THE AVERAGE AND HETEROGENEOUS EFFECTS OF TRANSPORTATION INVESTMENTS:
EVIDENCE FROM SUB-SAHARAN AFRICA 1960-2010

Remi Jedwab
Adam Storeygard

The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan
Africa 1960-2010
Remi Jedwab and Adam Storeygard
NBER Working Paper No. 27670
August 2020, Revised March 2021
JEL No. F15,F16,O18,O20,R11,R12,R4

**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 a 30-year elasticity of city population with respect to market access of about 0.08–0.13. 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 besides international ports matters little. Additional evidence points suggestively to rural-urban migration as the primary source of this population increase, though we cannot fully rule out natural increase or reallocation across cities.

Remi Jedwab
George Washington University
Department of Economics #310
2115 G Street, NW, Washington, DC 20052
jedwab@gwu.edu

Adam Storeygard
Department of Economics
Tufts University
8 Upper Campus Road
Medford, MA 02115
and NBER
Adam.storeygard@tufts.edu

A data appendix is available at http://www.nber.org/data-appendix/w27670

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 0.8–1.3% 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 similar across the first three decades of road-building, subsequently falling. Our results then suggest that effects are larger for more isolated cities, and market access changes to domestic rather than foreign cities, and weaker in politically favored and more agriculturally suitable areas. We provide suggestive evidence that these effects stem primarily from rural-to-urban migration, as opposed to natural increase or reallocation of population across cities.

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).[1] 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

---

[1]According to our data, there are only 3,500 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).

and ethnic groups within African countries and contribute to their high levels of spatial inequality and the persistent weakness of their states (Herbst, 2000). Today, sub-Saharan Africa is the region with 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.[2] 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). Some donors 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. Faber, 2014; Storeygard, 2016; Jedwab and Moradi, 2016; Donaldson and Hornbeck, 2016; Donaldson, 2018, for a comprehensive overview of the literature, see Redding and Turner 2015). More generally, a large literature has looked at how market access affects the growth of cities among other units of analysis (e.g. Redding and Sturm, 2008; Jaworski and Kitchens, 2019).[3] Another large literature has looked at the effect of large highway projects

---

[2]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).

[3]Much of this recent literature has estimated the effects of transport costs and infrastructure investment within a general equilibrium trade model. This is not feasible in our environment, where data are substantially less available, even compared to the middle-income developing

on a variety of outcomes (e.g. Ghani et al., 2016). Finally, a smaller literature has emphasized the specific role of road quality, which is the main source of variation in this work (e.g. Casaburi et al., 2013; Gertler et al., 2015; Asher and Novosad, 2020). This paper makes several contributions to this literature.[4]

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 nearly 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., 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, 2020). 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 (2020) cost 1.8% of India's GDP. We also focus on cities, which represent a large share of overall economic activity.[5]

Our second contribution is methodological. We develop a novel identification strategy, relying on the variation in market access induced by roads built far

---

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

[4]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.

[5]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.

away. Our strategy is most closely related to Donaldson and Hornbeck (2016) and the contemporaneous work of Jaworski and Kitchens (2019), but we depart from their approach especially in further ruling out several potential channels for reverse causality.[6] Our strategy of isolating variation from non-local changes to 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). However, to our knowledge this strategy has been used rarely to study the effects of infrastructure, notable recent exceptions being Schlenker and Walker (2016) and Jaworski and Kitchens (2019).[7]

Third, we consider heterogeneous effects along multiple dimensions. These 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), 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 (2020), who find that roads cause outmigration from villages with low agricultural productivity. (iii) Cities less

---

[6]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, including least-cost construction routes, 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 details 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.

[7]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.

likely to be politically favored see bigger impacts. This is consistent with the literature documenting political motivations in the allocation of roads (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 outsize role, while access to foreign cities in neighboring countries does not, consistent with the overseas nature of much African trade, and perhaps not surprising given the non-cooperative nature of road-building across borders (Felbermayr and Tarasov, 2015). While the stronger role of access to world markets is in line with Fajgelbaum and Redding (2014) and Baum-Snow et al. (2019), the differential seems to be smaller in Africa, possibly due to oligopolistic intermediaries (Atkin and Donaldson, 2015).

Finally, our work also builds on the literature considering how cities in developing countries grow. Previous work on transport and city growth in Africa has emphasized the earlier railroad revolution (Jedwab and Moradi, 2016; Jedwab et al., 2017) 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.[8] City growth is the best available proxy for local economic development in our context, and an object of interest in its own right.[9] Previous work has

---

[8]Fay and Opal (2000); Gollin, Jedwab and Vollrath (2016); Jedwab and Vollrath (2019).

[9]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. Henderson et al. (2017) use information on populations for subnational units of 89 censuses in 29 countries, but these data are not consistently available back to the 1960s for most countries. 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 and have limited geographic information before the mid-1990s (we use their retrospective data on fertility and mortality for a limited sample of cities below). For a subsample of decades, we consider night lights, available

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 construct a new spatial dataset of road surface and city population every 10 years between 1960 and 2010 for the 42 countries of mainland sub-Saharan Africa (c. 2010). Appendix Sections A.1–A.6 provide further details on the data.

### 1.1. Roads, 1960-2010

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

The Michelin maps report highways and intercity paved and improved roads comprehensively, but their coverage of earthen roads is less complete, with some maps changes clearly due to coverage decisions 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. Figures 1a and 1b show roads circa 1960 and 2010.[10]

from 1992, as a proxy for city income, following Storeygard (2016).

[10]See Section A.1 in the Appendix for further details on the road data. Appendix Figure A.3

Michelin uses four sources to create the maps: (i) the previous Michelin map, (ii) government road censuses/maps, (iii) direct information from its tire stores across Africa, and (iv) correspondence from road users including truckers.[11] The latter two sources of information are especially important, and new to this literature.[12] Michelin has been producing maps since 1910, with its first map for West Africa appearing in 1938. As one of the largest tire companies in the world since the early 1970s (Rajan et al., 2000), unlike other organizations producing maps, Michelin has long maintained a large network of stores distributing its tires, in addition to its maps. Many truck drivers in Africa use both, and are in regular contact with this network. Because inaccurate characterization of road surface leads to delays or truck damage, truckers complain to the store managers when the information is inaccurate, and the store managers relay this information to Michelin cartographers. Michelin also focuses on road surfaces whereas other maps classify roads as primary/secondary or major/minor, which is less informative about road quality. We are unaware of another source of maps with similarly broad coverage over such a long period.

We believe that this process leads to generally consistent information across countries and time, but this does not mean that the evolution of every road segment is perfectly characterized. This has several implications. First, changing conditions may be reflected on maps with a lag. The lag is unlikely to be long because: (i) Michelin dealers collect data on ongoing projects and their maps are intended to reflect the year a road will open and (ii) periods between maps are generally short. Second, Michelin's network is more sparse in some countries and periods. Country-year fixed effects should ameliorate the effect of this to some extent. Coverage of the early 1960s is more limited; as we show, results are robust to excluding the decade most likely to be 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

---

shows the Michelin map for Sierra Leone in 1969 and the associated GIS map.

[11]This paragraph is based on our discussions with Michelin employees.

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

may underestimate recent changes. Results are robust to excluding the 2000s.

## 1.2.    City Location and Population, 1960-2010

We obtained location and decadal population estimates of cities in 33 countries from *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa*.[13] They 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 c. 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 c. 2000 using satellite imagery. If two distinct cities in 1970 later 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 comprehensive database for 6 southern African countries not in Africapolis, and a more limited database for the remaining three countries of mainland sub-Saharan Africa: South Africa, Swaziland, and Lesotho. Comparable cities data covering the range of populations and years we use are not available for these three countries, so they are not included in our analysis sample. 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. 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 correct some locations based on Google Earth, GeoNet, and Wikipedia, aggregating multiple administrative cities into one agglomeration using more recent satellite imagery from Google Earth.[14]

The raw population data sources used for each of the 237 available country-

---

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

[14]Jedwab and Moradi (2016) use the Africapolis data and similar but not identical data for the other six countries. City data details are in Appendix Section A.2.

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 decadal 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.[15]

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.[16] We show below that results are similar when we include estimates below 10,000 where available.

### 1.3.   Data integration and unit of analysis

Our cross-sectional unit of analysis is a 0.1 by 0.1 degree grid square ($\approx$ 11 x 11 km). There are 199,814 such cells in our 42 countries. Using these units dramatically simplifies computation compared to the full vector road network, and avoids problems due to missing topological information, concerning which

---

[15]In our sample, the initial and final populations of a decade are both at least 5 years from a population data source for about 6% of country-decades. Alternatively, the initial *or* final populations of a decade are more than 5 years from a population data source for about 64% of country-decades. We thus investigate how results change when dropping these country-decades.

[16]Further, 98% of the 1,721 extensive margin cases are growth from below 10,000 to above it; the remaining 2% are declines to below it. Since we do not have data on the exact population of all city-decades with a population below 10,000, we ignore the extensive margin in our analysis.

segments connect to each other and which do not, in vector roads datasets.[17]
We assign to each grid square in each year a speed of travel for the fastest road
segment type falling in the grid square in the year, or a baseline speed if no
roads are present. We assume 80, 60, 40, 12, and 6 km/hour 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.[18] The
urban population of each cell is the population of the city whose location falls
within it, or in the small minority of cells with multiple cities, the sum of their
populations.[19] In our econometric analysis, we focus on the cells with a total city
population of at least 10,000 in two consecutive decades since 1960 in the 39 of
these countries with comprehensive cities data.

### 1.4.  Other data

We compile several additional datasets and assign them to cells: (i) navigable
rivers; (ii) the mean and standard deviation of altitude; (iii) average rainfall 1900–
1960; (iv) land suitability for food/cash/all crops today (assuming low input
levels and no irrigation); (v) open mines (including fields) 1960–2010 (N = 288);
(vi) the identity of national and provincial capitals in 1960 (N = 346) and 2010 (N =
481); (vii) 1960 province identifiers; (viii) railroad lines and when each was built;
(ix) 65 and 44 international ports in 1960 and 2005 respectively; (x) 466 airports
in 2007; (xi) 837 customs posts circa 2010; (xii) natural parks covering 26,252 cells
circa 2015; (xiii) historical ethnic group identifiers; (xiv) the locality of origin
and ethnicity of the 189 heads of state 1960–2010; (xv) night lights 1992–2010;
(xvi) conflict locations 1990–2010; (xvii) 1980–2009 fertility and infant mortality
histories from Demographic and Health Surveys (DHS) for 16 countries.

   We also obtained country-year-level data on: (i) population, per capita GDP,

---

[17]Mean road length for 90,628 cells with a road in Nelson and Deichmann (2004) is 12.1 km.
This suggests that our 11 x 11 km cells have on average about 1 road crossing them fully.

[18]We show below that results are similar if we use alternative speeds.

[19]In 95% of cell-decades in our estimation sample, a cell-decade contains a single city-decade,
so this summing only affects the remaining 5% of cell-decades. Thus, conceptually, we think of
these cells as cities, and sometimes refer to cells containing a city as "cities" for ease of exposition.
We effectively assume each city falls entirely in one cell, but verify below that results are similar
if we drop the largest cities, which are the most likely to span multiple cells.

polity score, refugee count, and total fertility, infant mortality, crude birth and crude death rates; and (ii) whether the country was still a colony, experienced an international/civil war, or suffered a multi-year drought in each decade. Appendix Section A.5 provides details.

## 1.5.  Aggregate Patterns in Road Building and Urban Growth

Appendix Figure A.4 shows aggregate lengths of highways and paved and improved roads over time. 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 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.[20]

Figures 1c and 1d 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

---

[20]For example, the Kenyan government has invested heavily in rebuilding the Mombasa-Nairobi road (Burgess et al., 2015). Jedwab and Storeygard (2019) discuss country-level patterns in greater detail, and report evidence that our data are consistent with other sources at the country level. Specifically, in an unbalanced panel of 609 country-years between 1960 and 2002 that we report in common with Canning and Farahani (2007), our two independent measures of paved roads are correlated at 0.7.

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.

## 2.    Empirical Methods

We study how increased market access to other cities affects city population growth in 2,126 urban cells in 39 sub-Saharan African countries sampled every ten years 1960–2010.[21] 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.[22] Origin cell $o$'s *market access* (*MA*) in year $t$, is $MA_{ot} = \sum_{d \neq o} P_{dt} \tau_{odt}^{-\theta}$, where $P$ is urban population, $d$ indexes destination cells, $\tau_{odt}$ is travel time from cell $o$ to cell $d$, 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. Combining these estimates, we apply a baseline value of $1.27 \times 3 = 3.8$ for $\theta$ to assumed travel times and consider alternatives below. In our analysis, we focus on how changes in the road network affect travel times.[23]

---

[21]These 2,126 cells are the subset of the 2,789 noted above that have an urban population over 10,000 in at least two consecutive decades, so that changes are well-defined.

[22]The model of Donaldson and Hornbeck (2016) includes both firm and consumer market access (MA) terms, similarly to how Redding and Venables (2004) define supplier and market access. However, these two terms cannot be separately identified in their model, so our empirical specification, like theirs, includes a single MA term that reflects benefits to firms and consumers.

[23]Appendix Table A.9 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).

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.[24]  When a map is not available for a given year, we interpolate speeds between the closest map years before and after.[25]

## 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:

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

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:

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

We further control for initial log population in the first-difference specification to account for any divergence (convergence) if large cities grow faster (slower) than small cities, due to local increasing returns/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 two-dimensional spatial polynomial in longitude and latitude interacted with year fixed effects to control for unobservables correlated across space within country-decades. This is

---

Appendix Figure A.8 shows how results vary with respect to the choice of $\theta$.

[24]See Appendix Section A.3 for details. Appendix Figure A.5 shows how we obtain market access changes for Sierra Leone between 1970 and 1980.  Appendix Figure A.6 shows the change in market access between 1960 and 2010 for the 187,900 cells of the 39 sample countries.

[25]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 show later that results are similar when dropping the decade plausibly affected by 1960s road-building.

equivalent to the main specification of Donaldson and Hornbeck (2016, eq. (13)), using population as an outcome instead of land rents.[26] We will nonetheless show that the results are similar if no additional controls are added except for the country-decade fixed effects. Finally, we cluster standard errors by province (using province boundaries c. 1960). We will show that estimated standard errors clustered by country are usually larger but modestly so and 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:

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

a log-transformed spatial lag specification, where the log is applied element-wise.[27] Then, using (3), (2) becomes:

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

Changes in market access come from either changes in the population of other cities (weighted by travel times $T$ in $t$) or changes in travel times $T$ to these other cities (weighted by the population of cities 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

---

[26]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.

[27]An explicit matrix formulation appears in Appendix Section A.6.

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 likely reduce the bias due to measurement error.[28]

**Instrument fixing population.** We first propose 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:

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

**Instrument also excluding local road changes.** An important problem with the instrument in (5) 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 from $o$ 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:

---

[28]Building on a prominent recent literature studying shift-share identification strategies, Borusyak and Hull (2020) highlight a novel form of omitted variable bias that may affect this approach. Specifically, they show that in principle, even randomly placed roads can affect market access systematically due to underlying geography. However, addressing this potential bias requires variation such as infrastructure plans or policy discontinuities that, as noted above, are not available in our context (or that of Donaldson and Hornbeck (2016) and most others in this literature).

$$\Delta_R^{out,j} \ln MA_{ot} = \ln\left(\sum_{d:\delta(d,o)\geq j} P_{d,\mathbf{t\text{-}10}}\tau_{od,\mathbf{t}}^{-\theta} + \sum_{d:0<\delta(d,o)<j} P_{d,\mathbf{t\text{-}10}}\tau_{od,\mathbf{t\text{-}10}}^{-\theta}\right) - \ln\left(\sum_{d\neq o} P_{d,\mathbf{t\text{-}10}}\tau_{od,\mathbf{t\text{-}10}}^{-\theta}\right) \quad (6)$$

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 2a illustrates 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.

$\Delta_R^{out,j} \ln MA_{ot}$ (or more simply IV $j$) 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. Since we generally include two lags of market access, we use two lags of the instrument as well, for a total of three instruments for three endogenous variables.

**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 2a, 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 2a, 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 2a, 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 2a.

*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 2a, 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 2a, this means dropping city $o$ in year $t$ if in any decade between $t-30$ and $t$, $r_1$ ($r_6$) 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 at least 100 thousand people in $t-10$. Formally, this is the convex hull of the exclusion circle and that nearest large city. Figure 2b shows this exclusion zone 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 (and 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, and 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 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 $d$, these 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.[29] Alternatively, we use population in 1960 (or t-10) to define not just the instruments but also the main change in market access variable: $MA_{ot} =$

---

[29]The second lag of the change in market access for the period 1980–1990 already uses 1960 population in (6). We thus run this test on a sample dropping the 1980s as well.

$\sum\limits_{d \neq o} P_{d,1960} \tau_{odt}^{-\theta}$. 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.

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 Panel A 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 multiplied by 100, so that coefficient can be interpreted as elasticities multiplied by 100. 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 10% increase in market access in each decade, across these three decades, is thus over 0.3%.[30] 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 1980–2010.[31] Appendix Table A.3 contains descriptive statistics for this main sample.

---

[30]To see this, note that our results imply that a change in market access between $t - 30$ and $t - 20$ has impacts in three decades: $t - 30$ to $t - 20$, $t - 20$ to $t - 10$, and $t - 10$ to $t$, so the overall 30-year elasticity is the sum of the contemporaneous elasticity and the two lags.

[31]The 2-lag specification has lower Aikike and Bayesian information criteria than the 3-lag one.

In column 5, we investigate reverse causality by adding a lead to the column 3 specification; it is small and insignificant, though as it is imprecisely measured, it is not statistically distinguishable from either of the included lags. The last row of coefficients in Table 1 Panel A 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 3.4–4.2%.

## 3.2. Instrumental Variables (IV) Results

**IV Results.** Table 1 Panel B reports the results of the IV specifications where the instruments defined in equation (6) are used to disentangle the causal effect of market access due to roads on city growth.[32] Columns 1–3 instrument with changes in market access due only to roads built more than 5, 10 and 15 cells away, thus excluding changes due to roads built nearby as well as recent city growth everywhere. As expected, the instrument is stronger at lower radii, because it includes road changes closer to the city. Effects are larger than in the OLS specification, with 30-year elasticities between 8.3% and 16.5%, increasing with the radius. Alternatively, these results imply that a one standard deviation increase in market access growth is associated with a 0.43-0.85 standard deviation increase in city population growth. The first lag is larger than the contemporaneous term for IV10 and IV15, though not significantly so. This may be because the contemporaneous term includes roads build late in a decade, with little time to have an impact. Moreover, Section 3.4. shows that population effects are slower to develop than effects on night lights.

If we add a lead, instrumented analogously, it is not significantly different from zero. Thirty-year elasticities are somewhat smaller, between 6.6% and 9.1% (Appendix Table A.5, Panel A). However, this specification requires dropping the 2000s, and doing that alone has a similar effect (Appendix Table A.8, row 14). A third lag, entered separately, is also small and insignificant, and changes the estimated overall effect little, though its inclusion does result in a weaker instrument and a smaller and less precisely estimated contemporaneous term

---

[32]The corresponding first stage regressions are reported in Appendix Table A.4.

(Appendix Table A.5 Panel B).

**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 -0.06***. Using log distance to the capital and the two largest cities in 1960 as an alternative measure of initial remoteness gives similar results.

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, we expect that IV strategies measure local average treatment effects that are no larger than the average effect.

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. We thus cannot know the extent to which our instruments reduce measurement error.

Finally, since larger exclusion circles remove more local road-building, instruments based on them are more likely to satisfy the exclusion restriction, so IV15 should be more excludable than IV5. However, the higher-radius instruments are also weaker, especially in several robustness checks, which means any violation of the exclusion restriction creates a larger bias. Thus, to be conservative, we emphasize the IV5–IV10 results.

In summary, we expect that the IV coefficients are higher than the OLS most likely due to omitted variable bias and potentially due to measurement error, though it is possible that for the weakest instruments, especially IV15, the large magnitudes reflect violations of the exclusion restriction.

**Magnitude.** We find a 30-year elasticity of 0.08–0.13, with a slightly wider range in robustness checks below. This implies that a 10% increase in market access will increase city populations by 0.8 to 1.3%. How might one increase market access by 10%, given our assumptions? In a symmetric context where all roads are unpaved, paving a random 3.1% of them, or improving 3.5%, would increase market access by 10%. In a context where all roads are improved, paving 7.4% of them would similarly increase MA by 10%.

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 estimate discrete effects of rail construction on agricultural land prices, which allows for a comparison to a wider literature. As noted by Redding and Turner (2015), these discrete effects 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 2, we investigate how results change when we account for: (i) selected non-local road changes; (ii) growth hubs; and (iii) correlated regional growth. Rows are structured like Table 1 Panel B but only report overall 30-year effects. In many cases, the 15-cell instrument becomes somewhat weaker.

**Excluding selected non-local road changes.** Rows 1–3 exclude cities with any *co-investment, radial extension outward*, and *radial extension inward*, respectively, as defined above. In each case, we define the inner ring between 2 and 3 cells from the city, and the outer ring 15–16 cells away. Sample sizes fall by 40–65%, but results are generally consistent with the baseline. Row 4 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 5 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.[33]

**Dropping potential growth hubs.** Row 6 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

---

[33]Appendix Section A.8 and Appendix Table A.6 report several variants of these specifications.

construction. Instead, they rely on small cities, which were more likely to be connected incidentally. Results are similar to baseline.[34]

**Correlated regional growth.** In row 7, population is fixed at its 1960 level in constructing the instruments, and in row 8, in constructing the instrumented market access (*MA*) as well. Row 9 reports 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 ($\approx 90,000km^2$) is at the the 50th and 88th percentiles, respectively, of the province and ethnic area size distributions. In other words, 50% of provinces and 88% of ethnic areas are smaller that our largest exclusion zones. In row 9, 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 6.1–8.7%, and the IV10 and IV15 instruments considerably weaker.[35]

To control even more flexibly for any nearby trends, in row 10 we add province-year fixed effects, using provinces from 1960, the beginning of our sample period. There are 287 provinces and 804 province-years in our sample of 4725 observations, hence only about 6 observations per province-year on average, with considerable variation depending on the density of provinces across countries. Two hundred fifty of the 804 province-years have exactly one city-year, so they do not help to identify the effect of interest. Another 125 have

---

[34]Appendix Section A.8 and Appendix Table A.7 report specifications dropping several other sets of cities with features that might be driving road construction, including nearby national cities, isolation, mines, cash crops, presidential origins, ports, airports, border crossings and national parks and unusually large positive or negative changes in population or MA, and controlling for local conflict and neighboring country's growth, among other potentially correlated factors. Appendix Section A.8 and Appendix Table A.8 show that results are similar 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

[35]Appendix Section A.8 and Appendix Table A.6 report variants on these specifications.

two city-years. Thus, roughly half of the province-years, which provided roughly 10% of the degrees of freedom on the baseline specification now provide less than 3%. OLS and IV5 coefficients fall substantially, and IV10 and IV15 are again quite likely to be weak.

Given the variation in the way countries define provinces, we try an alternative strategy in row 11. Here we divide each country into four quadrants relative to the median latitude and longitude of the country, and control for quadrant-year fixed effects. There are 138 quadrants and 395 quadrant-years represented in the sample. OLS, IV5 and IV10 coefficients fall by a quarter to a third from baseline but remain precisely estimated. The IV15 instrument is again quite weak and its coefficient falls further.

**Other specification and sample checks.** Appendix Section A.9 and Appendix Table A.8 drop various sets of countries and country-years for which population and/or road data are likely to be measured with the most error. It also reports results for samples varying the city population threshold, including a fully balanced panel. Results are generally very similar to baseline.

In Appendix Section A.10 and Appendix Table A.9, we further consider alternative specifications and functional forms. We show the effects are broadly robust to: (i) removing controls; (ii) replacing the country-decade fixed effects with decade fixed effects; (iii) clustering by country; (iv) weighting observations by initial population; (v) using alternative speeds; (vi) allowing railroad travel; (vii) using alternative values of $\theta$; (viii) adding uniform costs of crossing borders; (ix) using iceberg costs with an example container value; (x) excluding sample countries bordering non-sample regions in South Africa, North Africa or the Arabian Peninsula, as their market access may be differentially underestimated. Defining market access changes due to paving alone and improving alone suggests that if anything market access changes due to improving have slightly larger effects (Appendix Table A.9). This difference is not statistically significant, and of course it is sensitive to the speed assigned to each surface.[36]

---

[36]As we discuss in the same Appendix Section and Table, regressing $\Delta_{t-10}^{t} \ln(\text{Population})$ on $\Delta_{t-30}^{t} \ln(\text{Market Access})$ results in a standardized effect that is about half as large as baseline.

Appendix Table A.10 considers alternative specifications that control for 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 an impact of either of these alternative measures, and little effect on the main urban market access coefficient, though both new variables are likely to be highly mismeasured.

### 3.4.   Effects on Night Lights

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

To explore this idea, we consider changes in night lights as a proxy for overall output.[37] The sample includes 3,591 observations, for the periods 1992–2000 and 2000–2010 only, because lights data begin in 1992. Appendix Table A.11 panel A shows that the 30-year overall effects on population for this restricted sample are similar to those for the full samples.

Table 3 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 1, in each IV specification, market access has substantially larger effects on lights than on population. The effect size is comparable to what Alder (2019) 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. Appendix Table A.11 panel B shows that the effect on night lights per capita is also only contemporaneous in IV estimates. In panel C, the effect on population controlling for contemporaneous night lights change are also substantially

---

[37] 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.

positive but limited to the two lags. 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.[38]

### 3.5. Channels: Induced Urbanization, Reallocation and Natural Increase

Results shown thus far do not allow us to distinguish 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*). In this section we provide evidence about the sources of this growth and discuss important limits of that evidence. Broadly speaking, we find some evidence against reallocation and evidence for a small role of natural increase. This suggests to us that to the extent we can rely on this evidence, the residual category of induced urbanization is likely to be responsible for the plurality of the effect.

Our first piece of evidence against reallocation is that only 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 increased only to 28% by 2010. This means that the pool of rural potential migrants was always 3-4 times as large as the pool of urban potential migrants.[39]

Table 4 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 $t-30$. With low urbanization rates, they 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,

---

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

[39]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.

restricting to the bottom quartile (<10%) reduces magnitudes somewhat more, though in the case of IV15, this may be driven by instrument weakness in a small sample. Furthermore, using only the lowest-urbanization decile (<7%) of countries in row 4, results are more similar to the full sample (though again the IV15 and now IV10 instruments are quite weak).

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

From the perspective of central place theory (Christaller, 1933), longer-distance migration is especially likely to the largest cities. Appendix Table A.12 repeats the tests of rows 5–8, restricting the sample to mega-cells that do not contain the capital or any of the 5 largest cities or regional capitals of each country in 1960. These mega-cells are the least likely to be destinations of long distance migration, especially if there are ethnic differences across mega-cells. Results are again noisy but similar. They do not rule out urban 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.

To consider the possibility that the effect on population is driven by urban natural increase, we replace population growth in our baseline specification with

the rate of natural increase (RNI) based on the infant mortality rate (IMR) and the crude fertility rate (CFR) for women 15–33 estimated for 751 cell-decades using data for 16 countries from the Demographic and Health Surveys. We convert each city's CFR to a crude birth rate (CBR) using its country-years's share of women 15–33 in the overall population. We convert the IMR to a crude death rate (CDR) using the country-year-level regression relationship between IMR and CDR in our overall sample. Appendix Section A.4 describes the details of generating these estimates. The city-decade-specific RNI is then simply the CBR minus the CDR. Given the small values of RNI involved (so that the approximation $x \approx \ln(1 + x)$ works well), we interpret these percentage changes as good approximations of the natural increase component of $\Delta \ln(\text{population})$.

Table 4 row 9 reports results of our baseline specification for this limited sample of 751 cell-decades. Results are remarkably similar to baseline on this small sample, though the instruments are quite weak. Row 10 reports results for the same sample with RNI as the dependent variable. Focusing on IV5, the effect of market access on RNI is positive and about one third of the effect of market access on population growth. The OLS regression shows no effect on RNI. The other, weaker, instruments suggest that about half of induced population growth is RNI.

There are at least two broad reasons to be cautious about this result. The first is likely upward bias in the share of induced population growth assigned to RNI because we cannot know whether the births we measure took place in the city, or rather to mothers that subsequently moved to the city. This means that births we assign to the city could instead be births that happened elsewhere, and thus likely represent migrants instead, to the extent that children move with their mothers. While the converse is also true (mothers who gave birth in the city and then moved away are not captured), to the extent that urban-to-rural migration is less prevalent (and rural fertility rates are higher), the overall role of natural increase is likely to be overstated. The second is that there is a lot of noise in this result, not all of which is likely to be captured in the standard errors because this is a small subset of our overall sample, roughly one sixth of

the total. Among other reasons, all the instruments, including IV5, are quite weak in this sample (but of course not differentially so between RNI and total population change, so this does not imply bias in the RNI share). In addition, all of the demographic estimates are based on relatively small samples of births in each city, and the translation from CFR and IMR to CBR and CDR relies on relationships in national-level data that are themselves interpolated.

To the extent that this result is correct, it is inconsistent with the standard theoretical prediction associated with the demographic transition that if MA increases income it should therefore reduce fertility. A possible reason for this is that the induced migrants have high fertility relative to existing urban residents, because migrants tend to be young, in high fertility age groups. (Note that this is separate from the measurement issue discussed above – here we are talking about births that happened after a migrant moved to the city.)

Alternatively, in Row 11, we report the baseline regression on the same DHS sample with controls for contemporaneous CFR and IMR. Results are similar to baseline, with the same caveats as above: a small sample, weak instruments, and measurement error in demographic rates (which could now bias the coefficients of interest in either direction). They however suggest more strongly that natural increase is not driving the bulk of our main effect.

## 4.   Results: Heterogeneous Effects

Transport investments may have different effects depending on the local context in which they take place. Table 5 explores heterogeneity of results with respect to several factors highlighted in recent literature on economic geography, structural change, and political economy. As in Tables 2 and 4, each row shows 30-year estimates of a variant of equation (2). We control for the dummy variable shown at left and interact it with contemporaneous and lagged changes in market access, and the analogous instruments. For the IV5 estimation strategy the table reports the 30-year coefficient for the dummy=0 group, the dummy=1 group, and the difference; for IV10 and IV15, which are weaker instruments, 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. 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.** Row 1 of Table 5 compares cities above and below the median in distance to the country's top (capital, largest and second largest) cities in 1960. More remote cities see stronger effects of a marginal improvement. This is consistent with decreasing marginal returns to transportation investments, and suggests that remoteness raises their returns. Appendix Table A.13 also shows that cities initially (in t − 30) smaller than their country's median city generally see larger effects. Core-periphery models predict that reduced trade costs increase the size of big cities more than small cities. If anything, reduced trade costs lead to a decentralization of urban population in our context.[40]

**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 attract workers into secondary and tertiary sectors. Row 2 shows variation with respect to a measure of agricultural land suitability within one cell of the city, cutting the sample at 25% percent suitability.[41] The weaker IV10 and IV15 instruments suggest that cities near worse land are more positively affected by increases in market access. This is consistent with a smaller opportunity cost of leaving poor farmland, and with

---

[40]Appendix Section A.12 and Appendix Table A.13 show that differentials based on each of the following are mostly in the consistent direction: (i) distance below the median 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; (ii) access below the median to paved/improved roads (in 1960), railroads (1960), ports (1960 or 2005), or airports (2007); (iii) market access below the country's 25th, 50th or 75th percentile or the continental median; and (iv) population below the country's 25th or 75th percentile in $t - 30$, or the continental median, or alternatively dropping the top cities in 1960.

[41]FAO's Global Agroecological Zones (GAEZ) dataset 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.5).

cities in less agricultural areas specializing in more transport intensive activities that benefit more from the roads. The OLS and IV5 estimates are much noisier and in opposite directions.[42]

**Political Geography.** Row 3 of Table 5 allows for a differential effect for city-decades of road-building that may have been favored because they were within 150 km of the place of origin of a head of state in power for at least two years in the decade (the mean decade-specific tenure). We use 150 km because this represents a 3–4 hour driving time from the hometown given a driving speed of 40-60 kph (what we assume for improved/paved roads). 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 IV15 is 0.103). However, the instruments are even weaker here than in rows 1 and 2. In Appendix Section A.12, we report more results that overall suggest stronger negative effects for leaders with a longer tenure and whose regime is not democratic.[43]

A leader's origin penalty 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 grow faster than the country's other cities 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, regional capitals see if anything larger effects of market access on their growth, consistent with, for example, complementarity between government services and transport-sensitive activities (see Appendix Section A.12 and Appendix Table A.14).

---

[42]We find generally analogous effects when we cut the sample at 75% suitability, define suitability based on food crops or cash crops only, and when we study interaction effects with rainfall (see Appendix Section A.12 and Appendix Table A.14). We find less evidence of a differential for cities closer to mines, a sector whose labor intensity varies.

[43]Appendix Section A.12 and Appendix Table A.14 show that the differential is consistently negative, though often not precisely estimated, when varying the distance, the minimum tenure and other aspects of the specification.

**First stage heterogeneity and local average treatment effects (LATEs).** The results from Table 5 suggest that market access has stronger effects on the population of more remote cities. As noted 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 1 are larger than their OLS counterparts. In that case, we would expect the overall average effect to be lower than the LATE.

To consider this possibility, we investigate heterogeneity by remoteness (far from top 1960 cities) in the first stage in Appendix Table A.15 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 less remote cities. This suggests that our main effects capture LATEs that that are concentrated among *less* remote places, where results from Table 5 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 1.[44]

**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.9 show that adding substantial uniform border costs has little effect on results. In Table 6, 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

---

[44]Panels B and C show analogous results based on an alternative form of remoteness (less than median 1960 MA; panel B) and city size (population below the national median in $t-30$; panel C). 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.

are instrumented as above, calculating access only to the set of destinations under study. There are six endogenous variables (two market accesses × three lags) and six instruments, so instrument sets are weaker.

Row 1 of Table 6 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 smaller, indistinguishable from zero in all IV specifications.[45]

Row 2 investigates the effects of overland vs. overseas market access. We treat market access to 44 major ports in 2005 as a proxy for overseas market access because ports are the primary conduits of international trade. Overseas market access thus capture road changes and population changes for the cities with a 2005 port. While we do not have comprehensive historical measures of port traffic, the port cities' 2010 populations are highly correlated (at 0.68) with their port traffic volume (in 20-foot equivalent units) in 2005. We thus believe that the population of a port city is a reasonable proxy for port traffic.[46] Overland market access captures access to cities without a major port in 2005. The two measures are correlated at 0.31. The six instruments always include IV5 and its two lags for overland, while the remaining three for overseas differ by column as shown. [47]

Rising access to overland markets generally increases the size of cities. Point estimates suggest that overseas access (to non-African countries) is generally more important that foreign access (to other African countries), but it is only

---

[45]Connections to wealthier cities/countries may be more important than connections to poorer cities/countries. However, Appendix Section A.12 and Appendix Table A.16 show that results are similar when market access is constructed weighting destination city populations by their country's contemporaneous per capita GDP.

[46]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.

[47]We control for log distance to the coast interacted with country-year change 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).

significantly different from zero for IV10 and IV15.[48]

**Summary.** We find suggestive evidence that the effect of market access is stronger for: (i) remote (and small) cities; (ii) cities whose hinterlands do not have a comparative advantage in agriculture; and (iii) cities less likely to be politically favored, except 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 are due primarily to induced urbanization, and that they 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.13, 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, our estimates attribute 4–9% of the intensive margin increase in the urban share between 1960 and 2010 to these road upgrades. Applying our estimated elasticities to the proposed TAH network suggests a somewhat smaller effect on the urbanization projected over the subsequent 30 years.

---

[48]Appendix Section A.12 and Appendix Table A.16 show effects are similar when: (i) dropping the ports themselves and nearby cities; (ii) including the ports in calculating overland market access and its instrument; (iii) using 1960 populations; and (iv) using 1960 instead of 2005 ports.

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 and to consider what an optimal road network would look like, given the region's heterogeneity in physical, economic, and political geography.

## References

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: ADB, 2012–5.

Alder, Simon, "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," mimeo, University of North Carolina 2019.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, March 2020, *110* (3), 797–823.

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*, 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.

Baum-Snow, Nathaniel, 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.

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.

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

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Exogenous Shocks: Theory and Applications," Working Paper 27845, National Bureau of Economic Research September 2020.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Under-investment 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 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.

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.

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.

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.

_ , 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.

Felbermayr, Gabriel J. and Alexander Tarasov, "Trade and the Spatial Distribution of Transport Infrastructure," CESifo Working Paper Series 5634, CESifo Group Munich November 2015.

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.

Gollin, Doug, Rémi Jedwab, and Dietrich Vollrath, "Urbanization with and without Industrialization," *Journal of Economic Growth*, 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.

Henderson, Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102*, 994–1028.

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

Herbst, Jeffrey, *States and Power in Africa: Comparative Lessons in Authority and Control*, Princeton, NJ: Princeton University Press, 2000.

Jaworski, Taylor and Carl T. Kitchens, "National Policy for Regional Development: Historical Evidence from Appalachian Highways," *The Review of Economics and Statistics*, 2019, *101* (5), 777–790.

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," *American Economic Journal: Macroeconomics*, 2019, *11* (1), 223–75.

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

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*, 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.

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 Anthony Venables, "Economic geography and international inequality," *Journal of International Economics*, 2004, *62* (1), 53–82.

_ 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.

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.

_ , *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 and cities in the 39-country sample, 1960 and 2010

(a) Roads c. 1960                              (b) Roads c. 2010



(c) Cities in 1960                             (d) Cities in 2010
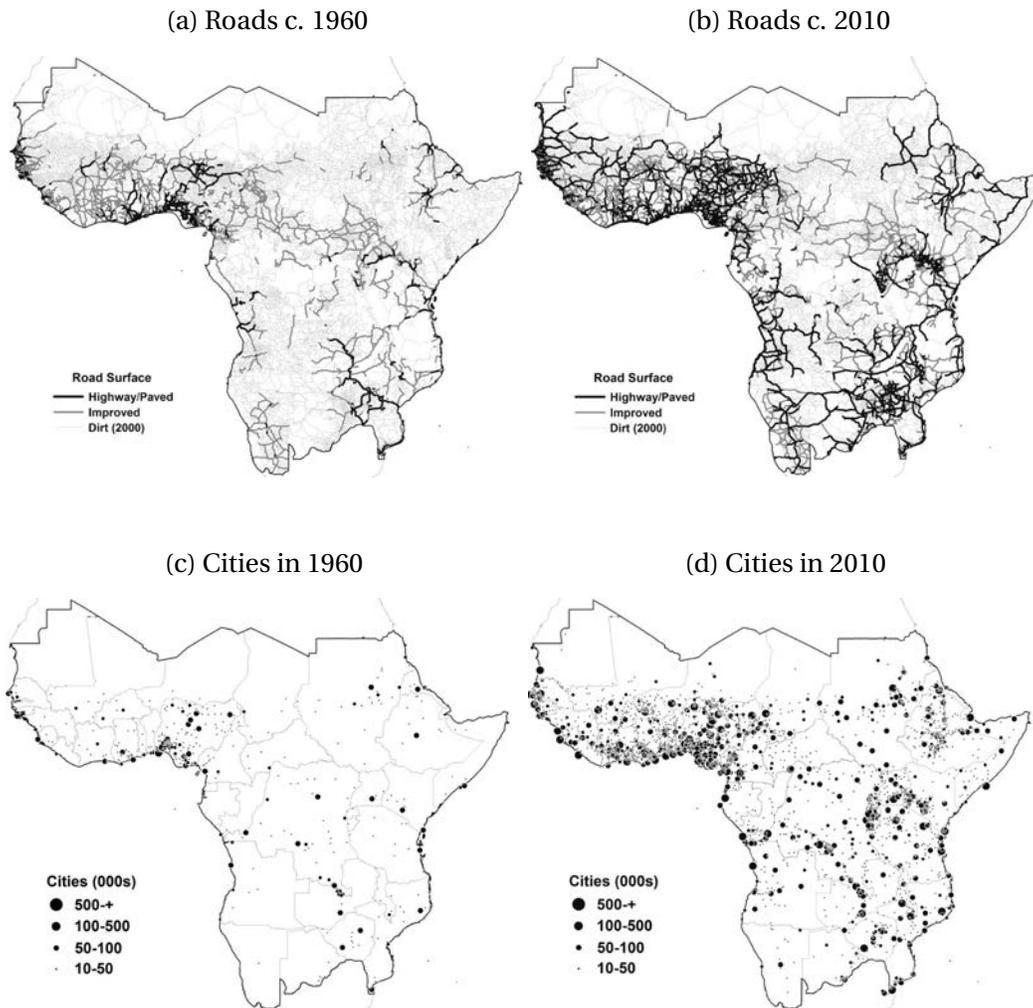


*Notes:* Subfigures 1a and 1b show the roads in the 39 sub-Saharan African countries of our sample in 1960 and in 2010 respectively. Roads are classified into four categories: *highways, paved, improved,* and *dirt.* See Appendix Section A.1 for details. Subfigures 1c and 1d show the cities (defined 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 Section A.2 for details.

Figure 2: Identification strategies

(a) Main



(b) Convex hull
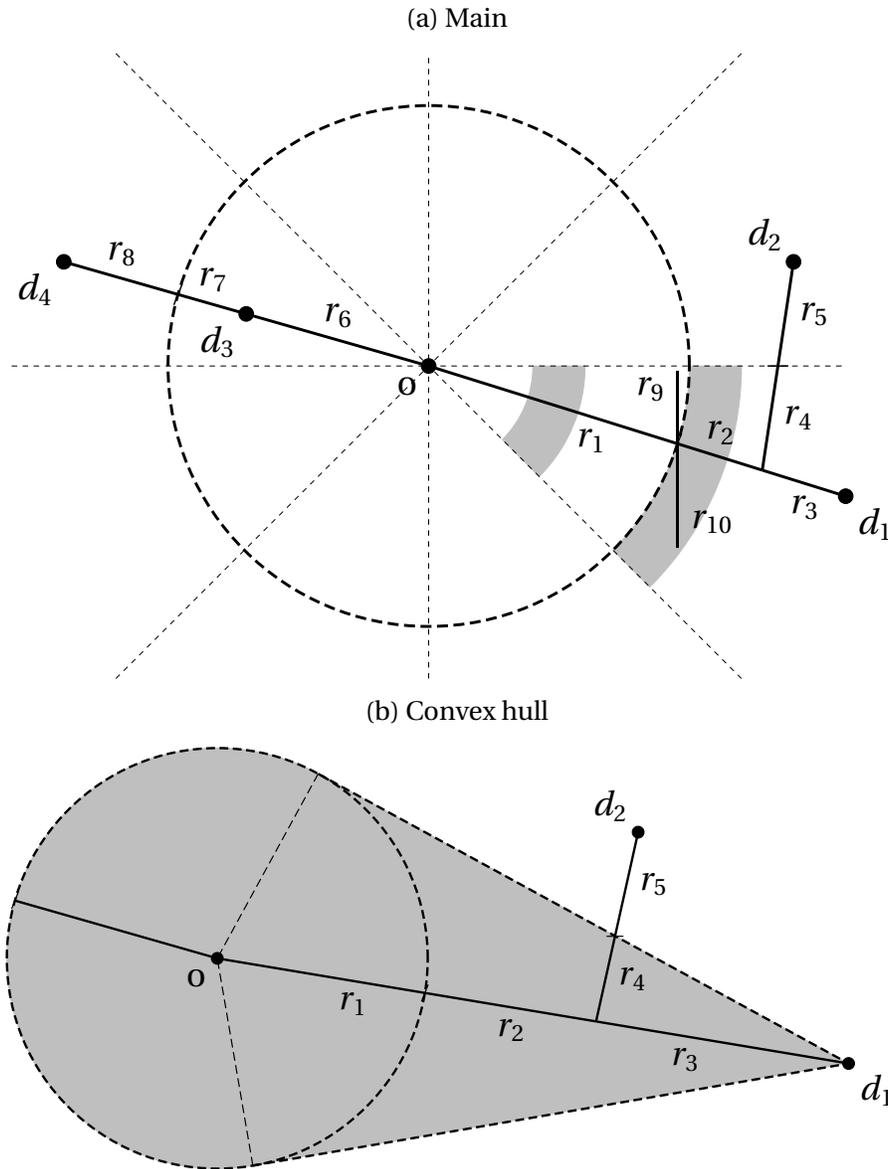


*Notes:* Panel (a): We instrument city *o*'s change in market access to other cities (here, $d_1$–$d_4$) with the change in market access due to road changes far away (i.e., outside the dashed circle) while fixing population of all cities at their initial levels. The exclusion restriction is that road changes outside (here, $r_2$, $r_3$, $r_4$, $r_5$, $r_8$ and $r_{10}$) are exogenous. Robustness checks exclude (i) 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); (ii) 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 (iii) cities where an inner road is upgraded in the decade after an outer road in the same octant is improved. Panel (b) In the convex hull strategy, all roads built near the line 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

| Panel A: OLS | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $\Delta_{t-10}^{t}$ ln Market Access | 1.33*** [0.38] | 1.33*** [0.42] | 1.56*** [0.37] | 1.57*** [0.46] | 1.52*** [0.41] |
| $\Delta_{t-20}^{t-10}$ ln Market Access | | 0.98*** [0.27] | 1.18*** [0.31] | 1.49*** [0.36] | 1.10*** [0.34] |
| $\Delta_{t-30}^{t-20}$ ln Market Access | | | 0.73*** [0.24] | 0.80** [0.32] | 0.76** [0.30] |
| $\Delta_{t-40}^{t-30}$ ln Market Access | | | | 0.31 [0.25] | |
| $\Delta_{t}^{t+10}$ ln Market Access | | | | | 0.72 [0.56] |
| Overall Effect ($t-40$ to $t$) | 1.33*** [0.38] | 2.31*** [0.58] | 3.47*** [0.63] | 4.18*** [0.83] | 3.39*** [0.71] |
| Observations | 5,906 | 5,472 | 4,725 | 3,630 | 2,607 |
| Adj. R-squared | 0.23 | 0.19 | 0.17 | 0.16 | 0.19 |

| Panel B: IV | IV: Exclude 5 (1) | IV: Exclude 10 (2) | IV: Exclude 15 (3) |
|---|---|---|---|
| $\Delta_{t-10}^{t}$ ln Market Access | 3.09*** [1.10] | 4.45** [1.82] | 5.55* [2.99] |
| $\Delta_{t-20}^{t-10}$ ln Market Access | 3.04*** [0.87] | 5.56*** [1.50] | 6.68** [2.62] |
| $\Delta_{t-30}^{t-20}$ ln Market Access | 2.23** [0.88] | 2.88** [1.38] | 4.27** [1.94] |
| Overall Effect ($t-30$ to $t$) | 8.35*** [2.16] | 12.89*** [3.23] | 16.49*** [4.60] |
| Observations | 4,725 | 4,725 | 4,725 |
| 1st stage Kleibergen-Paap F | 98.56 | 42.94 | 11.90 |

*Notes:* Each column is a separate regression of $100\Delta_{t-10}^{t}$ ln(urban population) 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-change fixed effects, ln(urban population) in $t-10$, and third order polynomials in longitude and latitude interacted with year fixed effects. In columns 2–4 of Panel B, measures of $\Delta$ln 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 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 2: Main Robustness Checks

|  | OLS | IV: Excl. 5 | IV: Excl. 10 | IV: Excl. 15 |
|---|---|---|---|---|
|  | (1) | (2) | (3) | (4) |
| (1) Co-Investment: Inner: 2, Outer: 15 | 3.65*** | 10.21*** | 14.39** | 19.22* |
| (N=2,260; F: _; 60.6; 12.2; 5.1) | [0.98] | [3.34] | [6.24] | [9.97] |
| (2) Radial Ext. Out.: Inner: 2, Outer: 15 | 3.86*** | 9.60*** | 13.75*** | 17.93*** |
| (N=1,603; F: _; 76.6; 21.4; 5.1) | [0.92] | [2.44] | [3.89] | [6.20] |
| (3) Radial Ext. In.: Inner: 2, Outer: 15 | 3.11*** | 8.76*** | 10.94*** | 13.17*** |
| (N=2,867; F: _; 107.2; 52.9; 9.9) | [0.85] | [2.34] | [3.43] | [4.86] |
| (4) Excl.  Changes Convex Hull 100k+ |  | 8.00*** | 11.83*** | 14.01*** |
| (N=4,725; F: _; 43.7; 35.6; 12.0) |  | [2.19] | [3.07] | [4.59] |
| (5) Excl. ΔTranscontinental Road |  | 8.22*** | 12.78*** | 14.43*** |
| (N=4,725; F: _; 44.6; 27.8; 6.7) |  | [2.32] | [3.16] | [5.12] |
| (6) Excl. Nat'l, Regional & Top 5 Cities | 3.61*** | 7.73*** | 12.96*** | 15.30** |
| (N=3,799; F: _; 87.8; 8.3; 7.3) | [0.75] | [2.56] | [4.56] | [6.63] |
| (7) Fix Population to 1960 in IVs |  | 7.54*** | 11.81*** | 16.20*** |
| (N=4,723; F: _; 49.4; 18.8; 4.6) |  | [1.90] | [2.73] | [3.97] |
| (8) Fix Pop. to 1960 in Market Access | 3.18*** | 9.83*** | 16.45*** | 24.71*** |
| (N=4,723; F: _; 134.3; 24.0; 9.7) | [1.10] | [2.54] | [3.95] | [6.04] |
| (9) Control Regional Growth | 2.34*** | 6.06*** | 7.97** | 8.69* |
| (N=3,498; F: _; 31.2; 8.2; 4.2) | [0.65] | [2.23] | [3.29] | [4.61] |
| (10) Province (1960)-Year FE | 1.34* | 5.54 | 13.68* | 13.13 |
| (N=4,725; F: _; 44.1; 5.5; 2.2) | [0.76] | [3.38] | [7.34] | [11.24] |
| (11) Quadrant-Year FE | 2.38*** | 6.28*** | 8.81** | 7.87 |
| (N=4,725; F: _; 97.4; 16.8; 6.6) | [0.63] | [2.33] | [4.11] | [6.91] |

*Notes:* This table is structured like Table 1 but only reports the overall effect. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 3: Effect of Market Access on Night Lights

|  | (1) OLS | (2) IV: Excl. 5 | (3) IV: Excl. 10 | (4) IV: Excl. 15 |
|---|---|---|---|---|
| $\Delta_{t-10}^{t}\ln$ MA | 0.39 | 22.29** | 43.98*** | 69.37*** |
|  | [3.06] | [9.97] | [11.98] | [18.56] |
| $\Delta_{t-20}^{t-10}\ln$ MA | 1.70 | 12.52 | 8.82 | 5.84 |
|  | [2.90] | [8.12] | [12.50] | [16.91] |
| $\Delta_{t-30}^{t-20}\ln$ MA | 0.84 | 3.90 | 0.71 | -3.01 |
|  | [2.11] | [4.32] | [7.33] | [10.21] |
| Overall Effect | 2.93 | 38.70*** | 53.51*** | 72.19*** |
|  | [5.22] | [11.39] | [18.31] | [26.49] |
| 1st stage Kleibergen-Paap F |  | 53.24 | 29.74 | 9.837 |

*Notes:* See Table 1. Outcome variable is $100\Delta_{t-10}^{t}\ln$ (Light Intensity). N = 3,591. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 4: Population Reallocation across Cities

| | OLS | IV: Excl. 5 | IV: Excl. 10 | IV: Excl. 15 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| (1) Baseline | 3.47*** | 8.35*** | 12.89*** | 16.49*** |
| (N=4,725; F: _; 98.6; 42.9; 11.9) | [0.63] | [2.16] | [3.23] | [4.60] |
| (2) Urbanization≤ 18% (50th %ile) t-30 | 2.98*** | 9.32*** | 13.27*** | 17.31*** |
| (N=2,279; F: _; 57.2; 22.8; 8.5) | [0.80] | [2.45] | [3.54] | [5.27] |
| (3) Urbanization≤ 10% (25th %ile) t-30 | 2.02* | 5.83** | 7.94** | 7.64* |
| (N=1,250; F: _; 27.3; 6.9; 3.2) | [1.05] | [2.44] | [3.52] | [4.57] |
| (4) Urbanization≤ 7% (10th %ile) t-30 | 3.21** | 10.23*** | 13.76** | 15.08** |
| (N= 715; F: _; 10.7; 4.5; 1.2) | [1.41] | [3.59] | [5.40] | [7.38] |
| (5) 3x3 Mega-Cells | 6.22*** | 7.87*** | 11.09** | 9.42 |
| (N=3,948; F: _; 41.7; 4.4; 1.1) | [0.83] | [2.76] | [4.61] | [6.69] |
| (6) 5x5 Mega-Cells | 6.95*** | 5.45** | 8.27** | 10.34 |
| (N=3,316; F: _; 8.6; 7.9; 4.2) | [1.03] | [2.64] | [4.10] | [6.29] |
| (7) 7x7 Mega-Cells | 8.04*** | 12.91*** | 16.74*** | 15.94** |
| (N=2,778; F: _; 21.6; 3.5; 1.2) | [1.13] | [3.47] | [5.59] | [8.07] |
| (8) 9x9 Mega-Cells | 9.29*** | 4.10 | 10.18* | 10.18 |
| (N=2,320; F: _; 17.8; 10.3; 5.0) | [1.27] | [3.34] | [5.72] | [8.94] |
| (9) DHS Sample | 4.47*** | 9.29** | 10.49* | 11.51 |
| (N=751; F:_; 8.1; 5.4; 4.0) | [1.51] | [3.58] | [5.81] | [7.03] |
| (10) DHS Sample: nat. incr. as dep. var. | 0.15 | 3.59** | 6.50*** | 7.81*** |
| (N=751; F:_; 8.1; 5.4; 4.0) | [0.45] | [1.56] | [2.35] | [2.77] |
| (11) DHS Sample: control for nat. incr. | 4.47*** | 9.30*** | 10.59* | 11.66* |
| (N=751; F:_; 8.5; 5.3; 3.8) | [1.51] | [3.44] | [5.61] | [6.79] |

*Notes:* This table is structured like Table 2. Rows 2–4 limit the sample 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. Rows 9–11 use the sample of cell-decades with DHS data and drop the $\ln(pop_{t-10})$ control. In rows 9 and 11 the dependent variable is $100\Delta_{t-10}^{t}\ln$ (urban population); in row 10 it is 100*RNI. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 5: Heterogeneous Effects of Market Access on Urban Population

|  | OLS Diff. | Col. (2)–(4): IV5 | | | IV10 Diff. | IV15 Diff. |
|  | | 0 | 1 | Diff. | | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| (1) > Med. Dist. Top Cities | 7.54*** | 2.16 | 9.21*** | 7.05*** | 13.41*** | 21.13*** |
| (F: _; 41.3; 10.5; 2.1) | [1.30] | [2.65] | [2.20] | [2.66] | [3.74] | [5.76] |
| (2) Crop Suitability < 25% | -0.92 | 6.95*** | 11.75*** | 4.80 | 11.28* | 20.54*** |
| (F: _; 17.0; 8.6; 3.9. Sh: 0.16) | [1.34] | [2.12] | [4.32] | [4.42] | [6.03] | [7.87] |
| (3) Leader's Origin 150km | -2.20* | 9.68*** | 1.68 | -8.00** | -7.35 | -9.88 |
| (F:_; 12.6; 5.7; 5.2, Sh: 0.24) | [1.26] | [2.20] | [3.94] | [4.00] | [5.21] | [6.29] |

*Notes:* Each row reports results from variants of Table 1 (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 cell-decade-changes with the dummy equal to one ("Sh") are reported in the left column. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table 6: Effect of Foreign versus Domestic Market Access

|  |  | OLS (1) | IV: Excl. 5 (2) | IV: Excl. 10 (3) | IV: Excl. 15 (4) |
|---|---|---|---|---|---|
| (1) | Domestic Market Access | 3.12*** | 6.37*** | 8.34*** | 10.51*** |
|  | | [0.56] | [1.89] | [2.72] | [3.71] |
|  | Foreign Market Access | 2.01* | 2.16 | 1.36 | 0.30 |
|  | | [1.07] | [2.94] | [3.10] | [3.40] |
|  | First stage Kleibergen-Paap F | | 22.76 | 8.45 | 4.03 |
| (2) | Overland Market Access | 2.89*** | 6.79*** | 5.34** | 3.69 |
|  | | [0.68] | [2.53] | [2.56] | [2.76] |
|  | Overseas Market Access | 4.08 | 3.03 | 8.89* | 15.87** |
|  | | [2.83] | [3.64] | [4.84] | [6.96] |
|  | First stage Kleibergen-Paap F | | 47.04 | 29.56 | 8.88 |

*Notes:* Each column contains summed 30-year coefficients from two separate regressions. In row (1), market access to domestic and foreign cities, and their lags, are entered separately (N=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 (N=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. In addition to our standard controls from Table 1 we control for log distance to the coast interacted with country-year change FE. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.