years to reallocate, suggest that the resulting migration is costly. In the interim away from equilibrium, the increase in market access could produce an increase in welfare, via lower prices and increased productivity and wages. Unfortunately, in this data-poor context, we do not have panel data on wages, prices, or amenities at the city-level.

To explore this idea, we consider changes in night lights as a proxy for overall output.[29] The sample includes 3,591 observations, for the periods 1992–2000 and 2000–2010 only, because lights data begin in 1992. Appendix Section A.9 and Appendix Table A.8 show that the 30-year overall effects on population for this restricted sample are similar to those for the full samples.

Table 4 reports results for night lights. While OLS estimates of the market access coefficients are not large and much less precisely specified than in the

[29] Henderson et al. (2011) show that in a worldwide sample of countries, as well as a subsample of developing countries, changes in night lights are correlated with changes in GDP.

population regressions of Table 2, in each IV specification, market access has substantially larger effects on lights than on population. Furthermore, the effects are entirely in the first decade. Consistent with this, IV estimates of the effect of market access on lights per person in Appendix Table A.8 are substantially positive and restricted to the decade of road construction.

These results suggest that roads increase economic activity relatively quickly (in the first decade), while the population effects take much longer to evolve (over three decades). This is consistent with migration flows that take substantial time to develop. While we cannot say exactly why economic activity increases, the results are consistent with a transition to spatial equilibrium.

3.5. Net Creation vs. Reorganization of Economic Activity

Results thus far have not distinguished between different sources of population growth from the perspective of an individual city. Increased market access could induce a city to grow by attracting rural residents (what we call *induced urbanization*), by attracting urban residents from other cities (*urban reallocation*), or by increasing its differential of births over deaths (*urban natural increase*). 18% of sub-Saharan African population was urban (based on localities above 10,000) in 1980 at the beginning of the regression sample, and this number has increased only to 28% by 2010, so the pool of rural potential migrants was always 3-4 times as large as the pool of urban potential migrants.[30]

Table 5 provides some further evidence distinguishing between the first two possibilities. In rows 2–4, we restrict to country-decades with successively smaller urban shares in year $t-30$. These countries with low urbanization rates have the most limited sources of potential urban-urban migrants, and are therefore least likely to see reallocation across cities. In row 2, restricting to country-years under the median urbanization rate (≈18%) has very little effect on results. In row 3, restricting to the bottom quartile (≈10%) reduces magnitudes somewhat more, though in the case of IV15, this may be driven by instrument weakness in a small sample. Furthermore, using only the low-urbanization

[30]This differs from the context of middle- or high-income countries like China (urban share ≈ 55% today) and the U.S. (80%). Urban reallocation is mechanically more likely there.

decile (≈7%) of countries in row 4, results are more similar to the full sample (though again the IV15 and now IV10 instruments are quite weak).

Row 5–8 offer a more direct test of local reallocation. Each row repeats the baseline regression on successively larger units of analysis, created by aggregating individual cells into mutually exclusive square blocks, or *mega-cells*. In row 5, each unit is a 3x3 square of the original units. Because some such 3x3 squares contain multiple cities, the sample size shrinks. By row 8, the average 9x9 square contains approximately two cities. If all urban growth induced by roads was pure reallocation within such 9x9 grid squares, we would expect no effect on this sample. Effects do on average become smaller and noisier, with weaker instruments, as is expected given the smaller sample size. However, they are broadly of the same magnitude as baseline results, suggesting that the majority of the effect is not due to local reallocation. We cannot distinguish reallocation between cities across larger distances using this method, as aggregation to larger squares produces small sample sizes and weak instruments.

From the perspective of central place theory (Christaller, 1933), this kind of long-distance migration is especially likely to the largest cities. Rows 9–12 repeat the tests of rows 5–8, restricting the sample to mega-cells that do not contain the capital or any of the 5 five largest cities or regional capitals of each country in 1960. This restricts the test to mega-cells that are unlikely to be destinations of long distance migration, especially if there are ethnic differences across mega-cells. Results are noisy but similar. They do not rule out reallocation, but they are broadly inconsistent with the story that our results are driven mostly by urban residents migrating up the urban hierarchy to the largest cities.

While no direct evidence can help us to distinguish between rural-urban migration and urban natural increase, theory tells us that if anything, natural increase should operate in the opposite direction. If market access increases labor demand and therefore wages, this should decrease both fertility and mortality Galor (2012). However, variation in urban rates of natural increase across African countries in the period under study was driven primarily by variation in birth rates, whereas urban mortality was much lower and much more uniform across both countries and cities within countries (Jedwab et al., 2017b; Jedwab and Vollrath, 2017). If this in turn means that mortality is unlikely to change with market access, then the fertility channel would dominate, and if anything, increased market access should be more likely to decrease urban population growth. However, without existing panel city-level data on fertility and mortality, we cannot formally test this hypothesis.[31]

Overall, this limited evidence suggests that urban growth was primarily driven by induced urbanization. However, without panel data on historical local demographic patterns, these results must be taken with great caution. In the analysis of the aggregate urban effects of roads that were upgraded in 1960–2010, we will thus consider different urban reallocation scenarii.

4. Results: Heterogeneous Effects

Transport investments may have different effects depending on the local context in which they take place. Table 6 explores heterogeneity of results with respect to several factors highlighted in recent literature on economic geography, structural change, and political economy. As in Tables 3 and 5, each row shows 30-year estimates of a variant of equation (2), in which we control for the dummy variable shown at left and interact it with the contemporaneous and lagged changes in market access, and the analogous instruments. For the IV5 estimation strategies the table reports the 30-year coefficient for the dummy=0 group, the dummy=1 group, and the difference; for IV10 and IV15, for which instruments are generally weaker, only the difference is reported. At left, each row also reports first stage Kleibergen-Paap F-statistics and the share ("Sh") of the dummy=1 group. F-statistics suggest that instruments are somewhat weaker here than above.

These exercises are very demanding on the data, with six endogenous variables and six instruments per regression. All in all, differences shown are illustrative of broadly consistent general patterns but not all are robustly significantly different from zero across the four specifications.[32]

[31] See footnote 18 on the limits of the Demographic and Health Surveys for our purposes.
[32] Appendix Section A.10 and Appendix Tables A.12-A.13 show that most of the results described

Economic Geography. Rows 1–3 of Table 6 show variation with respect to three economic geography characteristics. Core-periphery models predict that reduced trade costs increase the size of big cities more than smaller cities. However, row 1 shows that cities initially (in $t-30$) smaller than their country's median city generally see larger effects. If anything, reduced trade costs lead to a decentralization of urban population in our context.[33]

Rows 2 and 3 consider dummy variables proxying for economic remoteness as of 1960: below median market access in the country, and above median Euclidean distance from the "top" (capital, largest or second largest) cities in each country. Cities with worse market access see stronger effects of a marginal improvement. This is consistent with decreasing marginal returns to transportation investments, and suggests that remoteness raises their returns.[34]

Physical Geography. Sub-Saharan Africa has a large agricultural workforce, and much urbanization reflects workers moving out of agriculture. Cities in regions with differing levels of agricultural suitability may thus be more or less able to take advantage of better transport to diversify into secondary and tertiary sectors. Rows 4 and 5 show variation with respect to a measure of agricultural land suitability within one grid cell of the city, cutting the sample at 75% and 25% percent suitability.[35] In both cases, cities with worse land are more positively affected by increases in market access. In row 4 cities in areas where land suitability is under 25% grow relatively faster when they are better connected to other cities. Conversely, cities in areas where land suitability is over 75%

---

below are similar if we use market access changes and instruments based on 30-year periods. This ignores information on timing, but also reduces the number of instruments and endogenous variables to two, so the instrument set is stronger. The only result that changes substantially is for smaller cities (see below), possibly because we lose too much relevant information on them.

[33]Differentials based on the country's 25th or 75th percentile population, dropping the top cities (capital, largest and second largest) in 1960 and 2010, or using the continental population median are all in the same direction (see Appendix Section A.10 and Appendix Table A.9).

[34]Appendix Section A.10 and Appendix Table A.9 show that differentials based on each of the following are in the same direction: (i) the country's 25th or 75th percentile market access or using the continental market access median; (ii) access to paved/improved roads (in 1960), railroads (1960), ports (1960 or 2005), or airports (2007); and (iii) distances to the country's top cities in both 1960 and 2010, the continent's top cities in 1960, or dropping the top cities in 1960.

[35]GAEZ defines crop-specific land suitability based on soils, terrain and climate. Overall land suitability here is the maximum suitability across all potential crops (see Appendix Section A.4).

grow relatively slower when market access increases. The significance of the differences (for IV5 and IV15) are striking given that the high suitability group represents only 5% of the sample and its coefficients are imprecisely estimated as a result. This is consistent with cities in less agricultural areas specializing in more transport intensive activities that benefit more from the roads.[36]

Political Geography. Rows 6 and 7 of Table 6 show variation with respect to two political geography characteristics. Row 6 allows for a differential effect for city-decades of road-building that may have been favored because they were within 150 km of the place of origin of a head of state in power for at least two years in the decade (the mean decade-specific tenure). We use 150 km, because this represents a 3–4 hour driving time from the hometown given a driving speed of 40-60 kph (what we assume for improved/paved roads).[37] The differential is negative, suggesting that changes in market access have smaller effects when roads are built towards the cities surrounding the place of origin of a head of state (the p-value for the coefficient of the difference for IV10 is 0.103). In Appendix Section A.10, we report more results that overall suggest stronger negative effects for leaders with a longer tenure and whose regime is not democratic.[38]

This is surprising given that such areas were likely to also get complementary public investments and subsidies, which should increase the returns to transportation investments. The uninteracted effect of the leader favoritism dummy has a positive and significant coefficient between 0.05 and 0.07 (not shown), implying that cities around the leader's place of origin grows faster than

[36]We find generally similar effects as for land suitability when we study the interaction effects with rainfall (see Appendix Section A.10 and Appendix Table A.10). We also do not find any differential for cities closer to mines, a sector that may or may not be labor-intensive.

[37]A circle of radius 150 km has approximately the same area as 584 (11x11 km) cells. In the maps of ethnic boundaries based on Murdock (1959) and Weidmann et al. (2010), ethnic groups occupy on average 249 and 926 cells respectively, with the mean of the two equal to 587.

[38]Appendix Section A.10 and Appendix Tables A.10 and A.12 show that the differential is consistently negative (with varying precision) when: (i) using the 90th percentile tenure (9 years); (ii) using the mean and 90th percentile tenures for non-democratic leaders only; (iii) using the mean and 90th percentile tenures for a distance of 250 km from the hometown, i.e. a 4-6 hour drive from it; (iv) using the mean and 90th percentile tenures and the map of ethnic boundaries based on Murdock (1959) to identify politically connected areas; and (v) using the mean and 90th percentile tenures (5 and 20 years, respectively) for a single 30-year period instead.

other cities in the country controlling for market access. It is however consistent with the idea that such roads were politically but not economically optimal.

Conversely, and unlike large cities in general in Row 1, regional capitals see if anything larger effects of increased market access on their growth, consistent with, for example, complementarity between government services and transport-sensitive activities. The differential is only significantly different from zero when considering 2010 regional capitals (row 8), whose status could have been jointly determined along with road locations, not 1960 regional capitals (rows 7), but the sign is consistent. Overall, this suggests that roads built for different kinds of "political" reasons may have different effects.

Foreign, domestic, overland and overseas. The effect of market access may also depend on what markets are being accessed. Measures of market access shown so far assume that crossing a border is costless, but that crossing an ocean is infinitely costly. Results in Appendix Table A.7 show that adding substantial uniform border costs has little effect on results. In Table 7, we decompose market access, first into access to *domestic* cities versus *foreign* cities within sub-Saharan Africa, and then into access *overland* to the rest of sub-Saharan Africa versus access to *overseas* markets, proxied by access to cities with a port. For market access to foreign cities, we construct an instrument restricting attention to roads built outside the country rather than outside a radius (IV-Foreign); all other terms are instrumented as above. There are six endogenous variables (two market accesses × three lags) and six instruments, so instruments are weaker.[39]

Row 1 of Table 7 reports the effects of domestic vs. foreign market access. The six instruments always include IV-Foreign and its two lags, while the remaining three differ by column as shown. Access to domestic markets consistently increases the size of cities. The impact of access to foreign cities is both smaller and less precisely measured.[40] A one standard deviation change in domestic (foreign) market access is associated with a 0.37–0.61 (0.06–0.10) standard

[39] Appendix Section A.10 and Table A.13 show that when averaging information into 30-year changes, instruments are strong, and the relative roles of foreign and overseas MA decrease.

[40] Other combinations of instruments are generally weaker. Domestic and foreign market access changes are correlated, but only at 0.27.

deviation change in city population growth.[41]

Row 2 investigates the effects of overland vs. overseas market access. We treat market access to 44 major ports in 2005 as a proxy for overseas market access because ports are the primary conduits of international trade. Overseas market access thus capture road changes and population changes for the cities with a 2005 port. While we do not have comprehensive historical measures of port traffic, the port cities' 2010 populations are highly correlated (at 0.68) with their port traffic volume (in 20-foot equivalent units in 2005). We thus believe that the population of a port city is a good proxy for port traffic.[42] Overland market access captures access to cities without a 2005 port. The two measures are correlated at 0.31. The six instruments always include IV5 and its two lags for overland cities, while the remaining three for overseas cities differ by column as shown. [43]

Rising access to overland markets consistently increases the size of cities. Unlike what we found for foreign access (to other sub-Saharan African countries), overseas access (to non-African countries) has positive and significant coefficient estimates, but only when the exclusion radius is 10-15 cells. The coefficient is then higher than for overland access, but only because the variance of overland market access is larger than that of overseas market access in the sample. A one standard deviation increase in overland and overseas market access change, are respectively associated with 0.28–0.43 and 0.01–0.23 standard deviation increases in city population growth.[44]

[41]Connections to wealthier cities/countries may be more important than connections to poorer cities/countries. However, Appendix Section A.10 and Appendix Table A.11 show that results do not change in an alternative specification with a measure of market access that weights the population of each destination city by its country's contemporaneous per capita GDP.

[42]We use a list of 44 major ports from 2005 rather than a 1960 list because several small colonial ports declined after independence, and new ports emerged and grew fast before 2005. We thus believe that the 2005 list better represents the overall location of ports during the 1960-2010 period. For 36 ports with the relevant data, in 1960 log population was correlated with log exports and imports at 0.63 and 0.74, respectively. As noted below, results are similar using 1960 ports.

[43]Other combinations of instruments are weaker. We control for log distance to the coast interacted with country-year fixed effects, as we do not want overseas access to capture trends specific to coastal areas. In many countries, coastal and hinterland areas have distinct geographies and histories and have experienced different evolutions after 1960 (Austin, 2007).

[44]Appendix Section A.10 and Appendix Table A.11 show effects are robust to: (i) dropping the ports themselves; (ii) including the ports in calculating overland market access and its

Summary. To summarize, we find suggestive evidence that the effect of market access is stronger for: (i) small cities; (ii) remote cities; (iii) cities whose hinterlands do not have a comparative advantage in agriculture; (iv) cities less likely to be politically favored, unless it is for administrative reasons. Market access to domestic cities matters more than access to foreign cities, but international ports do matter. Although these results vary somewhat across specifications, they provide suggestive evidence that transportation investments may be heterogeneous depending on the context in which they are placed.

5. Aggregate Effects

We quantify aggregate effects from two perspectives: in terms of new urban residents induced to move to the city during the sample period due to roads built, and in terms of new predicted urban residents due to the proposed Trans-African Highway (TAH) network. As noted below, each requires strong assumptions.

Road Upgrades 1960–2010. What do our results say about the overall effect of road building on urbanization in the 39 sample countries between 1960 and 2010? For $t = \{1970, 1980, 1990, 2000, 2010\}$, we define:

$$\widehat{\Delta \ln P_{ot}} = \widehat{\beta_{lag0}} \Delta_R \ln MA_{ot} + \widehat{\beta_{lag1}} \Delta_R \ln MA_{ot-10} + \widehat{\beta_{lag2}} \Delta_R \ln MA_{ot-20}$$

where $\Delta_R \ln MA_{ot}$, as defined above, includes only changes in roads, not population, between $t-10$ and t, and the $\hat{\beta}$ terms are lag-specific estimated effects of changes in market access due to roads only. $\Delta_R \ln MA_{ot-10}$ and $\Delta_R \ln MA_{ot-20}$ are assumed to be zero for $t = 1970$ and $t = \{1970, 1980\}$, respectively, in the absence of data about road-building in the 1940s and 1950s.[45] Then counterfactual log population in year t (i.e. in the absence of changes between $t-10$ and t caused by roads built between $t-30$ and t) can be defined

instrument; (iii) fixing the populations of the cities with a 2005 port to their levels in 1960 when calculating overseas market access and its instrument; and (iv) using 1960 instead of 2005 ports.

[45]Alternatively, we ignore the effects of $\Delta_R \ln MA_{ot-10}$ for $t = 2020$ and the effect of $\Delta_R \ln MA_{ot-20}$ for $t = \{2020, 2030\}$, in the absence of data about city growth after 2010.

as:

$$\ln \tilde{P}_{ot} = \ln P_{ot} - \widehat{\Delta \ln P_{ot}}$$

Thus, absolute population change due to those roads is: $\tilde{\Delta} P_{ot} = P_{ot} - \tilde{P}_{ot}$. Summing across all cities and decades, our estimated contribution of road building to city growth 1960–2010:

$$\sum_t \sum_o \tilde{\Delta} P_{ot} = 5.6 \text{ to } 11.6 \text{ million}$$

depending on whether we use the average IV5 or IV15 estimates (see row 19 of Table 3 where we fix population to $t-10$ in the instrumented market access variables, in order to capture only changes in market access due to roads). Between 1960 and 2010, the total urban population of the 39 countries increased by 203.5 million, of which 171 million reflected the intensive margin growth we study. The 5.6–11.6 million new urban residents thus represent 3–7% of intensive margin growth. If these "extra" urban residents had stayed in rural areas, the 39 countries' overall urbanization rate would be 0.7–1.5 percentage points lower than its actual rate of 23.8% in 2010. In other words, our estimates attribute 5–10% of the intensive margin increase in the urban share to these road upgrades.

Now, allowing for heterogeneity in the simple binary way that we have increases the range of effects to 3.3–15.5 million new urban residents, accounting for 2–9% of the intensive margin growth in urban population or 0.4–0.19 percentage points of the urbanization rate and 3–13% of the intensive margin increase in the urban share.[46]

These estimates are conservative in the narrow sense that they apply changes to individual decades, rather than compounding them, and because they do not include the contribution of roads to extensive margin urban growth. However, they also assume no reallocation, which would reduce the estimated aggregate

[46]These estimates exclude the heterogeneity regressions for which the first stage F-statistic is below 5. More details on the methodology can be found in Appendix Section A.11. The heterogeneous effects when we fix population to $t-10$ in the instrumented market access variables can be found in Appendix Tables A.14 and A.15. The results on urban growth and the urban share can be found in Appendix Table A.16.

effects, or other general equilibrium effects, which could increase or decrease them. While we found little evidence of reallocation in Section 3.5. above, we certainly cannot rule it out.

Trans-African Highways 2010–2040. Another way to interpret these results is in the context of proposed roads. The idea of a Trans-African Highway (TAH) network has been discussed since at least the early 1970s and was operationalized in a proposal 30 years later by the United Nations Economic Commission for Africa and the African Development Bank (ADB and UNECA, 2003). Using a map of the TAH network we constructed from this document (Appendix Figure A.7b), we find that its complete implementation would require construction of 44,000 km of highways in sub-Saharan Africa, 42,000 km of which are in the 39 sample countries. In our data, there are only 1,490 km of highways in the 39 countries in 2010 . By comparison, India had 24,000 km (Government of India, 2016) and China 111,900 km (Government of China, 2016). The TAH network would thus represent a 2,740% increase in highway length.[47]

Assuming travel speeds of 80 (or alternatively 100) kph along the TAH roads, we estimate by how much the market access of each city in 2010 would have increased (due to roads only) had the TAH roads been built by 2010.[48] We then use the same methodology as described above to estimate the potential aggregate effects of the TAH network between 2010 and 2040.[49] Our average effects imply that the urban population of the 39 sample countries would increase by 2.7–11.8 million, similar in magnitude to the 1960–2010 effect estimates. However, the urban population of the 39 countries was 223.3 million in 2010, so this is a 1–5% increase, a small fraction of the 206% overall increase

[47]Using construction cost data from Collier et al. (2015), we estimate that the cost of building the TAH network is 19–24% of 2010 regional GDP (vs. 61.5% of 2000 regional GDP for 1960-2010 road upgrades).

[48]80 kph is the baseline highway speed in the rest of our analysis. 100 kph reflects the possibility that the TAH will be built to a higher standard than existing highways.

[49]We use the 30-year change in market access specification for this exercise because city populations in the intermediate years 2020 and 2030 are unknown. The average effects when using 30-year changes in market access, and fixing population to t-30 in the market access variable(s), and not just the instrument(s), are reported in row 24 of Appendix Table A.7. The heterogeneous effects using the same specification are shown in Appendix Tables A.17-A.18

predicted by United Nations (2015).[50] Our estimates thus imply that road-induced increased market access could raise the urbanization rate by 0.2-0.7 percentage points.[51] Allowing for heterogeneity expands these ranges to 1.5–13.4 million new urban residents, representing a 0.7–6.0% increase in the urban population; and a 0.1–0.8 percentage point higher 2040 urban share.

Again, these results do not account for urban reallocation, which could be more important given a higher initial urbanization rate, other general equilibrium effects, and extensive margin growth.

6. Conclusion

We find that increased market access due to road construction in Africa since 1960 has accelerated city growth, not only at the time of construction but in the subsequent two decades as well. We report suggestive evidence that effects differ by context. They are larger for smaller and more isolated cities, and market access changes to domestic rather than foreign cities, and weaker in politically favored and more agriculturally suitable areas. Under the scenario of no reallocation across cities, for which we provide some evidence, these effects represent a substantial fraction of urbanization in the 1960–2010 period, though it points to a more limited role for proposed highways in future urbanization driven more by other factors. Several mechanisms could be driving the results. Most theoretical and empirical work has focused on reductions in the cost of transporting goods. However, other work show that reduced intercity transport costs encourage the flow of information and labor. Future work will be needed to disentangle these channels. Another question that we are leaving for future research is what an optimal road network would look like, given the region's heterogeneity in physical, economic, and political geography.

[50] United Nations (2015) base their estimates on national urban definitions, which are on average less restrictive than ours. In addition, their estimates include both intensive margin growth and extensive margin growth.

[51] More details on the methodology can be found in Appendix Section A.12. The results on urban growth and the urban share can then be found in Appendix Tables A.19-A.20.

References

Acemoglu, Daron, Simon Johnson, and James Robinson, "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review*, 2005, *95* (3), 546–579.

ADB and UNECA, "Review of the Implementation Status of the Trans African Highways and the Missing Links," Technical Report, African Development Bank and United Nations Economic Commission For Africa 2003.

African Development Bank, *Annual Report*, Abidjan: African Development Bank, 2012–2015.

Ahlfeldt, Gabriel, Steve Redding, Daniel Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Alder, Simon, "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," mimeo, University of North Carolina March 2015.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

__ and __ , "The Welfare Effects of Transportation Infrastructure Improvements," mimeo, Dartmouth University 2016.

Asher, Samuel and Paul Novosad, "Market Access and Structural Transformation: Evidence from Rural Roads in India," mimeo, Dartmouth 2016.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Working Paper, MIT July 2015.

Austin, Gareth, "Labour And Land In Ghana, 1874-1939: A Shifting Ratio And An Institutional Revolution," *Australian Economic History Review*, 03 2007, *47* (1), 95–120.

Bairoch, Paul, *Cities and Economic Development: From the Dawn of History to the Present*, Chicago: The University of Chicago Press, 1988.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Paper 17897 2012.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, May 2007, *122* (2), 775–805.

__ , Loren Brandt, Vernon Henderson, Matthew Turner, and Qinghua Zhang, "Roads, Railroads and Decentralization of Chinese Cities," *Review of Economics & Statistics*, 2017, *99* (3), 435–448.

__ , Vernon Henderson, Matthew Turner, Qinghua Zhang, and Loren Brandt, "Highways, Market Access, and Urban Growth in China," mimeo, Brown University 2017.

Behrens, Kristian, W. Mark Brown, and Theophile Bougna, "The world is not yet flat : transport costs matter!," Policy Research Working Paper Series 7862, World Bank October 2016.

Berg, Claudia N., Uwe Deichmann, Yishen Liu, and Harris Selod, "Transport Policies and Development," Policy Research Working Paper 7366, World Bank 2015.

Blimpo, Moussa, Robin Harding, and Leonard Wantchekon, "Public Investment in Rural Infrastructure: Some Political Economy Considerations," *Journal of African Economies*, 2013, *22*, ii57–ii83.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, September 2014, *82* (5), 1671–1748.

Burgess, Robin, Rémi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel, "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, June 2015, *105* (6), 1817–51.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," Unpublished 2013.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, July 2000, *30* (4), 457–490.

Christaller, Walter, *Die zentralen Orte in Suddeutschland* 1933.

Coşar, A. Kerem and Banu Demir, "Domestic Road Infrastructure and International Trade: Evidence from Turkey," *Journal of Development Economics*, 2016, *118*, 232–244.

Collier, Paul, Martina Kirchberger, and Måns Söderbom, "The Cost of Road Infrastructure in Low and Middle Income Countries," *The World Bank Economic Review*, 2015, *30* (3), 522–548.

Coquery-Vidrovitch, Catherine, *The History of African Cities South of the Sahara: From the Origins to Colonization*, Princeton: Markus Wiener, 2005.

DeLong, J Bradford and Andrei Shleifer, "Princes and Merchants: European City Growth before the Industrial Revolution," *Journal of Law and Economics*, 1993, *36* (2), 671–702.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, forthcoming.

__ and Richard Hornbeck, "Railroads and American Economic Growth: A 'Market Access' Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Duranton, Gilles and Matthew A. Turner, "Urban growth and transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440.

__ , Peter M. Morrow, and Matthew A. Turner, "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 2014, *81* (2), 681–724.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Fajgelbaum, Pablo and Stephen J. Redding, "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914," Technical Report 2014.

Fay, Marianne and Charlotte Opal, "Urbanization without growth: a not-so-uncommon phenomenon," Policy Research Working Paper 2412, World Bank 2000.

Feyrer, James, "Distance, Trade, and Income: The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," NBER Working Paper 15557 December 2009.

Galor, Oded, "The Demographic Transition: Causes and Consequences," *Cliometrica, Journal of Historical Economics and Econometric History*, January 2012, *6* (1), 1–28.

Gertler, Paul J., Marco Gonzalez-Navarro, Tadeja Gracner, and Alex Rothenberg, "Road Quality and Local Economic Activity: Evidence from Indonesia's Highways," mimeo, Berkeley 2015.

Ghani, Ejaz, 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," *Economic Journal*, 03 2016, *126* (591), 317–357.

Gollin, Doug, Rémi Jedwab, and Dietrich Vollrath, "Urbanization with and without Industrialization," *Journal of Economic Growth*, March 2016, *21* (1), 35–70.

Government of China, *Transportation Overview*, Beijing, China: Ministry of Transport of the People's Republic, 2016.

Government of India, *Basic Road Statistics of India 2013-14 and 2014-15*, Delhi, India: Ministry of Road Transport & Highways, 2016.

Gwilliam, Ken, *Africa's Transport Infrastructure: Mainstreaming Maintenance and Management*, Washington D.C.: World Bank, 2011.

Hanson, Gordon, "Regional adjustment to trade liberalization," *Regional Science and Urban Economics*, July 1998, *28* (4), 419–444.

Henderson, J. Vernon, Mark Roberts, and Adam Storeygard, "Is urbanization in Sub-Saharan Africa different?," Policy Research Working Paper 6481, World Bank 2013.

Henderson, Vernon, Adam Storeygard, and David N. Weil, "A Bright Idea for Measuring Economic Growth," *American Economic Review*, May 2011, *101* (3), 194–99.

__ , __ , and Uwe Deichmann, "Has climate change driven urbanization in Africa?," *Journal of Development Economics*, January 2017, *124*, 60–82.

Jedwab, Rémi and Adam Storeygard, "Economic and Political Factors in Infrastructure Investment: Evidence from Railroads and Roads in Africa 1960–2015," mimeo 2017.

__ and Alexander Moradi, "The Permanent Economic Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics & Statistics*, 2016, *98* (2), 268–284.

__ and Dietrich Vollrath, "The Urban Mortality Transition and Poor Country Urbanization," mimeo, George Washington University 2017.

__ , Edward Kerby, and Alexander Moradi, "History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya," *The Economic Journal*, 2017.

__ , Luc Christiaensen, and Marina Gindelsky, "Demography, Urbanization and Development: Rural Push, Urban Pull and... Urban Push?," *Journal of Urban Economics*, 2017, *98*, 6–16.

Knight, Brian, "Parochial interests and the centralized provision of local public goods: evidence from congressional voting on transportation projects," *Journal of Public Economics*, March 2004, *88* (3–4), 845–866.

Konadu-Agyemang, Kwadwo and Kwamina Panford, *Africa's Development in the Twenty-First*

Century: Pertinent Socioeconomic and Development Issues, Farnham,UK: Ashgate, 2006.

Limão, Nuno and Anthony Venables, “Infrastructure, geographical disadvantage, transport costs, and trade,” *World Bank Economic Review*, 2001, *15* (3), 451–479.

Michaels, Guy, “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System,” *Review of Economics and Statistics*, November 2008, *90* (4), 683–701.

Morten, Melanie and Jaqueline Oliveira, “The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City,” mimeo, Stanford University 2017.

Murdock, George, *Africa: Its Peoples and their Culture.*, New York: McGraw-Hill, 1959.

Nelson, Andrew and Uwe Deichmann, “African Population database documentation,” Technical Report, United Nations Environment Programme and CIESIN 2004.

O’Connor, Anthony, *The Geography of Tropical African Development: 2nd Edition*, Headington Hill Hall, Oxford: Pergamon Press Ltd., 1978.

Pedersen, Poul Ove, “The Freight Transport and Logistical System of Ghana,” Working Paper 01.2, Centre for Development Research in Copenhagen 2001.

Rajan, Raghuram, Paolo Volpin, and Luigi Zingales, “The Eclipse of the U.S. Tire Industry,” in Steven Kaplan, ed., *Mergers and Productivity*, University of Chicago Press, 2000, pp. 51–92.

Redding, Stephen and Daniel Sturm, “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 2008, *98* (5), 1766–97.

Redding, Steven and Matthew Turner, “Transportation Costs and the Spatial Organization of Economic Activity,” in Gilles Duranton, Vernon Henderson, and William Strange, eds., *Handbook of Urban and Regional Economics*, Vol. 5, Elsevier, 2015.

Rothenberg, Alexander D., “Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia’s Highways,” mimeo, RAND Corporation 2013.

Stanig, Piero and Leonard Wantchekon, “The Curse of good Soil? Land Fertility, Roads and Rural Poverty in Africa,” mimeo, African School of Economics 2015.

Storeygard, Adam, “Farther on Down the Road: Transport Costs, Trade and Urban Growth in sub-Saharan Africa,” *Review of Economic Studies*, 2016, *83*, 1263–1295.

Tanzi, Vito and Hamid Davoodi, *Roads to Nowhere: How Corruption in Public investment Hurts Growth*, Washington: International Monetary Fund, 1998.

United Nations, *World Urbanization Prospects: The 2014 Revision* 2015.

Wasike, Wilson S.K., “Road Infrastructure Policies in Kenya: Historical Trends and Current Challenges,” Working Paper 1, Kenya Institute for Public Policy Research and Analysis 2001.

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

World Bank, *Road Deterioration in Developing Countries: Causes and Remedies.*, Washington, DC: World Bank, 1988.

— , *Africa’s infrastructure: a time for transformation*, Washington, DC: World Bank, 2010.

— , “World Bank Approves US$228 Million for Abidjan-Lagos Transport Corridor,” Press Release March 2010. http://www.worldbank.org/en/news/press-release/2010/03/24/world-bank-approves-us228-million-abidjan-lagos-transport-corridor.

— , *World Development Indicators*, Washington: World Bank, 2015.

— , *Annual Report*, Washington: World Bank, 2016.

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

Table 1: Average Effect of Market Access on Urban Population: OLS

Dependent Variable:	(Δ_{t-10}^{t} ln Urban Population)/100				
	(1)	(2)	(3)	(4)	(5)
Δ_{t-10}^{t} ln Market Access	1.34***	1.27***	1.58***	1.63***	1.50***
	[0.32]	[0.32]	[0.35]	[0.44]	[0.38]
Δ_{t-20}^{t-10} ln Market Access		1.02***	1.23***	1.55***	1.11***
		[0.24]	[0.26]	[0.34]	[0.30]
Δ_{t-30}^{t-20} ln Market Access			0.81***	0.89***	0.79***
			[0.23]	[0.29]	[0.27]
Δ_{t-40}^{t-30} ln Market Access				0.27	
				[0.23]	
Δ_{t}^{t+10} ln Market Access					0.67
					[0.49]
Overall Effect	1.34***	2.29***	3.62***	4.33***	3.40***
($t-40$ to t)	[0.32]	[0.45]	[0.59]	[0.83]	[0.65]
Observations	5,906	5,472	4,725	3,630	2,607
Adj. R-squared	0.23	0.19	0.17	0.16	0.19

Notes: Each column is a separate OLS regression of (Δ_{t-10}^{t}ln urban population)/100 on the change in market access measures shown, where t indexes years 1960 to 2010. "Overall Effect" is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, ln urban pop$_{t-10}$, and third order polynomials in longitude and latitude interacted with year fixed effects. Robust SEs, clustered by cell, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 2: Average Effect of Market Potential on Urban Population: IVs

Dependent Variable:	(Δ_{t-10}^{t} ln Urban Population)/100			
	OLS	IV: Exclude 5	IV: Exclude 10	IV: Exclude 15
	(1)	(2)	(3)	(4)
Δ_{t-10}^{t} ln Market Access	1.58***	2.98***	4.59***	5.75*
	[0.35]	[1.00]	[1.76]	[2.95]
Δ_{t-20}^{t-10} ln Market Access	1.23***	3.28***	5.76***	7.34***
	[0.26]	[0.87]	[1.59]	[2.46]
Δ_{t-30}^{t-20} ln Market Access	0.81***	2.57***	3.38**	4.60**
	[0.23]	[0.86]	[1.39]	[1.95]
Overall Effect	3.62***	8.83***	13.74***	17.69***
($t-30$ to t)	[0.59]	[1.89]	[3.31]	[4.64]
First stage Kleibergen-Paap F		114.00	41.86	17.41

Notes: Each column is a separate regression of (Δ_{t-10}^{t}ln urban population)/100 on the change in market access measures shown, where t indexes years 1990 to 2010, for 4,725 cell-years. "Overall Effect" is the sum of the contemporaneous effect and all lags shown. Each regression includes the same controls as Table 1. In columns 2–4 measures of Δln Market Access that exclude road surface changes within the radius shown (5, 10 and 15 cells respectively) instrument for the market access change measures. Robust SEs, clustered by cell, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 3: Robustness Checks: Overall Effects

	OLS (1)	IV: Excl. 5 (2)	IV: Excl. 10 (3)	IV: Excl. 15 (4)
(1) Baseline	3.62***	8.83***	13.74***	17.69***
(N=4,725; F: _; 114.0; 41.9; 17.4)	[0.59]	[1.89]	[3.31]	[4.64]
(2) Co-Investment: Inner: 2-3, Outer: 5-6	4.47***	10.56**		
(N=1,890; F: _; 17.4; _; _)	[1.17]	[4.93]		
(3) Co-Inv.: Inner: 2-3, Outer: 10-11	4.29***	10.82***	14.21**	
(N=2,197; F: _; 30.0; 10.6; _)	[1.03]	[3.70]	[6.34]	
(4) Co-Inv.: Inner: 2-3, Outer: 15-16	3.77***	10.63***	15.81**	21.31**
(N=2,260; F: _; 45.6; 13.3; 3.9)	[0.99]	[3.57]	[7.09]	[10.60]
(5) Radial Extension: Inner: 2-3, Outer: 5-6	4.23***	9.59***		
(N=2,098; F: _; 102.4; _; _)	[0.89]	[2.55]		
(6) Radial Ext.: Inner: 2-3, Outer: 10-11	3.55***	8.97***	13.33***	
(N=1,804; F: _; 78.0; 16.5; _)	[0.91]	[2.45]	[4.67]	
(7) Radial Ext.: Inner: 2-3, Outer: 15-16	4.22***	10.09***	14.81***	19.65***
(N=1,603; F: _; 66.0; 17.1; 6.7)	[0.94]	[2.62]	[4.78]	[6.90]
(8) Excl. Transcontinental Road Changes	3.62***	8.91***	13.65***	15.74***
(N=4,725; F: _; 51.9; 21.6; 8.7)	[0.59]	[2.34]	[3.91]	[6.03]
(9) Excl. National, Regional & Top 5 Cities	3.59***	8.41***	14.18***	17.28**
(N=3,801; F: _; 119.8; 18.5; 9.4)	[0.72]	[2.39]	[4.99]	[7.36]
(10) Excl. ≤100 km from Any Mine	3.81***	9.76***	15.97***	21.95***
(N=3,202; F: _; 97.6; 24.0; 14.0)	[0.72]	[2.37]	[4.38]	[6.31]
(11) Excl. ≤100 km from Cash Crop Cells	3.80***	9.15***	13.83***	17.72***
(N=4,606; F: _; 115.8; 42.2; 16.3)	[0.60]	[1.90]	[3.31]	[4.63]
(12) Excl. ≤100 km from President's Origin	3.93***	10.29***	17.58***	24.54***
(N=3,032; F: _; 95.7; 22.5; 8.6)	[0.75]	[2.22]	[4.23]	[6.84]
(13) Incl. All City-Level Controls	3.10***	8.44***	13.52***	19.19***
(N=4,725; F: _; 105.5; 28.1; 12.2)	[0.58]	[1.98]	[3.88]	[5.86]
(14) Incl. Two Lags of Population Growth	2.59***	5.17**	8.07**	12.94**
(N=2,264; F: _; 55.9; 28.4; 12.1)	[0.68]	[2.18]	[3.67]	[6.20]
(15) Fix Population to 1960 in IVs	3.62***	7.56***	11.96***	16.88***
(N=4,723; F: _; 68.6; 26.3; 6.6)	[0.59]	[1.62]	[2.67]	[4.15]
(16) Fix Pop. to 1960 in IVs & Drop 1980s	4.04***	6.59***	10.66***	16.31***
(N=3,629; F: _; 26.8; 7.0; 3.0)	[0.76]	[2.09]	[3.51]	[5.91]
(17) Fix Pop. to 1960 in Market Access (MA)	3.18***	9.80***	16.67***	25.40***
(N=4,723; F: _; 233.0; 70.1; 23.7)	[1.03]	[2.05]	[3.55]	[5.62]
(18) Fix Pop. to 1960 in MA & Drop 1980s	2.41*	8.38***	15.87***	26.97***
(N=3,629; F: _; 235.2; 66.9; 23.3)	[1.39]	[2.50]	[4.45]	[7.36]
(19) Fix Pop. to t-10 in Market Access (MA)	3.41***	10.34***	15.72***	20.96***
(N=4,725; F: _; 151.6; 47.1; 22.8)	[0.95]	[2.23]	[3.80]	[5.74]

Notes: This table is structured like Table 2 but only reports the overall effect. Rows 2–4: We remove the observations for which road building occurred 2-3 cells from the city and 5-6, 10-11 and 15-16 cells from the city within the same octant, respectively. Rows 5–7 analogously remove cities where investment in an outer zone octant occurs in an octant where the inner zone already has a paved or improved road. *, **, *** denote significance at the ten, five, and one percent level, respectively.

Table 4: Effect of Market Access on Night Lights

	(1) OLS	(2) IV: Excl. 5	(3) IV: Excl. 10	(4) IV: Excl. 15
Δ_{t-10}^{t} ln Market Access	0.69	23.45***	43.30***	67.11***
	[2.85]	[8.79]	[10.33]	[19.72]
Δ_{t-20}^{t-10} ln Market Access	2.05	12.07	7.73	5.56
	[2.28]	[7.79]	[11.34]	[16.82]
Δ_{t-30}^{t-20} ln Market Access	1.95	5.13	2.19	-1.71
	[1.87]	[4.94]	[7.61]	[10.77]
Overall Effect	4.69	40.65***	53.22***	70.96***
	[4.18]	[11.30]	[17.37]	[26.85]
First stage Kleibergen-Paap F		48.12	24.99	10.25

Notes: See Table 2. Outcome variable is Δ_{t-10}^{t}ln (Light Intensity). N=3,591 cell-decades. *, **, *** = 10, 5, 1% significance.

Table 5: Investigation of Population Reallocation across Cities

	OLS (1)	IV: Excl. 5 (2)	IV: Excl. 10 (3)	IV: Excl. 15 (4)
(1) Baseline	3.62***	8.83***	13.74***	17.69***
(N=4,725; F: _; 114.0; 41.9; 17.4)	[0.59]	[1.89]	[3.31]	[4.64]
(2) Urbanization≤18% (50th %ile) in t-30	3.35***	9.68***	14.22***	18.85***
(N=2,279; F: _; 76.9; 27.9; 12.3)	[0.80]	[2.54]	[4.05]	[5.62]
(3) Urbanization≤10% (25th %ile) in t-30	1.99**	6.18**	8.03**	7.41
(N=1,250; F: _; 36.4; 12.7; 4.2)	[0.91]	[2.60]	[3.95]	[4.92]
(4) Urbanization≤7% (10th %ile) in t-30	2.73**	10.49**	12.54**	14.37*
(N= 715; F: _; 12.2; 4.8; 1.4)	[1.25]	[4.09]	[6.18]	[8.14]
(5) 3x3 Mega-Cells	5.96***	8.54***	12.94**	12.28
(N=3,948; F: _; 33.0; 6.6; 1.0)	[0.78]	[3.20]	[5.30]	[7.98]
(6) 5x5 Mega-Cells	6.65***	7.25**	8.52*	9.84
(N=3,316; F: _; 11.0; 12.9; 4.2)	[0.96]	[3.07]	[5.00]	[6.87]
(7) 7x7 Mega-Cells	7.52***	12.53***	16.90**	16.61*
(N=2,778; F: _; 34.4; 4.8; 1.1)	[1.10]	[3.39]	[6.57]	[9.35]
(8) 9x9 Mega-Cells	9.01***	4.09	10.30	11.97
(N=2,320; F: _; 26.0; 10.0; 3.0)	[1.17]	[3.85]	[6.40]	[10.70]
(9) 3x3 Excl. National, Regional & Top 5	6.51***	9.15***	14.81**	18.40**
(N=3,068; F: _; 22.0; 15.3; 5.8)	[0.97]	[3.46]	[6.34]	[8.22]
(10) 5x5 Excl. National, Regional & Top 5	7.09***	8.33**	9.58	8.75
(N=2,468; F: _; 10.2; 11.3; 6.2)	[1.21]	[3.31]	[5.96]	[7.78]
(11) 7x7 Excl. National, Regional & Top 5	7.68***	12.14***	15.15**	14.33
(N=1,976; F: _; 42.4; 4.0; 1.1)	[1.38]	[3.49]	[7.49]	[12.06]
(12) 9x9 Excl. National, Regional & Top 5	9.49***	5.09	9.69	9.72
(N=1,563; F: _; 28.7; 10.7; 2.9)	[1.56]	[3.76]	[6.65]	[10.76]

Notes: This table is structured like Table 3. Rows 2–4 limit to countries below the urbanization rates shown. Rows 5–8: Baseline regressions for mega-cells that are a 3x3, 5x5, 7x7 or 9x9 square of the original 1x1 cells, respectively. The instruments are defined for the central cell of the mega-cell, where defined. Rows 9–12 show the same regressions on a sample dropping 1960 national and region capital cities and the five largest in each country. *, **, *** = 10, 5, 1% significance.

Table 6: Heterogeneous Effects of Market Access on Urban Population

	OLS Diff. (1)	Col. (2)–(4): IV5 0 (2)	1 (3)	Diff. (4)	IV10 Diff. (5)	IV15 Diff. (6)
(1) < Median Pop. t-30	3.10**	3.57	8.27***	4.70	11.57**	13.51**
(F: _; 27.6; 23.4; 8.2. Sh: 0.56)	[1.22]	[3.92]	[2.07]	[4.13]	[5.84]	[6.72]
(2) < Median 1960 MA	6.74***	-0.03	10.72***	10.75	24.02**	35.79***
(F: _; 9.3; 8.8; 6.7. Sh: 0.49)	[1.51]	[7.68]	[2.53]	[8.15]	[10.46]	[10.73]
(3) > Med.Dist.Top 1960 Cities	6.50***	1.76	9.89***	8.13***	15.04***	22.87***
(F: _; 46.0; 7.8; 1.8. Sh: 0.49)	[1.30]	[2.09]	[2.17]	[2.56]	[3.77]	[5.37]
(4) Land Suitability <25%	-0.93	6.80***	14.32***	7.51	14.59**	24.90***
(F: _; 9.2; 21.9; 6.6. Sh: 0.16)	[1.49]	[1.78]	[5.01]	[5.21]	[7.32]	[9.42]
(5) Land Suitability >75%	-1.60	9.38***	-2.15	-11.53**	-11.76	-18.99*
(F: _; 56.6; 20.1; 8.3. Sh: 0.05)	[1.94]	[1.94]	[4.90]	[5.09]	[8.70]	[10.33]
(6) Leader's Origin 150km t-10,t	-1.74	10.05***	2.72	-7.32*	-8.27	-11.56*
(F: _; 15.8; 12.1; 8.8. Sh: 0.24)	[1.19]	[1.93]	[3.99]	[4.07]	[5.06]	[6.10]
(7) Provincial Capital in 1960	0.08	7.93***	10.96***	3.03	2.59	7.24
(F: _; 9.8; 20.2; 5.2. Sh: 0.16)	[1.21]	[2.22]	[3.14]	[3.56]	[5.06]	[6.92]
(8) Provincial Capital in 2010	1.78	5.08**	11.91***	6.83**	9.36**	14.13**
(F: _; 22.9; 8.8; 4.0. Sh: 0.23.)	[1.18]	[1.98]	[2.93]	[3.23]	[4.70]	[6.23]

Notes: Each row reports results from variants of Table 2 (N=4,725), where the three market access variables are interacted with the dummy variable shown at left. IV5 results show the 30-year ($t-30$ to t) effect for both groups, along with the differential between them. The OLS, IV10 and IV15 columns show the differential only. The 1st stage F-statistics ("F") and the share of city-years with the dummy equal to one ("Sh") are reported in the left column. *, **, *** = 10, 5, 1% significance.

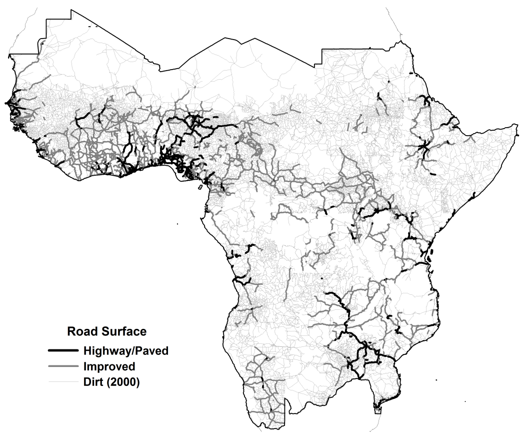
Table 7: Effect of Foreign versus Domestic Market Access

		OLS (1)	IV: Excl. 5 (2)	IV: Excl. 10 (3)	IV: Excl. 15 (4)
(1)	Domestic Market Access	3.17***	6.30***	7.93***	10.19***
		[0.56]	[2.01]	[2.76]	[3.66]
	Foreign Market Access	2.10*	3.90	3.32	2.51
		[1.23]	[3.22]	[3.33]	[3.56]
	First stage Kleibergen-Paap F		29.12	9.86	4.78
(2)	Overland Market Access	3.02***	7.72***	5.98***	4.44*
		[0.63]	[2.32]	[2.24]	[2.37]
	Overseas Market Access	3.39	1.67	8.16*	14.07**
		[2.37]	[4.03]	[4.85]	[6.17]
	First stage Kleibergen-Paap F		42.84	36.58	21.38

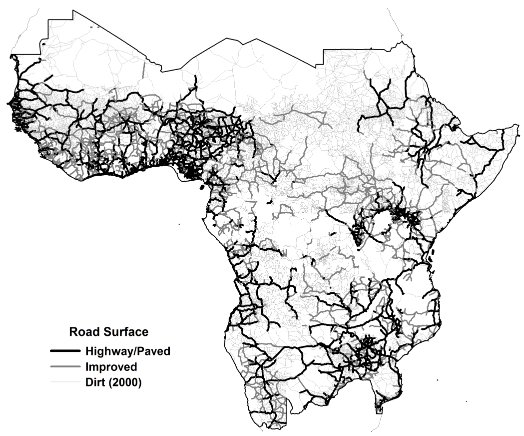
Notes: Each column contains summed coefficients from two separate regressions. In row (1), market access to domestic and foreign cities, and their lags, are entered separately (Obs.: 4,697). The six instruments always include IV-Foreign and its two lags for foreign cities, while the remaining three (for domestic) differ by column as shown. In row (2), market access to overland and overseas cities, and their lags, are entered separately (Obs.: 4,723). The six instruments always include IV5 and its two lags for overland cities, while the remaining three (for overseas) differ by column as shown. We control for log distance to the coast interacted with country-year FE. *, **, *** = 10, 5, 1% significance.

Figure 1: Road network maps in the 39-country sample, 1960 and 2010

(a) Roads Circa 1960



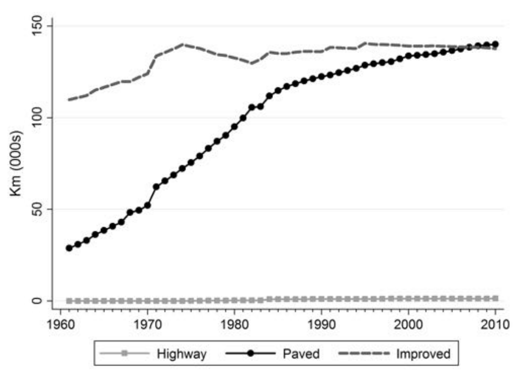
(b) Roads Circa 2010



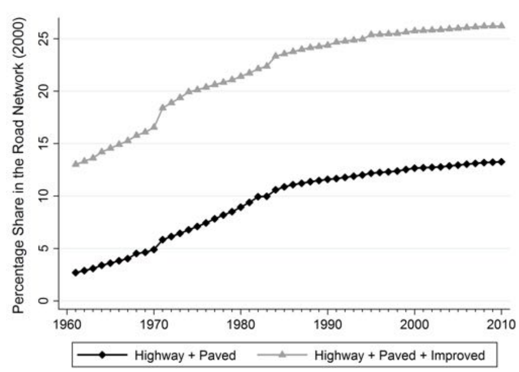
Notes: Subfigures 1a and 1b show the roads in the 39 sub-Saharan African countries of our sample in 1960 and in 2010 respectively. Roads are classified into four categories: *highways, paved, improved,* and *dirt*. See Appendix for details on data sources.

Figure 2: Road network evolution by type in the 39-country sample, 1960–2010

(a) Total Length of Each Type (Km)



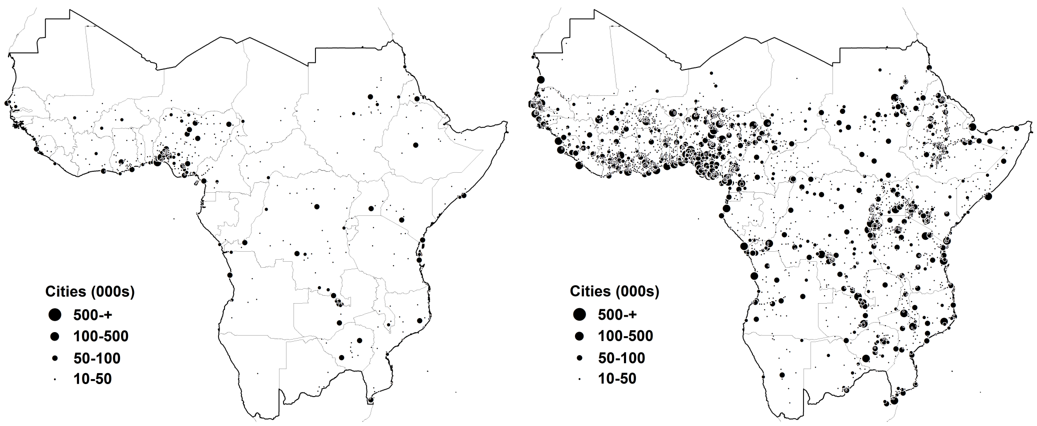
(b) Fraction of Each Type (%)



Notes: Total road network is defined circa 2004 based on Nelson and Deichmann (2004). See Appendix for details on data sources.

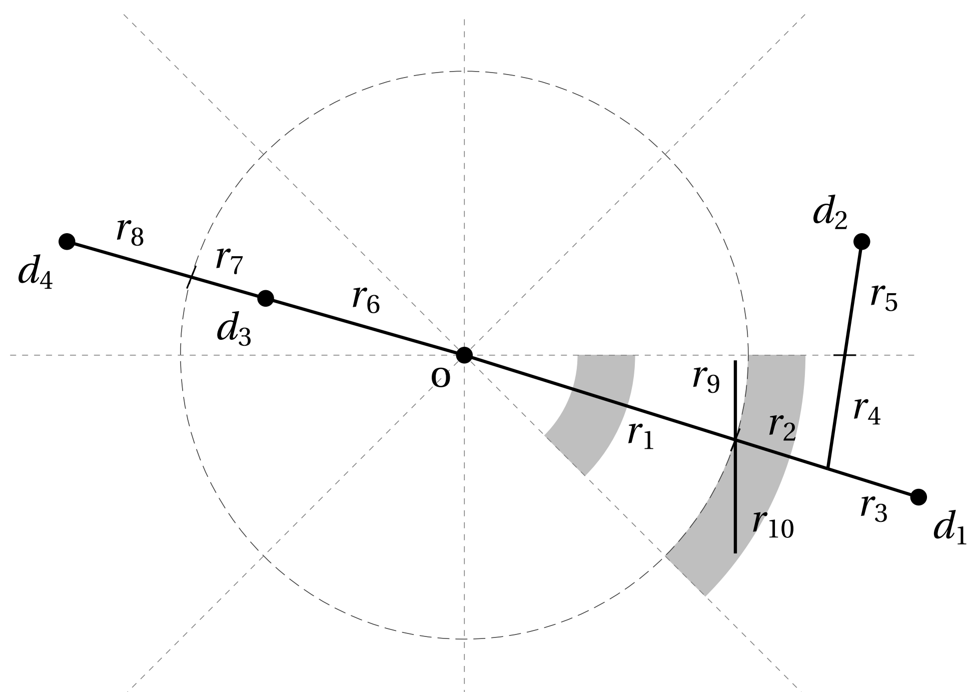
Figure 3: City population growth in the 39-country sample, 1960–2010

(a) Cities in 1960 (b) Cities in 2010



Notes: Subfigures 3a and 3b show the cities (defined as localities with population over 10,000 inh.) in our main 39-country sample in 1960 (N = 418) and in 2010 (N = 2,859) respectively. See Appendix for more details on data sources.

Figure 4: Identification strategies



Notes: See Section 2.