

**Institute for International Economic Policy Working Paper Series  
Elliott School of International Affairs  
The George Washington University**

**The Average and Heterogeneous Effects of Transportation  
Investments: Evidence from sub-Saharan Africa 1960-2010  
IIEP-WP-2019-8**

**Remi Jedwab  
Department of Economics, George Washington University  
Adam Storeygard  
Department of Economics, Tufts University**

**March 2019**

Institute for International Economic Policy  
1957 E St. NW, Suite 502  
Voice: (202) 994-5320  
Fax: (202) 994-5477  
Email: [iiep@gwu.edu](mailto:iiep@gwu.edu)  
Web: [iiep.gwu.edu](http://iiep.gwu.edu)

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

Remi Jedwab and Adam Storeygard\*

March 15, 2019

## 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 0.06–0.18. 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 Simon Alder, Treb Allen, Pol Antras, Nathaniel Baum-Snow, Kristian Behrens, Gharad Bryan, Kerem Cosar, Victor Couture, Dave Donaldson, Esther Duflo, 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, Georgetown, Harvard, HEC Montreal, IGC (LSE), IMF, McGill, Merced, NBER SI, NEUDC (Brown), North Dakota, Sciences Po, Stanford, UB/IEB, UdM, 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 Higuruchi, 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 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 0.6–1.8% 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, 2016b).<sup>1</sup> This combination of low urbanization and poor connectivity means that many people lack access to national and global markets (Limão and Venables, 2001). High transport costs still separate regions and ethnic groups within African countries and contribute to their high levels of spatial inequality and the persistent weakness of their states (Herbst, 2000; Acemoglu and Robinson, 2010). Today, sub-Saharan Africa is the region with

---

<sup>1</sup>According to our data, there are only 3,700 km of highways in Sub-Saharan Africa vs. 24,000 km in India (Government of India, 2016) and 111,900 km in China (Government of China, 2016).

the highest urban primacy rate (28% of its urban residents live in their country's largest city) and the second highest level of inequality (gini of 43%) (World Bank, 2016b). While road construction was rapid in the 1960s and 1970s post-Independence, it slowed substantially in the subsequent decades.

African countries have begun to make large transportation investments again.<sup>2</sup> Governments and international donors, including new donors like China through its expanding Belt and Road Initiative, describe them as having the potential to transform their regions, and highlight the potential of road projects to develop remote regions and reduce spatial inequality (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). Donors increasingly consider these projects in the context of a Trans-African Highway (TAH) system. It is thus imperative to consider the effect of earlier road construction on the economic geography of the region as a whole, 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, 2018) (for comprehensive overviews of the literature, see Redding and Turner (2015) and Berg et al. (2015)).<sup>3</sup> 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., 2017, 2019). Finally, a smaller literature has emphasized the specific role of road quality, which is the main source of variation in this work

---

<sup>2</sup>The Economist. 2015. “African roads and rails: All aboard.” Print edition, 28 February. Transport also accounted for 14% of World Bank lending, and 22% of African Development Bank disbursements 2012–2015 (World Bank, 2016a; African Development Bank, 2012–5).

<sup>3</sup>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).

(e.g. Casaburi et al., 2013; Gertler et al., 2015; Asher and Novosad, 2018). This paper makes several contributions to this literature.<sup>4</sup>

First, we document the development and effects 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 cross-sections instead of our six over 50 years, cannot. In particular, the effects take place over a period of 30 years and these long-run effects are about three times larger than short-run effects in the first decade. 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 35,000 km of highways (Faber, 2014; Baum-Snow et al., 2017, 2019). 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 (Bryan et al., 2014; Stanig and Wantchekon, 2015; Asher and Novosad, 2018). However, while rural (earthen) roads programs impact villages, they are far less costly than intercity road investments. Estimates of road-building costs from Collier et al. (2015) suggest that road upgrades in our sample cost 17% of endline regional GDP. By comparison, the large rural road program studied by Asher and Novosad (2018) cost 1.8% of India's GDP. We also focus on cities, which represent a large share of overall economic activity.<sup>5</sup>

Our second contribution is methodological. We develop a novel identification strategy, relying on the variation in market access induced by roads built far away. Our strategy is inspired by the market access approach of Donaldson and Hornbeck (2016) but we depart from their framework in several respects, instrumenting for market access changes with components that are most likely to be exogenous, and further ruling out several several potential channels for reverse causality. Our strategy of isolating variation from non-local changes to

---

<sup>4</sup>Berg et al. (2018) uses our panel data set on roads to study the effects of market access on land cultivation in Africa. The first working paper version of our paper predates their paper.

<sup>5</sup>McKinsey (2011) estimates that cities' contribution to sub-Saharan African GDP will be 63% by 2025. The cities in our sample account for two-thirds of night lights in our 39 sample countries.

the road network is akin to using “friends of friends” to estimate peer effects (e.g., Bramoullé et al., 2009; Calvo-Armengol et al., 2009; Mian et al., 2011; Goldsmith-Pinkham and Imbens, 2013). However, to our knowledge this strategy has been used rarely to study the effects of infrastructure, notable exceptions being Schlenker and Walker (2016) and Chiovelli, Michalopoulos and Papaioannou (2018).<sup>6</sup>

Indeed, with respect to the typology of identification strategies introduced by Redding and Turner (2015), our sample of 39 countries is not a context in which comprehensive planned or historical networks are available, our broad 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. These econometric approaches are most often applied to one country at a time, and results depend on the specificities of the natural experiment studied. More precisely, strategies based on planned/historical networks or accidental connections can explain well the location of road investments, but limit inference about the timing of effects. Randomized experiments and regression discontinuity designs have only been used to study rural roads, since their implementation is generally not politically feasible for intercity roads. In contrast, our identification strategy, although not as “clean”, has the advantage of being implementable for most types of transportation infrastructure and in most contexts, which could facilitate the comparison of effects across countries and over time. In principle, market access also accounts for aggregate effects and displacement of economic activity, unlike other strategies.

Third, we consider a wide variety of heterogeneous effects, which have received less attention in the empirical literature, and may be especially important given Africa’s diverse physical, economic, and political geography. We find suggestive evidence for three forms of heterogeneity. (i) The effect of market access is stronger for cities that are small and remote. 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,

---

<sup>6</sup>This strategy can ameliorate but not solve the “spatial reflection problem” that every location is affected by, and in turn affects, other locations. See Gibbons et al. (2015) for a discussion.

2013), but less so with recent papers on China (Faber, 2014; Baum-Snow et al., 2019). (ii) Effects are stronger for cities surrounded by poor farm land. This is consistent with Ricardian internal trade models, and echoes Asher and Novosad (2018), who find that roads cause outmigration from villages with low agricultural productivity. (iii) Cities less likely to be politically favored see bigger impacts. 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 sited based on political rather than economic returns, they may be less beneficial (Tanzi and Davoodi, 1998). We use a new dataset reporting place of origin of the 189 heads of state of 39 countries 1960–2010. To our knowledge, this is the first such dataset covering virtually all of sub-Saharan Africa. If anything, these results suggest that upgrading roads has decentralized economic activity in Africa, but not all non-primate towns benefited from such investments.

Access to ports plays an outside role, while access to foreign cities in neighboring countries does not, consistent with the overseas nature of much African trade. While the stronger role of access to world markets is in line with Fajgelbaum and Redding (2014) and Baum-Snow et al. (2017), the differential seems to be smaller in Africa, possibly due to oligopolistic intermediaries (Atkin and Donaldson, 2015).

Another important recent literature has estimated the effects of transport costs and infrastructure investment within a general equilibrium trade model (Allen and Arkolakis, 2014, 2016; Fajgelbaum and Redding, 2014; Alder, 2017; Donaldson and Hornbeck, 2016; 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.<sup>7</sup> 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

---

<sup>7</sup>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).

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.<sup>8</sup> Finally, while we use city size as a proxy for local economic development, city growth per se is an object of interest. Previous work has shown how urbanization in developing countries has effects on productivity (Meijers et al., 2016; Chauvin et al., 2017), access to amenities (Gollin et al., 2017; Jedwab and Vollrath, 2019), and democratization (Glaeser and Steinberg, 2016).

## 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 ( $\approx$  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.5 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.

---

<sup>8</sup>Fay and Opal (2000); Henderson, Roberts and Storeygard (2013); Gollin, Jedwab and Vollrath (2016); Jedwab, Christiaensen and Gindelsky (2017b); Jedwab and Vollrath (2019).

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.<sup>9</sup>

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.<sup>10</sup> The latter two sources of information are especially important, and new to this literature.<sup>11</sup> 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

---

<sup>9</sup>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.

<sup>10</sup>This paragraph is based on our discussions with Michelin employees.

<sup>11</sup>Burgess et al. (2015) use these data for Kenya 1964–2002 alone.

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*.<sup>12</sup> These sources generated estimates using various sources including population censuses, “non-native” population censuses, administrative counts, demographic studies, 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.<sup>13</sup> 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

---

<sup>12</sup><http://www.africapolis.org>; 15 countries are from part I and 18 from part II.

<sup>13</sup>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.

Wikipedia, aggregating multiple administrative cities into one agglomeration using more recent satellite imagery from Google Earth.<sup>14</sup>

The raw population data sources used for each of the 237 available country-years are listed in Appendix Table A.1. For example, for Botswana, we used the Population Censuses of 1956, 1964, 1971, 1981, 1991, 2001 and 2011. Of these 237, 177 take population counts for each individual city directly from population censuses. In robustness checks, we show that results are similar when we limit analysis to country-decades with estimates that are the most likely to be reliable, i.e. estimates based on a census year not too far from the actual year used in the regressions. Since the source dates vary across countries, population figures for all cities are exponentially interpolated and extrapolated between raw data years to obtain estimates for 1960, 1970, 1980, 1990, 2000 and 2010.<sup>15</sup>

The resulting sample includes population estimates for all 2,911 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. However, over our sample period 1960–2010, 84% of urban population growth, representing 171 of 203 million new urban residents, was on this intensive margin only.<sup>16</sup>

---

<sup>14</sup>City data details are in Appendix Section A.2. Sources are in Appendix Table A.1.

<sup>15</sup>Given two dates  $t$  and  $s$ , predicted population at time  $v$  is equal to  $= \exp((v - t) * \log(\exp((\log(pop_s) - \log(pop_t))/(s - t)) + \log(pop_t)))$ . In our sample, the initial and final populations are both at least 5 years from a population data source for about 6% of country-decades. Alternatively, the initial *or* final populations are more than 5 years from a population data source for about 64% of country-decades. We will thus investigate how results change when dropping these country-decades.

<sup>16</sup>98% of the 1,721 extensive margin cases are growth from below 10,000 to above it, with the remaining 2% representing declines to below it. Since we do not have data on the exact population of all city-years with a population below 10,000, we ignore the extensive margin in our analysis. However, we will show results are similar if we use estimates below 10,000 when available.

### 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.<sup>17</sup>

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.<sup>18</sup> 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).<sup>19</sup>

---

<sup>17</sup>For example, the Kenyan government has invested heavily in rebuilding the Mombasa-Nairobi road (Burgess et al., 2015). See Jedwab and Storeygard (2019) for evidence that our data are consistent with other sources at the country level, even for the most recent period.

<sup>18</sup>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.

<sup>19</sup>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

Our definition of market access follows Donaldson and Hornbeck (2016), who show that it summarizes direct and indirect effects of network changes in a large class of multi-region models. 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,  $\tau_{odt}$  is a travel time between cells  $o$  and  $d$ , and  $\theta$  is the elasticity measuring how trade volumes fall as travel times increase. Departing from Donaldson and Hornbeck (2016), we use travel times rather than iceberg trade costs, because no appropriate shipment value is available. We instead follow Duranton et al. (2014), whose central estimate of the elasticity of inter-city trade with respect to highway distance in the United States is -1.27, and Atkin and Donaldson (2015), whose results imply a trade cost-distance elasticity three times larger in Nigeria than in the United States. We combine these estimates and apply a baseline value of  $\theta = 3.8$  to assumed travel times. In our analysis, we focus on how changes in the road network affect travel times.<sup>20</sup>

Our unit of analysis is a 0.1 by 0.1 degree grid square ( $\approx 11 \times 11$  km). Using these units dramatically 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.<sup>21</sup> We assign to each grid square in each year a speed of travel

<sup>20</sup>Appendix Table A.10 considers an iceberg specification with a plausible shipment value. We are not aware of any work identifying an elasticity for intercity trade in Africa. Buys et al. (2010) report a trade-distance elasticity of -3.84 to -2.05 in a sample of country-pairs in sub-Saharan Africa. Elsewhere in the developing world, Morten and Oliveira (2017) report a trade-travel time elasticity of -2.65 across Brazilian meso-regions. We rely on the Duranton et al. (2014) estimate because it allows us to use the crosswalk to Africa inferred from Atkin and Donaldson (2015).

<sup>21</sup>Mean road length for 85,344 cells with a road in Nelson and Deichmann (2004) is 11.6 km. This suggests that our 11 x 11 km cells have on average about 1 road crossing them fully.

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.<sup>22</sup> The urban population of each cell is the population of the city in it, or in the small minority of cells with multiple cities, the sum of their populations.<sup>23</sup>

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.<sup>24</sup> When a map is not available for a given year, we interpolate speeds between the closest map years before and after.<sup>25</sup>

## 2.2. Baseline Specification

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

$$\ln P_{ot} = \beta_0 \ln MA_{ot} + \lambda_o + \rho_{ct} + \epsilon_{0ot} \quad (1)$$

which includes cell fixed effects  $\lambda_o$  and country-year fixed effects  $\rho_{ct}$  to account for time-invariant city characteristics and flexible national trends, respectively. 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:

---

<sup>22</sup>We show below that results hold if we use alternative trade elasticities or speeds.

<sup>23</sup>Our 2,911 cities fall within 2,789 distinct urban cells. We assume each city falls entirely in one cell, but verify that results are similar if we drop the largest cities that are most likely to span multiple cells.

<sup>24</sup>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.

<sup>25</sup>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 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.

$$\Delta \ln P_{ot} = \beta_0 \Delta \ln MA_{ot} + \Delta \rho_{ct} + \Delta \epsilon_{0ot}. \quad (2)$$

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/agglomeration effects or mean reversion, as it is standard in the literature on the growth of cities (Duranton and Puga, 2014). We also include a third-order polynomial in longitude and latitude interacted with year fixed effects to control for unobservables correlated across space within country-decades. This is equivalent to the main specification of Donaldson and Hornbeck (2016, eq. (13)), using population as an outcome instead of land rents.<sup>26</sup> We will nonetheless show that the baseline results, and the results when implementing our main robustness checks, hold if no additional controls are added except for the country-decade fixed effects. Finally, we cluster standard errors at the cell level. We will show that estimated standard errors clustered at the country level are modestly larger but do not affect inference substantially.

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:

$$\ln P_t = \beta_0 \ln T_t P_t + \epsilon_{0t} \quad (3)$$

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

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

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

---

<sup>26</sup>Land rent is the appropriate dependent variable, based on their model, but we do not have data on land rent. They show that population will also respond log-linearly to market access, but that the slope of this relationship will depend on the level of population mobility.

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.

Our first two identification strategies, excluding population change and excluding nearby road changes, are inspired by the econometric framework of Donaldson and Hornbeck (2016). The remaining strategies go farther in addressing the possibility of endogeneously built roads farther away, based on several different reasons why roads might be built to a city. Our instruments also reduce the bias due to measurement error.

**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:

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

**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 circle of radius  $j \in (5, 10, 15)$  cells (roughly 55, 111, or 167 km) of city  $o$ , we define a class of instruments IV  $j$ :

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

where  $\delta$  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 *outside* the exclusion circle ( $r_2$ ,  $r_3$ ,  $r_4$ ,  $r_5$  and  $r_8$ ) between  $t - 10$  and  $t$ . Any changes to local roads *inside* the exclusion circle ( $r_1$ ,  $r_6$  and  $r_7$ ) 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 circle 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.** Some types of non-local road building may also be endogenous to city  $o$ 's growth. We consider five such types. First, 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 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 circle ( $r_2$  and  $r_3$ ) may not satisfy the exclusion restriction because they are correlated with road changes inside the exclusion circle ( $r_1$ ).

Second, 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 outward*. 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.

Third, while the converse of this, inner roads built extending outer roads toward city  $o$ , are already excluded in our instrument, there is a subtle variant that requires a different solution. Specifically, a road built toward city  $o$  in anticipation of its growth may see faraway sections completed before near sections. We call this *radial extension inward*.

In order to address these three 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 notation, *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 outward* then implies  $\Delta R_{otq1} = 0$  but only because octant  $q$  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 octant 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. *Radial extension inward* implies that outer ring road building ( $\Delta R_{otq2} > 0$ ) anticipates city growth ( $\Delta P_{ot} > 0$ ) even before a connecting inner road is built ( $\Delta R_{otq1} = 0$ , but  $\Delta R_{o,t+10,q1} > 0$ ). Formally, we drop city  $o$  in years  $t$  to  $t+20$  if  $\exists q : \Delta R_{o,t+10,q1} > 0$  &  $\Delta R_{otq2} > 0$ . In Figure 4, this means dropping city  $o$  in year  $t$  if in any decade between  $t-30$  and  $t$ ,  $r_1$  (or  $r_6$ ) was upgraded in the decade after  $r_2$  ( $r_8$ ) was upgraded.

The fourth type of road building we exclude is along or near potentially important routes to large cities. Specifically, we create a larger exclusion zone for each city that extends along the shortest path to the nearest city of 100 thousand people in  $t-10$ . Formally, this is the convex hull of the exclusion circle and the nearest large city (100 thousand people in  $t-10$ ). Figure 5 shows what this would look like for city  $o$  if city  $d_1$  is the nearest large city. The resulting instrument relies on non-local road changes not directly targeted at city  $o$  (e.g.,  $r_5$ ).

Fifth, 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. Thirty-three of the 37 cities over 100 thousand people in 1960 were in the same cell as a transcontinental road (the other 4 cities are within 50 km of such a road). In addition, transcontinental road cells were much more likely to have been paved by 2010 (61% vs. 6% for non-transcontinental road cells). By making our instrument only rely on non-local changes that do not take place along transcontinental roads, we guarantee that identification does not come from these potentially endogenous connections to large cities.

**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. Lastly, cities far away from other cities might have roads built towards them precisely because of their isolation. We will thus drop isolated cities, and alternatively, countries where cities are relatively far from each other on average.

**Addressing correlated regional growth.** 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 all nearby 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  $d$ , the weights could also be endogenous. In that case, the instruments may not satisfy the exclusion restriction. One solution to this problem is to use the initial population of the other cities  $d$  in 1960, as opposed to  $t - 10$ , as weights in the instruments.<sup>27</sup> Alternatively, we use population in 1960 (or

---

<sup>27</sup>The second lag of the change in market access for the period 1980–1990 already used 1960 in

t-10) to define not just the instruments but also the main change in market access variable:  $MA_{ot} = \sum_{d \neq o} P_{d,1960} \tau_{odt}^{-\theta}$ . We do not use population in 1960 in our baseline specification because only half of the cities with a population of over 10,000 at some point since 1960 had a population above 10,000 in 1960. These cities are generally substantially larger than the full sample average. We will nevertheless show that results are similar when fixing population at 1960 levels in the IVs and implementing all robustness checks described in this section.

Finally, road investments far away from city  $o$  may be endogenous if city  $o$ 's region is large and road investments far away are targeted at other cities in that region because of their growth or decline. In this case, investments far away may be correlated with regional growth/decline which may also be correlated with city  $o$ 's growth/decline. We will address this issue by showing that regions are not large in our sample and by directly controlling for the growth of other cities in the same region as city  $o$ , using various definitions for "regions".

### 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. We thus only include two lags for the rest of the paper, so the sample contains the three decades 1980–

---

(6). We thus run this test on a sample dropping the 1980s as well.

2010.<sup>28</sup> Appendix Table A.3 contains descriptive statistics for this main sample.

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 3.5-4.5%.

### 3.2. Instrumental Variables (IV) Results

**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 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%, increasing with the radius. Alternatively, these results imply that a one standard deviation increase in market access growth is associated with a 0.46-0.88 standard deviation increase in city population growth. The first lag is larger than the contemporaneous term, though not significantly so. This may be because the contemporaneous term includes roads build late in a decade, allowing little time to have an impact. Moreover, Section 3.4. will show that population effects are slower to develop than effects on night lights.

As expected, the instrument is stronger at lower radii, because it includes road changes closer to the city. As shown in Appendix Table A.5, 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).

If we add a lead which we also instrument, the effects are smaller, generating 30-year elasticities between 6.8% and 9.9%, but the lead effect remains not significantly different from 0 (see Appendix Table A.4). However, we then lose one round of observations so our sample differs from the sample of Table 2.

---

<sup>28</sup>The 2-lag specification has lower Aikike and Bayesian information criteria than the 3-lag one.

**IV vs. OLS Estimates.** 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. Indeed, before the 1960s when most sample countries became independent, European powers disproportionately built roads around the capital city. For example, Herbst (2000, p.167) writes: “The colonialists essentially built the minimum number of roads necessary to rule [...] That means that colonies with large geographic masses had relatively low road stock because most of the road building under white rule was concentrated around the capital.” After 1960, African governments increasingly built roads towards remote regions. For example, O’Connor (1978, p.156) explains that “many new highways have been built into less developed districts [in African countries].” Consistent with this, in our sample market access is mean-reverting. When we regress log change in market access on log market access in  $t-10$  with our baseline controls, we find an elasticity of  $-0.025^{***}$ . Using log distance to the capital and the two largest cities in 1960 as an alternative measure of initial remoteness gives similar results.

Since larger exclusion circles are more likely to satisfy the exclusion restriction, and since IV estimates are biased towards the OLS when the exclusion restriction is not satisfied, the point estimate increases with the radius, and IV15 should be more reliable than IV5. However, the instrument also becomes weaker when one uses road changes farther away, which would also bias the IVs towards the OLS. The fact that estimates are higher for IV15 than for IV5 suggests that instrument weakness may not be that consequential in our context.

The higher IV could reflect heterogeneity in the overall effect, if the cities most likely to be impacted by road changes far away (the “compliers”) are cities for which changes in market access have stronger effects. However, below in Section 4., we show that the compliers are cities for which changes in market access have if anything smaller effects, especially with the larger-radius instruments. Therefore, our interpretation is that the IV strategies measure local average treatment effects that are lower than or equal to the average effect.

Finally, this downward bias may be the result of measurement error in the market access measure. If measurement error in road coverage is correlated across space, as is likely given that road projects often span more than one cell, instruments relying on nearer roads (i.e., IV5 for which the exclusion radius is 5 cells) may also be affected by the same measurement error. If so, instruments relying on roads farther away (e.g., IV15) could reduce measurement error because they (or their own classical measurement error) are less correlated with the measurement error of the instrumented market access variable. However, as we show in Appendix Section A.7, this also depends on the strength of the instruments, and IV15 is weaker. In the end, without data on the true road networks, we cannot estimate if our instruments solve for measurement error.

In summary, the IV coefficients are higher than the OLS most likely due to omitted variable bias and potentially due to measurement error.

**Magnitude.** We find 30-year elasticity of 0.06–0.18. This implies that a doubling of market access will increase city populations by 6 to 18%. How might one double market access, given our assumptions? In a symmetric context where all roads are unpaved, paving a random 21% of them, or improving 24%, would double market access. In a context where all roads are improved, paving half of them would similarly double MA.

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 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. We make no comparison to the literature identifying direct effects of nearby roads as our context lacks a source of variation identifying them separately. Section 3.4. compares effects on night lights to related literature.

### 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.<sup>29</sup> 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 outward* 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. In rows 8–10, cities with any *radial extension inward* 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 also fall, but results remain similar.<sup>30</sup>

<sup>29</sup>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.

<sup>30</sup>Radial extension inward requires dropping the 1980s because we don't know whether second lag changes in the 1960s follow inner changes in the 1950s. Excluding observations with potential extension inward decreases the sample size by a further 40–48%. This is less than the analogous 56–66% of observations with radial extension outward, consistent with the idea that on average road improvement proceeded outward away from cities over time, in line with the patterns of road decentralization described in Section 3.2.. Appendix Figure A.6 shows road building is in rings closer to cities relatively more in the 1960s–1970s (and the 2000s when very few roads were built) than in the 1980s–1990s (see Appendix Section A.8 for details).

Row 11 redefines the instrument excluding any road building within the convex hull of the 5-, 10- and 15-cell exclusion circle and the nearest city of at least 100 thousand. This convex hull is the area where a road to such a nearby large city is most likely to be built. Row 12 excludes 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  $t$ ). Results change little.<sup>31</sup>

**Dropping potential growth hubs.** Row 13 drops from the sample each country's 5 largest cities and national and regional capitals from 1960. This is akin to 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.<sup>32</sup> Rows 14–16 drop (i) cities within 100 km of a mine open at any time between 1960 and 2010 (11); (ii) cities within 100 km of a cell whose land suitability for cash crops is above 90% (12); and (iii) cities within 100 km of the hometown of any of the country's head of states between 1960 and 2010 (13). 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.8 and Appendix Table A.6). Likewise, results hold if we add many controls proxying for physical, economic and political geography (row 17).<sup>33</sup>

---

<sup>31</sup>Results are also broadly similar (see Appendix Section A.8 and Appendix Table A.6) 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). Results also hold if the convex hull is constructed for the nearest city of at least 50 thousand (see Appendix Section A.8 and Appendix Table A.6). The resulting exclusion zones are weakly smaller since the nearest 50+ city is at least as close as the nearest 100k+ city.

<sup>32</sup>Appendix Table A.6 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.

<sup>33</sup>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

Next, the effects are generally smaller if we: (i) drop “isolated” city-years, i.e. urban cells that are in the top 25% of the Euclidean distance to the nearest cell with a city at any point between 1960 and 2010 (row 18); and (ii) drop country-years where cities are relatively far from each other on average, i.e. country-years where the average Euclidean distance between cells with a city at any point between 1960 and 2010 is in the top 25% in the sample (row 19). The fact that the effects are smaller when excluding less connected cities or countries is consistent with results on heterogeneous effects presented later on.<sup>34</sup>

Finally, results hold if we drop countries that became independent late or country-decades in which the country experienced a war, received refugees or suffered a multi-year drought (see Appendix Section A.8 and Appendix Table A.9).

**Correlated regional growth.** In rows 20–23, population is fixed at its 1960 level in constructing the instruments, and in rows 22–23, in the instrumented market access (*MA*) as well. In rows 21 and 23, the 1980s are dropped because they are the only decade in which an included lag of  $\Delta MA$  uses the population of the other cities in 1960. In addition, Appendix Table A.8 shows that all the results of Table 3 hold if population is fixed at its 1960 level in the instruments. In row 24, 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 20–21, and IV15 estimates are larger in rows 22–24.

Lastly, we report results controlling for regional growth, i.e. the growth between  $t-10$  and  $t$  of the other cities belonging to the same region as the city. We define regions in three ways: administratively (1960 provinces), along ethnic lines (ethnic zones from Murdock (1959)), and geometrically (circles of different radii). Provinces and ethnic groups are not large in our sample. A 15-cell radius circle ( $7000 km^2$ ) is at the the 50th and 88th percentiles, respectively, of the province and ethnic area size distributions. In other words, 50% of provinces

---

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.

<sup>34</sup>We use cells that are “urban” at any point in 1960-2010 because we do not want our measures of city or country isolation to be mismeasured simply because we do not want have good population data below 10,000. Note that these results hold if we drop city-years (country-years) in the bottom 25% of (average) market access in 1960 or 2010 (not shown).

and 88% of ethnic areas are smaller than our largest exclusion zones. In row 25, we simultaneously control for total population growth of the other cities of the same 1960 province *and* ethnic area *and* circle of 15 cells. We drop cities for which one of these three growth rates is likely to be imprecisely measured because the region does not contain at least two other cities. The effects are somewhat smaller, at 6.3–8.5%.<sup>35</sup>

**Main specification checks.** Our baseline specification controls for initial log population, since it is standard in the literature. However, its potential endogeneity is a concern. Row 2 of Table 4 shows that the baseline IV results (row 1) are unchanged when removing this control. The IV estimates decrease only slightly when the other baseline control, the polynomial in longitude and latitude is also excluded (Row 3). Appendix Table A.7 shows that all of the results of Table 3 are stable when these controls are excluded.

Our baseline specification with three distinct periods of market access change allows effects to vary across three decades during and after road construction. Alternatively, in row 4, we collapse these three changes into one change from  $t - 30$  to  $t$ . The instrument set, with only one instrument for one endogenous variable, is substantially stronger, but point estimates fall by about two-thirds. This likely reflects two aspects of the change. First, pooling doubles the variance of change in market access across observations, while the variance of city growth remains the same. In standardized terms, 30-year market access has half as large an effect as the pooled 10-year changes (a one standard deviation in market access growth is associated with a 0.25-0.44 standard deviation increase in population growth vs. 0.46-0.88 when using 10-year market access). Second, and more importantly, there is meaningful decadal variation in market access during the  $(t-30,t)$  period that is being averaged away, imposing the restriction that changes in  $(t-30,t-20)$ ,  $(t-20,t-10)$  and  $(t-10,t)$  all have the same effects on population growth in  $(t-10,t)$ . However, Table 2 has shown that the contemporaneous effect differs from the lagged effects. Table 5 will show that effects differ even more strongly across lags when studying night

<sup>35</sup>In Appendix Table A.6, we show these results remain similar if we also control for growth in a 16-30 cell ring around the 15 cell circle.

light growth. Studying short-term effects is thus essential.

Rows 5 and 6 report specifications akin to Donaldson and Hornbeck (2016). In row 5, the reduced form effects of the instruments on city population growth are substantially larger, consistent with lower variance of the instruments. We prefer the IV specification because it is more comparable with the rest of the literature. Row 6 shows that results are similar if we weight observations by their initial population size in  $t - 10$ . Row 7 clusters standard errors at the country level, to account for the fact that much of the data is collected by country. The estimated standard errors are only slightly larger.

**Population data quality.** Our main sample has the advantage of applying a consistent population threshold across all countries and using the same years across all countries. This strategy has two important flaws. First, the sample is not balanced. Places that entered the sample earlier may have been different from other cities in ways that are correlated with road building. Second, it requires interpolation and extrapolation, sometimes several years away from censuses or other estimates. This affects both the dependent variable and the variables of interest. As long as this interpolation and extrapolation does not systematically overestimate or underestimate either of these, measurement error will be classical, biasing estimates downward. And country-decade fixed effects control for measurement error differences at the country-decade level.

Row 8 of Table 4 shows results are similar to the baseline (row 1) if we restrict to a balanced sample of cells a population over 10,000 in all three years. Likewise, results hold if we use alternative thresholds to define cities: 15,000 (row 9) and 20,000 (row 10). Rows 11–12 use additional population estimates for cell-years under 10,000 to increase the sample's balance. Row 11 uses all cell-years with non-zero population estimates. The sample size increases to 7,369, only about 10% less than a balanced sample. Row 12 adds to the baseline sample cell-years with non-zero population estimates only for the one year prior to crossing the 10,000 threshold. The longer the period a city has had its population recorded, despite it not being above 10,000, the more likely it is special in some unrecorded way. In both cases, IV estimates are generally slightly smaller.

Row 13 drops the 1980s, the one decade that uses data from the 1960s in our two-lag specification. Our population data for the 1960s include four countries in which we extrapolate city populations from a census of European residents only. The road data are also less complete. Coefficients are slightly larger than at baseline. In row 14, we drop the 2000s, for which both the population data and the road data may be incomplete. For example, the last raw year of city population data is before 2005 for 20 countries. Estimated coefficients fall somewhat more at the higher radii, but remain large and significant.

Rows 15–17 restrict the sample to country-decades with the population estimates most likely to be reliable. Row 15 restricts the sample to periods whose beginning and end populations are each based on at least two census populations, as opposed to other sources. Row 16 excludes country-decades for which the initial and final populations are both at least 5 years from a population data source. Results are similar to baseline. Row 17 excludes country-decades for which the initial *or* final populations are more than 5 years, respectively, from a population data source, reducing the sample by more than 50 percent. The point estimates are reduced as well, and while the 15-cell instrument is weakened substantially, the 5-cell and 10-cell instruments remain strong and suggest effects that are significant, if reduced by up to a third from baseline.

**Road data quality.** Rows 13 and 14 showed that results are similar when excluding data from the 1960s and the 2000s, two decades when the road data are most likely to be incomplete. In addition, we verify that results hold if we drop country-decades or countries which we believe to have poor road data. More precisely, for each of the 39 countries  $\times$  5 decades = 195 country-decades in the sample, we compute the growth rate in total paved road kilometers (recalling that paved roads were the main form of investment during our period), and compare it to an analogous estimate from official national sources compiled by Canning (1998), Canning and Farahani (2007) and World Bank (2015). The mean of the absolute value of this difference is about 0.2, meaning that we typically over- or under-estimate paved road building by 20%). Results are similar if we drop: (i) country-decades for which the absolute value difference is more than the mean

in our main sample (row 18); (ii) whole countries for which the mean absolute value difference is more than the mean in the sample (row 20). In row 19 and 21, we also drop country-decades or countries for which official data is missing.

**Other specification and sample checks.** In Appendix Section A.9 and Appendix Table A.10, we show the effects are robust to: (i) replacing the country-decade fixed effects with decade fixed effects; (ii) using alternative speeds; (iii) allowing railroad travel; (iv) using alternative distance elasticities of trade flows; (v) adding uniform costs of crossing borders; (vi) using iceberg costs; and (vii) excluding countries bordering South Africa, North Africa or the Arabian Peninsula, as their market access may be underestimated. Assigning larger (smaller) values of  $\theta$  mechanically decreases (increases) the coefficients on market access, as it shrinks (widens) the variation in market access, without altering the variation in city growth. However, as in Donaldson and Hornbeck (2016), the effect of a one standard deviation increase in  $\Delta_{t-10}^t \ln MA$ ,  $\Delta_{t-20}^{t-10} \ln MA$ , and  $\Delta_{t-30}^{t-20} \ln MA$  is quite stable for values of  $\theta \geq 2$  (Appendix Figure A.7). Our results thus do not depend on the chosen elasticity.<sup>36</sup>

Appendix Table A.11 considers alternative specifications that also include access to total (urban and rural) population and to mines, each of which might also make a location more attractive. There is no evidence of impacts, and little effect on the effect of urban market access, though both variables are likely to be highly mismeasured (see Appendix Section A.10 for details).

### 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.

---

<sup>36</sup>We verify that higher elasticities weaken our instruments (available upon request).

To explore this idea, we consider changes in night lights as a proxy for overall output.<sup>37</sup> The sample includes 3,591 observations, for the periods 1992–2000 and 2000–2010 only, because lights data begin in 1992. Appendix Section A.10 and Appendix Table A.12 show that the 30-year overall effects on population for this restricted sample are similar to those for the full samples.

Table 5 reports results for night lights. While OLS estimates of the market access coefficients are not large and much less precisely specified than in the population regressions of Table 2, in each IV specification, market access has substantially larger effects on lights than on population. The effect size is comparable to what Alder (2017) finds for the Golden Quadrilateral in India, but substantially larger than the 0.2–0.3 found by Chiovelli et al. (2018) in post-conflict Mozambique, where transport costs fall due to both road-building and landmine clearance. The effects are entirely in the first decade. Rows 15 and 16 of Table 4 show that IV estimates of the effect of market access on night lights controlling for population and on night lights per capita are also substantially positive. (Appendix Table A.12 shows that this effect is also 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 slow migration limiting the transition to spatial equilibrium, as Chauvin et al. (2017) show for India.<sup>38</sup>

15, 8 and 1 countries were still colonies and had migration restrictions imposed upon the movement of non-European Africans in the 1960s, the 1970s and the 1980s, respectively. This may have contributed to lower and slower effects of roads. However, Appendix Table A.9 shows results are similar if we drop these countries. Indeed, since we study the long-term effects of roads, we focus on city growth from 1980-2010, when almost all countries were independent.<sup>39</sup>

---

<sup>37</sup>Henderson et al. (2012) show that in a sample of developing countries, changes in night lights are correlated with changes in GDP, with an elasticity of about 0.3.

<sup>38</sup>Gollin et al. (2017) argue more broadly that urbanization trends in the developing world are inconsistent with simple static notions of spatial equilibrium.

<sup>39</sup>Henderson et al. (2018) find that ethnic diversity constrains the growth of primate cities, possibly by increasing the costs of migration to cities outside ethnic homelands. Therefore, the

### 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.<sup>40</sup>

Table 6 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 ( $\approx 18\%$ ) has very little effect on results. In row 3, restricting to the bottom quartile ( $\approx 10\%$ ) 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 decile ( $\approx 7\%$ ) of countries in row 4, results are more similar to the full sample (though again the IV15 and now IV10 instruments are quite weak).

---

migration response to increased market access might be slower in countries with more ethnic heterogeneity. We test this in Appendix Table A.13 (see Appendix Section A.10) where we interact the three change in market access variables with a dummy equal to one if the country's degree of ethnic fractionalization in 1960, from Alesina and Ferrara (2005), is higher than the mean degree of fractionalization in the sample. We consider effects separately for lights and population. In the bottom panel, we report the share of each lag in the overall effect. While the IV10 and IV15 instruments are quite weak, in general the contemporaneous effects are stronger on night lights than on population in the first lag. More interestingly, in general the population-minus-lights differential is largest in the first lag in low fractionalization countries, but in the second lag in the high fractionalization countries. The differences are not significant at conventional levels, and more fractionalized countries may be different in other ways, but the signs are consistent with this idea that population is slower to adjust in more fractionalized countries.

<sup>40</sup>This differs from the context of middle- or high-income countries like China (urban share  $\approx 55\%$  today) and the U.S. (80%). Urban reallocation is mechanically more likely there.

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.<sup>41</sup>

#### 4. Results: Heterogeneous Effects

Transport investments may have different effects depending on the local context in which they take place. Table 7 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 6, 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.

These exercises are demanding on the data, with six endogenous variables and six instruments per regression. Therefore, each type of heterogeneity is considered one at a time at baseline. 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.

**Economic Geography.** Rows 1–3 of Table 7 show variation with respect to economic geography characteristics. Core-periphery models predict that reduced trade costs increase the size of big cities more than small 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.<sup>42</sup>

---

<sup>41</sup>See footnote 19 on the limits of the Demographic and Health Surveys for our purposes. We also cannot learn about rural areas directly. As we describe in Appendix Section A.9, we do not believe that existing panel databases of total or rural populations are reliable enough.

<sup>42</sup>Differentials based on the country’s 25th or 75th percentile population, dropping the top

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.<sup>43</sup>

**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 cell of the city, cutting the sample at 75% and 25% percent suitability.<sup>44</sup> 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% grow relatively slower when market access increases (row 5). 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 also consistent with cities in less agricultural areas specializing in more transport intensive activities that benefit more from the roads.<sup>45</sup>

**Political Geography.** Rows 6–8 of Table 7 show variation with respect to two political geography characteristics. Row 6 allows for a differential effect for city-

---

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.11 and Appendix Table A.14).

<sup>43</sup>Appendix Section A.11 and Appendix Table A.14 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.

<sup>44</sup>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).

<sup>45</sup>We find generally similar effects as for land suitability when we study the interaction effects with rainfall (see Appendix Section A.11 and Appendix Table A.15). We also do not find any differential for cities closer to mines, a sector that may or may not be labor-intensive.

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). 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.11, we report more results that overall suggest stronger negative effects for leaders with a longer tenure and whose regime is not democratic.<sup>46</sup>

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 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.

**Multiple forms of heterogeneity.** The forms of heterogeneity we consider are

---

<sup>46</sup>Appendix Section A.11 and Appendix Table A.15 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.

potentially related to each other. In practice, the correlation matrix of the eight dummies shown at left contains only 4 coefficients of more than 0.1 in absolute value.<sup>47</sup> The dummies thus capture different city-years. Appendix Table A.16 reports results when we simultaneously include the market access measures interacted with both heterogeneity dummies in the three of the four most correlated pairs that are not mechanically related (we do not consider the 1960 and 2010 provincial capitals together). Appendix Table A.17 then reports results when we simultaneously include the market access measures interacted with two important heterogeneity dummies—population in  $t-30$  and market access in 1960—and the market access measures interacted with each other heterogeneity dummy one by one. Results for most forms of heterogeneity except land suitability are broadly similar to the baseline heterogeneity results in Table 7, though the instruments are weaker with 9 or 12 endogenous variables.

**First stage heterogeneity and local average treatment effects (LATEs).** The results from Table 7 suggest that market access has stronger effects on the population of smaller and more remote cities. As noted above in Section 3.2., the market access of remote cities could also be more strongly influenced by faraway road changes than the market access of less remote cities. In this case, both the first and second stage effects in our baseline IV regressions could be increasing with remoteness. Since IV estimates reflect the LATE, this is a possible explanation for why our baseline IV estimates in Table 2 are larger than their OLS counterparts. In that case, we would expect the overall average effect to be lower than the estimated LATE.

To consider this possibility, we investigate heterogeneity by remoteness (lower than national median market access in 1960) in the first stage in Appendix Table A.18 panel A. For each instrument set (IV5, IV10, and IV15), there are three first stage equations, one for each lag ( $t - 30$  to  $t - 20$ ,  $t - 20$  to  $t - 10$  and  $t - 10$  to  $t$ ). More remote cities have if anything a weaker first stage relationship than

---

<sup>47</sup>The strongest correlations are: (i) 0.71 between the 1960 provincial capital dummy (row 7) and the 2010 provincial capital dummy (row 8); (ii) 0.40 between the “below median market access in 1960” dummy (row 3) and the “above median distance to top cities in 1960” dummy (row 4); (iii) -0.31 or -0.26 between the “below median population in  $t-30$ ” dummy (row 1); and (iv) the 2010 provincial capital dummy (row 8) or the 1960 provincial capital dummy (row 7).

less remote cities. This suggests that our main effects capture LATEs that are concentrated among *less* remote places, where results from Table 7 suggest that market access has weaker effects. Thus we expect that the overall average effect of market access is if anything larger than the LATEs estimated in Table 2. Panels B and C show analogous results based on an alternative form of remoteness (farther than the median distance from the country's top cities in 1960; panel B) and city size (population below the national median in  $t - 30$ ; panel C).<sup>48</sup>

**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.10 show that adding substantial uniform border costs has little effect on results. In Table 8, 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  $\times$  three lags) and six instruments, so instruments are weaker.<sup>49</sup>

Row 1 of Table 8 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.<sup>50</sup> A one standard deviation change in domestic (foreign) market access is associated with a 0.37–0.61 (0.06–0.10) standard deviation change in city population growth.<sup>51</sup>

<sup>48</sup>The coefficients of interest are highlighted in bold in the table. If anything, the differential effects are more negative for IV15 than for IV5, another indication that the IV effect does not increase with the radius because it captures the LATEs of smaller and more remote places.

<sup>49</sup>When averaging information into 30-year changes, instruments are strong, and the relative roles of foreign and overseas MA decrease (available upon request).

<sup>50</sup>Other combinations of instruments are generally weaker. Domestic and foreign market access changes are correlated, but only at 0.27.

<sup>51</sup>Connections to wealthier cities/countries may be more important than connections to

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 captures 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.<sup>52</sup> 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.<sup>53</sup>

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.<sup>54</sup>

---

poorer cities/countries. However, Appendix Section A.11 and Appendix Table A.19 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.

<sup>52</sup>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.

<sup>53</sup>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).

<sup>54</sup>Appendix Section A.11 and Appendix Table A.19 show effects are robust to: (i) dropping the ports themselves; (ii) including the ports in calculating overland market access and its 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.

**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. Since we control for initial city size, this is not convergence per se, but these are different (and largely uncorrelated) components of being less developed. 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. Concluding Discussion

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.

In Appendix Section [A.12](#), using additional assumptions 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. Under the scenario of no reallocation across cities, for which we provide some evidence, our estimates attribute a sizable 5–10% of the intensive margin increase in the urban share between 1960 and 2010 to these road upgrades. Conversely, applying our estimated elasticities to the proposed TAH network suggests that it will only marginally increase the urbanization already projected over the subsequent 30 years.

Several mechanisms could be driving our 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.

## References

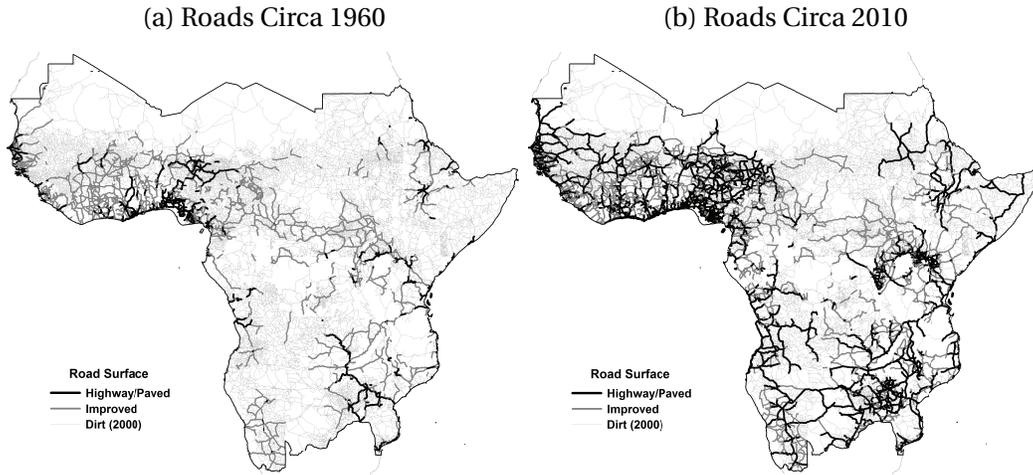
- Acemoglu, Daron and James A. Robinson, "Why is Africa Poor?," *Economic History of Developing Regions*, 06/2010 2010, 25 (1 June 2010), 21–50.
- , Simon Johnson, and James Robinson, "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review*, 2005, 95, 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/Tunis: African Development Bank, 2012–5.
- 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 December 2017.
- Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, September 2005, 43 (3), 762–800.
- 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, "Rural Roads and Local Economic Development," mimeo, Dartmouth 2018.
- 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, 435–448.
- , Vernon Henderson, Matthew Turner, Qinghua Zhang, and Loren Brandt, "Does Investment in National Highways Help or Hurt Hinterland City Growth?," *Journal of Urban Economics*, 2019.
- 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., Brian Blankespoor, and Harris Selod, "Roads and Rural Development in Sub-Saharan Africa," *The Journal of Development Studies*, 2018, 54 (5), 856–874.
- , 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.
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, May 2009, 150 (1), 41–55.
- 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.
- Buys, Piet, Uwe Deichmann, and David Wheeler, “Road Network Upgrading and Overland Trade Expansion in sub-Saharan Africa,” *Journal of African Economies*, 2010, 19 (3), 399–432.
- Calvo-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, “Peer Effects and Social Networks in Education,” *Review of Economic Studies*, 2009, 76 (4), 1239–1267.
- Canning, David, “A Database of World Stocks of Infrastructure: 1950–1995,” *World Bank Economic Review*, 1998, 12, 529–548.
- and Mansour Farahani, “A Database of World Stocks of Infrastructure: Update 1950–2005,” mimeo, Harvard School of Public Health Sep 2007.
- 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.
- Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio, “What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States,” *Journal of Urban Economics*, 2017, 98 (C), 17–49.
- Chiovelli, Giorgio, Stelios Michalopoulos, and Elias Papaioannou, “Landmines and Spatial Development,” NBER Working Papers 24758, National Bureau of Economic Research, Inc June 2018.
- 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*, April 2018, 108 (4-5), 899–934.
- and Richard Hornbeck, “Railroads and American Economic Growth: A ‘Market Access’ Approach,” *Quarterly Journal of Economics*, 2016, 131 (2), 799–858.
- Duranton, Gilles and Diego Puga, “The growth of cities,” in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 2, Amsterdam: North-Holland, 2014, pp. 781–853.
- 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,” NBER Working Papers 20217, National Bureau of Economic Research, Inc June 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.
- Gibbons, Stephen, Henry Overman, and Eleonora Patacchini, “Spatial Methods,” in Gilles Duranton, Vernon Henderson, and William Strange, eds., *Handbook of Urban and Regional Economics*, Vol. 5, Elsevier, 2015, chapter Chapter 3, pp. 115–168.
- Glaeser, Edward L and Bryce Millett Steinberg, “Transforming Cities: Does Urbanization Promote Democratic Change?,” Working Paper 22860, National Bureau of Economic Research November 2016.
- Goldsmith-Pinkham, Paul and Guido W. Imbens, “Social Networks and the Identification of Peer Effects,” *Journal of Business & Economic Statistics*, 2013, 31 (3), 253–264.
- Gollin, Doug, Rémi Jedwab, and Dietrich Vollrath, “Urbanization with and without Industrialization,” *Journal of Economic Growth*, March 2016, 21 (1), 35–70.
- Gollin, Douglas, Martina Kirchberger, and David Lagakos, “In Search of a Spatial Equilibrium in the Developing World,” CEPR Discussion Papers 12114, C.E.P.R. Discussion Papers June 2017.
- 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, “Measuring Economic Growth from Outer Space,” *American Economic Review*, April 2012, 102 (2), 994–1028.
- , —, and Uwe Deichmann, “Has climate change driven urbanization in Africa?,” *Journal of Development Economics*, January 2017, 124, 60–82.
- , Ulrich Eberle, Dominic Rohner, and Kurt Schmidheiny, “Ethnic Diversity and City Size,” mimeo, LSE 2018.
- Herbst, Jeffrey, *States and Power in Africa: Comparative Lessons in Authority and Control*, Princeton, NJ: Princeton University Press, 2000.
- Jedwab, Rémi and Adam Storeygard, “Economic and Political Factors in Infrastructure Investment: Evidence from Railroads and Roads in Africa 1960–2015,” mimeo, George Washington University 2019.
- 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.
- Jedwab, Remi and Dietrich Vollrath, “The Urban Mortality Transition and Poor-Country Urbanization,” *American Economic Journal: Macroeconomics*, January 2019, 11 (1), 223–75.
- Jedwab, Rémi, 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.
- McKinsey, *Urban world: Mapping the economic power of cities*, New York: McKinsey Global Institute, 2011.
- Meijers, Evert, Martijn Burger, and Gilles Duranton, “Determinants of city growth in Colombia,” *Papers in Regional Science*, March 2016, 95 (1), 101–131.

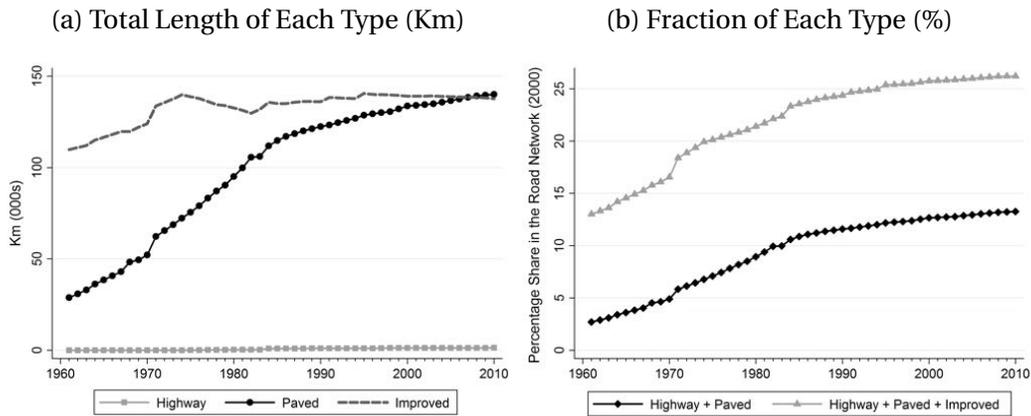
- Mian, Atif, Qamar Abid, and Asim Khwaja, "Bank Credit And Business Networks," Scholarly Articles, Harvard Kennedy School of Government 2011.
- Michaels, Guy, "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *Review of Economics and Statistics*, November 2008, 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.
- Schlenker, Wolfram and W. Reed Walker, "Airports, Air Pollution, and Contemporaneous Health," *The Review of Economic Studies*, 2016, 83 (2), 768–809.
- 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.
- World Bank, *Road Deterioration in Developing Countries: Causes and Remedies*. 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.
- , *World Development Indicators*, Washington: World Bank, 2016.
- Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, 128 (4), 1727–1785.

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



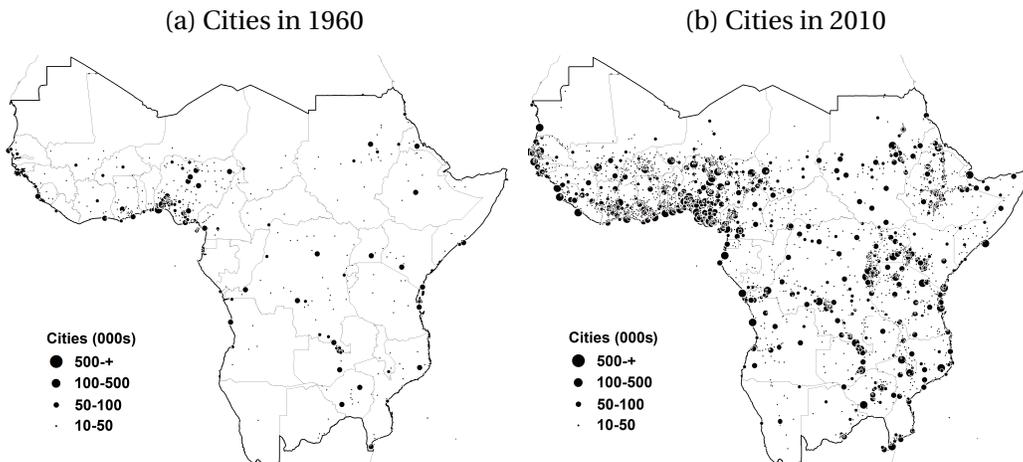
Notes: Subfigures 1a and 1b show the roads in the 39 sub-Saharan African countries of our sample in 1960 and in 2010 respectively. Roads are classified into four categories: *highways*, *paved*, *improved*, and *dirt*. See Appendix for details.

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



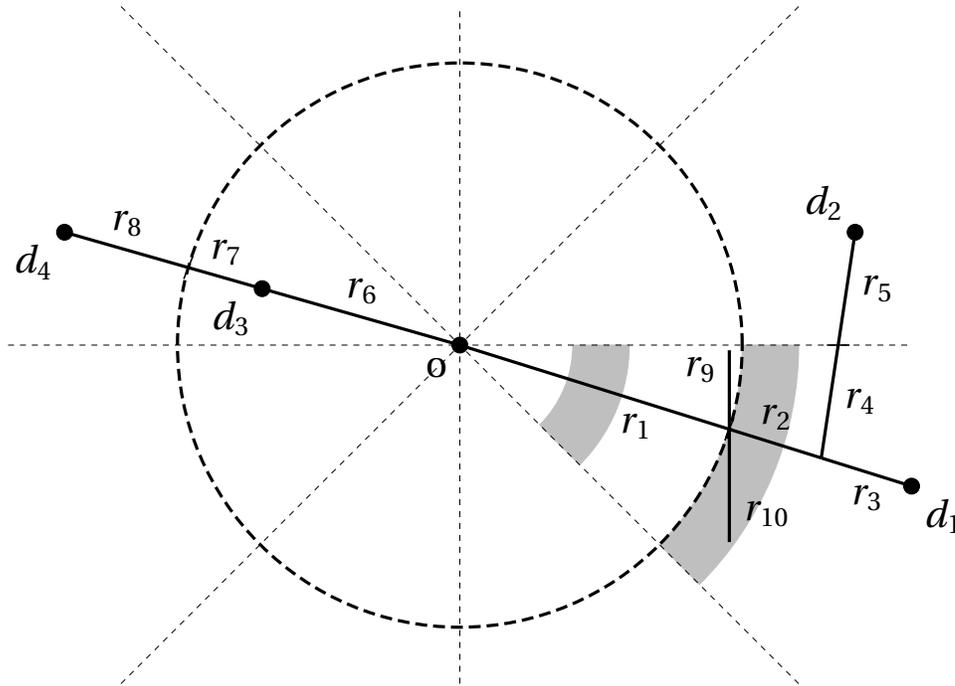
Notes: Total road network is defined circa 2004 based on Nelson and Deichmann (2004). See Appendix for details.

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



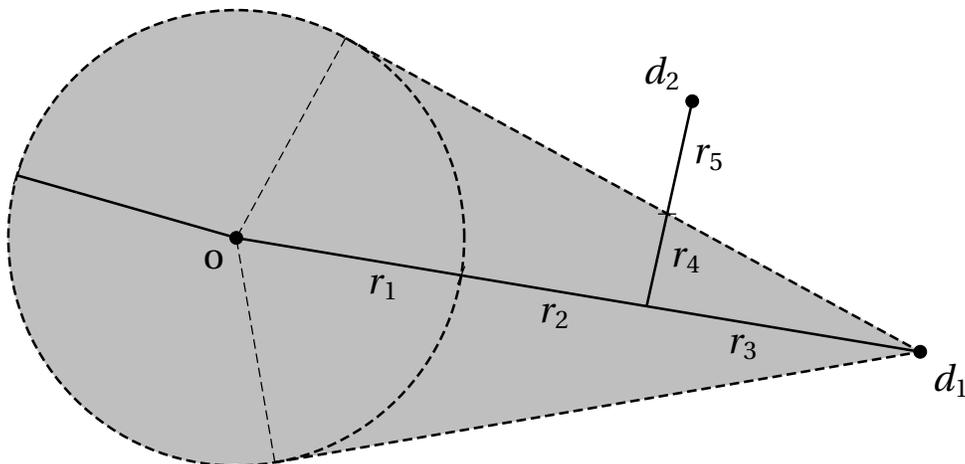
Notes: Subfigures 3a and 3b show the cities (defined as localities with population over 10,000 inh.) in our main 39-country sample in 1960 (N = 418) and in 2010 (N = 2,859) respectively. See Appendix for details.

Figure 4: Identification strategies



*Notes:* For each city  $o$ , our chief identification strategy consists of instrumenting its change in market access to other cities (here,  $d_1$ - $d_4$ ) with the change in market access due to road changes faraway (i.e., outside the dashed circle) while fixing population of the other cities at their initial levels. The exclusion restriction implies that roads changes outside (here,  $r_2, r_3, r_4, r_5, r_8$  and  $r_{10}$ ) are exogenous. Robustness checks exclude cities in which inner and outer roads, defined as passing through the cell-based equivalent of the inner and outer gray bars in the figure) are upgraded in the same octant in the same decade (e.g.  $r_1$  and  $r_2$ ; co-investment); cities where an outer road ( $r_2$ ) is built in an octant where an inner road ( $r_1$ ) is already paved or improved (radial extension outward); and cities where an inner road is upgraded in the decade after an outer road in the same octant is improved. See Section 2. for details.

Figure 5: Convex hull identification strategy



*Notes:* In the convex hull identification strategy, all roads built in the region of the path from city  $o$  to the largest city of 100,000 ( $d_1$ ) are excluded. Specifically, that region is the gray area, defined as the convex hull of city  $o$ 's exclusion circle and  $d_1$ . Here, any changes in  $r_1, r_2, r_3$ , and  $r_4$  are excluded, leaving only changes in  $r_5$ .

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

| Dependent Variable:                             | $(\Delta_{t-10}^t \ln \text{Urban Population})/100$ |                   |                   |                   |                   |
|---|---|-------------------|-------------------|-------------------|-------------------|
|   | (1)   | (2)               | (3)               | (4)               | (5)               |
| $\Delta_{t-10}^t \ln \text{Market Access}$      | 1.34***<br>[0.32]                                   | 1.27***<br>[0.32] | 1.58***<br>[0.35] | 1.63***<br>[0.44] | 1.50***<br>[0.38] |
| $\Delta_{t-20}^{t-10} \ln \text{Market Access}$ |   | 1.02***<br>[0.24] | 1.23***<br>[0.26] | 1.55***<br>[0.34] | 1.11***<br>[0.30] |
| $\Delta_{t-30}^{t-20} \ln \text{Market Access}$ |   |                   | 0.81***<br>[0.23] | 0.89***<br>[0.29] | 0.79***<br>[0.27] |
| $\Delta_{t-40}^{t-30} \ln \text{Market Access}$ |   |                   |                   | 0.27<br>[0.23]    |                   |
| $\Delta_t^{t+10} \ln \text{Market Access}$      |   |                   |                   |                   | 0.67<br>[0.49]    |
| Overall Effect<br>( $t - 40$ to $t$ )           | 1.34***<br>[0.32]                                   | 2.29***<br>[0.45] | 3.62***<br>[0.59] | 4.33***<br>[0.83] | 3.40***<br>[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  $(\Delta_{t-10}^t \ln \text{urban population})/100$  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,  $\ln \text{urban pop}_{t-10}$ , 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:                             | $(\Delta_{t-10}^t \ln \text{Urban Population})/100$ |                   |                    |                    |
|---|---|-------------------|--------------------|--------------------|
|   | OLS   | IV: Exclude 5     | IV: Exclude 10     | IV: Exclude 15     |
|   | (1)   | (2)               | (3)                | (4)                |
| $\Delta_{t-10}^t \ln \text{Market Access}$      | 1.58***<br>[0.35]                                   | 2.98***<br>[1.00] | 4.59***<br>[1.76]  | 5.75*<br>[2.95]    |
| $\Delta_{t-20}^{t-10} \ln \text{Market Access}$ | 1.23***<br>[0.26]                                   | 3.28***<br>[0.87] | 5.76***<br>[1.59]  | 7.34***<br>[2.46]  |
| $\Delta_{t-30}^{t-20} \ln \text{Market Access}$ | 0.81***<br>[0.23]                                   | 2.57***<br>[0.86] | 3.38**<br>[1.39]   | 4.60**<br>[1.95]   |
| Overall Effect<br>( $t - 30$ to $t$ )           | 3.62***<br>[0.59]                                   | 8.83***<br>[1.89] | 13.74***<br>[3.31] | 17.69***<br>[4.64] |
| First stage Kleibergen-Paap F                   |   | 114.00            | 41.86              | 17.41              |

Notes: Each column is a separate regression of  $(\Delta_{t-10}^t \ln \text{urban population})/100$  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  $\Delta \ln \text{Market Access}$  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 for the Identification Strategy**

|   | OLS<br>(1)        | IV:Excl. 5<br>(2)  | IV:Excl. 10<br>(3) | IV:Excl. 15<br>(4) |
|---|-------------------|--------------------|--------------------|--------------------|
| (1) Baseline<br>(N=4,725; F: _; 114.0; 41.9; 17.4)                                  | 3.62***<br>[0.59] | 8.83***<br>[1.89]  | 13.74***<br>[3.31] | 17.69***<br>[4.64] |
| (2) Co-Investment: Inner: 2-3, Outer: 5-6<br>(N=1,890; F: _; 17.4; _; _)            | 4.47***<br>[1.17] | 10.56**<br>[4.93]  |                    |                    |
| (3) Co-Inv.: Inner: 2-3, Outer: 10-11<br>(N=2,197; F: _; 30.0; 10.6; _)             | 4.29***<br>[1.03] | 10.82***<br>[3.70] | 14.21**<br>[6.34]  |                    |
| (4) Co-Inv.: Inner: 2-3, Outer: 15-16<br>(N=2,260; F: _; 45.6; 13.3; 3.9)           | 3.77***<br>[0.99] | 10.63***<br>[3.57] | 15.81**<br>[7.09]  | 21.31**<br>[10.60] |
| (5) Radial Ext. Out.: Inner: 2-3, Outer: 5-6<br>(N=2,098; F: _; 102.4; _; _)        | 4.23***<br>[0.89] | 9.59***<br>[2.55]  |                    |                    |
| (6) Radial Ext. Out.: Inner: 2-3, Outer: 10-11<br>(N=1,804; F: _; 78.0; 16.5; _)    | 3.55***<br>[0.91] | 8.97***<br>[2.45]  | 13.33***<br>[4.67] |                    |
| (7) Radial Ext. Out.: Inner: 2-3, Outer: 15-16<br>(N=1,603; F: _; 66.0; 17.1; 6.7)  | 4.22***<br>[0.94] | 10.09***<br>[2.62] | 14.81***<br>[4.78] | 19.65***<br>[6.90] |
| (8) Radial Ext. In.: Inner: 2-3, Outer: 5-6<br>(N=2,168; F: _; 32.8; _; _)          | 4.07***<br>[0.93] | 9.97***<br>[3.26]  |                    |                    |
| (9) Radial Ext. In.: Inner: 2-3, Outer: 10-11<br>(N=1,974; F: _; 30.6; 26.8; _)     | 4.43***<br>[1.08] | 10.24***<br>[3.44] | 15.00**<br>[6.32]  |                    |
| (10) Radial Ext. In.: Inner: 2-3, Outer: 15-16<br>(N=1,877; F: _; 39.4; 25.8; 6.4)  | 3.74***<br>[1.22] | 11.49***<br>[3.55] | 16.16***<br>[6.14] | 19.39**<br>[8.60]  |
| (11) Excl. Changes Convex Hull Nearest 100k+<br>(N=4,725; F: _; 63.8; 24.2; 13.7)   |                   | 8.38***<br>[1.87]  | 13.01***<br>[2.97] | 15.09***<br>[4.10] |
| (12) Excl. Transcontinental Road Changes<br>(N=4,725; F: _; 51.9; 21.6; 8.7)        |                   | 8.91***<br>[2.34]  | 13.65***<br>[3.91] | 15.74***<br>[6.03] |
| (13) Excl. National, Regional & Top 5 Cities<br>(N=3,801; F: _; 119.8; 18.5; 9.4)   | 3.59***<br>[0.72] | 8.41***<br>[2.39]  | 14.18***<br>[4.99] | 17.28**<br>[7.36]  |
| (14) Excl. ≤100 km from Any Mine<br>(N=3,202; F: _; 97.6; 24.0; 14.0)               | 3.81***<br>[0.72] | 9.76***<br>[2.37]  | 15.97***<br>[4.38] | 21.95***<br>[6.31] |
| (15) Excl. ≤100 km from Cash Crop Cells<br>(N=4,606; F: _; 115.8; 42.2; 16.3)       | 3.80***<br>[0.60] | 9.15***<br>[1.90]  | 13.83***<br>[3.31] | 17.72***<br>[4.63] |
| (16) Excl. ≤100 km from President's Origin<br>(N=3,032; F: _; 95.7; 22.5; 8.6)      | 3.93***<br>[0.75] | 10.29***<br>[2.22] | 17.58***<br>[4.23] | 24.54***<br>[6.84] |
| (17) Incl. All City-Level Controls<br>(N=4,725; F: _; 105.5; 28.1; 12.2)            | 3.10***<br>[0.58] | 8.44***<br>[1.98]  | 13.52***<br>[3.88] | 19.19***<br>[5.86] |
| (18) Excl. > 75th Pctile Dist. to Nearest City<br>(N=3,543; F: _; 27.5; 5.6; 2.3)   | 3.10***<br>[0.64] | 7.93***<br>[2.11]  | 10.38***<br>[2.48] | 9.40***<br>[2.61]  |
| (19) Excl. > 75th Avg. Dist. between Cities<br>(N=3,592; F: _; 28.9; 11.7; 7.4)     | 3.18***<br>[0.66] | 6.26***<br>[2.14]  | 9.86***<br>[3.34]  | 12.00***<br>[4.35] |
| (20) Fix Population to 1960 in IVs<br>(N=4,723; F: _; 68.6; 26.3; 6.6)              |                   | 7.56***<br>[1.62]  | 11.96***<br>[2.67] | 16.88***<br>[4.15] |
| (21) Fix Pop. to 1960 in IVs & Drop 1980s<br>(N=3,629; F: _; 26.8; 7.0; 3.0)        | 4.04***<br>[0.76] | 6.59***<br>[2.09]  | 10.66***<br>[3.51] | 16.31***<br>[5.91] |
| (22) Fix Pop. to 1960 in Market Access (MA)<br>(N=4,723; F: _; 233.0; 70.1; 23.7)   | 3.18***<br>[1.03] | 9.80***<br>[2.05]  | 16.67***<br>[3.55] | 25.40***<br>[5.62] |
| (23) Fix Pop. to 1960 in MA & Drop 1980s<br>(N=3,629; F: _; 235.2; 66.9; 23.3)      | 2.41*<br>[1.39]   | 8.38***<br>[2.50]  | 15.87***<br>[4.45] | 26.97***<br>[7.36] |
| (24) Fix Pop. to t-10 in Market Access (MA)<br>(N=4,725; F: _; 151.6; 47.1; 22.8)   | 3.41***<br>[0.95] | 10.34***<br>[2.23] | 15.72***<br>[3.80] | 20.96***<br>[5.74] |
| (25) Ctrls. Reg. Growth (Prov., Ethnic, Circle)<br>(N=3,498; F: _; 32.0; 12.5; 5.4) | 2.57***<br>[0.68] | 6.30***<br>[2.07]  | 7.40***<br>[3.14]  | 8.51***<br>[4.71]  |

*Notes:* This table is structured like Table 2 but only reports the overall effect. Robust SEs, clustered by cell, are in brackets. \*, \*\*, \*\*\* denote significance at the ten, five, and one percent level, respectively.

**Table 4: Robustness Checks: Specification, Measurement Error, and Mechanisms**

|   | OLS<br>(1)        | IV:Excl. 5<br>(2)   | IV:Excl. 10<br>(3) | IV:Excl. 15<br>(4) |
|---|-------------------|---------------------|--------------------|--------------------|
| (1) Baseline<br>(N=4,725; F: _; 114.0; 41.9; 17.4)  | 3.62***<br>[0.59] | 8.83***<br>[1.89]   | 13.74***<br>[3.31] | 17.69***<br>[4.64] |
| (2) No Pop.<br>(N=4,725; F: _; 114.0; 41.1; 17.4)   | 3.67***<br>[0.59] | 8.82***<br>[1.90]   | 13.75***<br>[3.32] | 17.77***<br>[4.69] |
| (3) No Pop. & 3rd Order Poly. Lon.Lat.<br>(N=4,725; F: _; 55.6; 18.2; 9.9)  | 3.54***<br>[0.58] | 7.73***<br>[1.83]   | 11.52***<br>[3.04] | 13.89***<br>[3.98] |
| (4) $\Delta_{t-30}^t \ln(\text{Market Access})$<br>(N=4,725; F: _; 472.9; 157.4; 88.9)  | 1.11***<br>[0.30] | 2.52***<br>[0.69]   | 3.81***<br>[1.11]  | 4.76***<br>[1.58]  |
| (5) Reduced-Form Effects of the IVs<br>(N=4,725; F: _; _; _)  | –<br>–            | 13.12***<br>[2.85]  | 19.74***<br>[5.20] | 23.33***<br>[8.56] |
| (6) Obs. Weighted by Initial Pop. t-10<br>(N=4,725; F: _; 65.2; 30.6; 8.6)  | 2.87***<br>[0.84] | 7.95***<br>[2.52]   | 10.58**<br>[4.16]  | 14.09**<br>[5.87]  |
| (7) SE (Cluster Country)<br>(N=4,725; F: _; 59.6; 40.4; 13.8)   | 3.62***<br>[0.88] | 8.83***<br>[2.53]   | 13.74***<br>[3.50] | 17.69***<br>[5.32] |
| (8) Balanced Sample (Always $\geq 10,000$ )<br>(N=3,264; F: _; 98.5; 35.5; 14.2)  | 3.35***<br>[0.64] | 7.76***<br>[1.83]   | 11.61***<br>[3.02] | 14.22***<br>[4.34] |
| (9) Population threshold = 15,000<br>(N=3,254; F: _; 86.2; 34.4; 12.9)  | 3.22***<br>[0.69] | 9.81***<br>[2.28]   | 15.34***<br>[3.64] | 19.94***<br>[5.27] |
| (10) Population threshold = 20,000<br>(N=2,471; F: _; 74.4; 26.5; 8.3)  | 2.71***<br>[0.72] | 8.02***<br>[2.58]   | 13.63***<br>[4.38] | 18.42***<br>[7.24] |
| (11) Pop. Estimate < 10,000 Available<br>(N=7,369; F: _; 122.3; 44.9; 27.0)   | 5.22***<br>[0.68] | 7.73***<br>[2.01]   | 9.99***<br>[2.82]  | 11.24***<br>[3.52] |
| (12) Pop. Est. < 10,000 Avail. One Prev. Year<br>(N=6,164; F: _; 105.0; 36.3; 20.5)   | 4.17***<br>[0.67] | 6.86***<br>[1.94]   | 10.37***<br>[3.07] | 12.48***<br>[3.99] |
| (13) Drop 1980s<br>(N=3,631; F: _; 48.3; 25.0; 10.3)  | 4.04***<br>[0.76] | 8.98***<br>[2.35]   | 14.95***<br>[4.36] | 21.41***<br>[6.91] |
| (14) Drop 2000s<br>(N=2,607; F: _; 99.7; 34.3; 13.7)  | 3.36***<br>[0.65] | 7.35***<br>[1.96]   | 10.84***<br>[3.22] | 10.94***<br>[4.16] |
| (15) 2 Censuses for Both Start & End Year<br>(N=3,414; F: _; 50.0; 26.1; 11.5)  | 3.31***<br>[0.66] | 10.97***<br>[2.59]  | 15.23***<br>[4.26] | 16.74***<br>[6.05] |
| (16) Excl. if Both Start & End $\geq 5$ Yrs from Source<br>(N=4,430; F: _; 122.4; 33.8; 16.7)                                     | 3.61***<br>[0.60] | 8.95***<br>[1.95]   | 13.92***<br>[3.39] | 17.40***<br>[4.72] |
| (17) Excl. if Start or End $\geq 5$ Yrs from Source<br>(N=1,711; F: _; 55.6; 18.2; 9.9)   | 1.41<br>[0.86]    | 6.78***<br>[2.58]   | 8.72**<br>[4.23]   | 6.41<br>[5.60]     |
| (18) Excl. if > Mean Diff. Official Road Data<br>(N=2,246; F: _; 75.3; 19.3; 8.3)   | 3.90***<br>[0.83] | 11.70***<br>[2.56]  | 18.91***<br>[4.97] | 24.91***<br>[7.11] |
| (19) Excl. if > Mean Diff. or Missing<br>(N=1,032; F: _; 13.5; 3.6; 3.1)  | 1.59<br>[1.25]    | 8.07**<br>[3.62]    | 20.23***<br>[7.78] | 22.66**<br>[9.34]  |
| (20) Excl. if Country > Mean Diff. Official<br>(N=3,659; F: _; 90.6; 31.1; 13.6)  | 4.11***<br>[0.70] | 10.23***<br>[2.28]  | 16.92***<br>[4.19] | 22.45***<br>[5.94] |
| (21) Excl. if Country > Mean Diff. or Missing<br>(N=3,258; F: _; 45.8; 25.7; 11.7)  | 4.66***<br>[0.76] | 11.71***<br>[2.79]  | 18.20***<br>[4.38] | 24.06***<br>[5.94] |
| (22) $\Delta_{t-10}^t \ln(\text{Light Int.})$ , Ctrl $\Delta_{t-10}^t \ln(\text{Urb. Pop.})$<br>(N=3,591; F: _; 48.4; 25.4; 10.0) | 2.87<br>[4.16]    | 37.47***<br>[11.50] | 48.25**<br>[17.92] | 64.56**<br>[28.01] |
| (23) $\Delta_{t-10}^t \ln(\text{Light Intensity Per Capita})$<br>(N=3,591; F: _; 48.1; 24.9; 10.3)                                | 1.58<br>[4.31]    | 33.34***<br>[11.80] | 37.24**<br>[17.95] | 49.09*<br>[27.92]  |

*Notes:* This table is structured like Table 2 but only reports the overall effect. Robust SEs, clustered by cell, are in brackets. \*, \*\*, \*\*\* denote significance at the ten, five, and one percent level, respectively.

**Table 5: Effect of Market Access on Night Lights**

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

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

**Table 6: Population Reallocation across Cities**

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

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

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