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

Rémi Jedwab† Adam Storeygard‡

October 2015

DO NOT QUOTE/CITE WITHOUT PERMISSION; PRELIMINARY AND INCOMPLETE

Abstract

This paper documents the effect of road construction on city population growth in sub-Saharan Africa 1960-2010, through the channel of increasing market access. Using changes in market access due to foreign or otherwise distant road construction as a source of exogenous variation in overall market access, we estimate a 30-year elasticity of city population with respect to market access of 0.05 to 0.20, larger than OLS estimates of 0.02–0.05, but somewhat smaller than estimates for other contexts. The effect is concentrated 10–30 years after road construction, closer to coasts and borders, and farther from a country's largest cities.

JEL classification: F15, O18, R11, R12, R4
Keywords: Transportation Infrastructure; Trade Costs; Roads; Urbanization; Cities; Africa; Market Access; Highways

*We thank James Fenske, Vernon Henderson, Harris Selod and seminar audiences at EPEN (CIDE-Princeton), GMU-GWU Economic History Workshop, IEA (Jordan), UND, the World Economic History Congress (Kyoto) and the World Bank for helpful comments, and Yasmin Abisourour, Karen Chen, Rose Choi, Erin McDevitt, and Emily Ryon for research assistance. We are grateful to to Uwe Deichmann and Siobhan Murray for sharing their roads data, François Moriconi-Ebrard from GEOPOLIS for his help with data collection, and Durwood Marshall for programming advice. We thank the World Bank's Strategic Research Program on Transport Policies for Sustainable and Inclusive Growth and Urbanisation in Developing Countries Programme and the Institute for International Economic Policy and the Elliott School of International Affairs at George Washington University for financial assistance. Storeygard thanks LSE, UC-Berkeley, and GWU for hospitality during the time this research was conducted.

†Department of Economics, George Washington University, 2115 G Street, NW, Washington, DC 20052, USA (e-mail: jedwab@gwu.edu).

‡Department of Economics, Tufts University, Medford, MA 02155, USA (e-mail: adam.storeygard@tufts.edu)

1 Introduction

Sub-Saharan Africa is the least urbanized major 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).[1] This combination of low urbanization and poor road connectivity mean that many people face relatively low access to national and global markets (Limão and Venables, 2001; Atkin and Donaldson, 2015).

While road construction was rapid in the 1960s and 1970s post-Independence, it slowed substantially in the subsequent three decades along with overall public investment. Gross capital formation averaged just under 25 percent of GDP in the region in the 1970s, falling to 21 percent in the 1980s and 17 percent in the 1990s, even as GDP itself was stagnating.[2]

African countries have begun to make large infrastructure investments again, with plans for even larger ones in the next decade, in roads and highways in particular.[3]. These projects are being carried out in diverse contexts, from sparsely populated Northern Kenya to the Abidjan-Lagos corridor, home to 35 million people along a 1000 km route. In the official documents of these projects, they are often described as having the potential to transform their regions. For example, the World Bank writes of the Abidjan-Lagos project: "The potential of the corridor to become a catalyst for economic growth and regional integration in the sub-region is well documented, and it is the hope and aspiration of the governments of the five countries, with assistance from the World Bank, to harnesses this potential for socio-economic development."

The African Development Bank in 2003 published a detailed report on a proposed 59,000 km Trans African Highway system (African Development Bank and United Nations Economic Commission For Africa, 2003). While such plans have been articulated in policy circles since at least the early 1970s, the Bank has highlighted this system when funding several recent projects.[4]

[1] While South Asia has slightly fewer roads per person, far more of them are paved.

[2] World Development Indicators, http://data.worldbank.org/indicator/NE.DAB.TOTL.ZS ; http://data.worldbank.org/indicator/NE.GDI.TOTL.ZS, accessed 2015/09/01.

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

[4] See for example, http://www.afdb.org/en/news-and-events/article/afdb-approves-fcfa-32-billion-loan-for-dakar-diamniadio-highway-4895/; http://www.afdb.org/en/news-and-events/article/ethiopia-kenya-us-326-5-million-for-road-project-4862/; http://www.afdb.org/en/news-and-events/article/

It is thus a good time to reconsider the effect of earlier road construction on the economic geography and growth of the region, with a view to understanding the effect of potential future projects.

In this paper, we consider the effect of roads built and upgraded between 1960 and 2010 on city growth during that period, as a result of increased market access to other cities. Using a variety of controls and instrumental variables strategies to account for potential endogeneity of road improvements, we find that a 10% increase in market access due to upgraded roads induces a 0.5–2% increase in city population, with the bulk of the effect in the 10 to 30 years after road construction. This effect is robust to a variety of alternate samples and identification strategies, and the inclusion of several controls. The OLS effect is smaller though precisely measured, suggesting that far from anticipating future growth, roads are built in otherwise lagging regions. This is consistent with a network that is expanding from the largest and most developed cities at independence to poorer, more remote places later.

Our work relates primarily to the literature on the effect of market access, and specifically intercity transport costs, on the growth of local areas (Chandra and Thompson, 2000; Haines and Margo, 2008; Bogart, 2009; Banerjee et al., 2012; Faber, 2014; Donaldson and Hornbeck, forthcoming; Storeygard, 2015; Jedwab and Moradi, forthcoming; Donaldson, forthcoming).[5] More generally, a large literature has looked at how market access affects the growth of neighborhoods (Ahlfeldt et al., forthcoming), cities (Redding and Sturm, 2008), regions (Hanson, 1998), and countries (Feyrer, 2009). Another large literature has looked at the effect of large highway projects on a variety of outcomes (Baum-Snow, 2007; Rothenberg, 2013; Baum-Snow et al., 2015; Coşar and Demir, 2015). Finally, a smaller literature has emphasized the specific role of road quality, which is the main source of variation in this work (Casaburi et al., 2013; Gertler et al., 2014)

The paper makes several contributions to this literature. First, we document the development of a continental paved road network from near its beginnings to the present, and this data richness allows us to consider the timing of effects in ways that previous work, which is mostly based on two or three cross-sections instead of our

afdb-approves-eur137-million-to-finance-tunisia-libya-highway-8199/;
http://www.afdb.org/en/news-and-events/article/enugu-bamenda-road-tool-of-regional-integration-and-afdb-success-story-12528/;
http://www.afdb.org/en/news-and-events/article/launch-of-flagship-road-project-in-kenya-a-dream-realized-9987/ , accessed 5 September 2015

[5] For a comprehensive overview of this literature, see Redding and Turner (2015). For a review focused on the developing world, see Berg et al. (2015).

six over 50 years, cannot. We also use the universe of intercity paved and improved roads, as opposed to highways alone as considered by most of the literature, and study an evolution of the road network rather than a revolution of the kind China has experienced in the past 25 years. To the extent that gradual evolution is more likely in the future of most developing regions, we believe that this is an instructive context. Second, building on Donaldson and Hornbeck (forthcoming), we use a novel identification strategy, relying on the variation in market access induced by roads built in foreign countries. Third, we consider a wide variety of heterogeneous effects, a subject that has received little attention in the literature.[6]

Our work also builds on the literature considering how cities in developing countries, and specifically sub-Saharan Africa, grow. Previous work on transport and city growth in Africa has emphasized railroads (Jedwab and Moradi, forthcoming; Jedwab et al., forthcoming) or variable costs of road transport (Storeygard, 2015), but not road construction, which is likely to provide the single largest source of decreasing transport costs in the future. Other empirical work on urbanization in Africa is primarily cross-country in nature (Fay and Opal, 2000; Bruckner, 2012; Henderson, Roberts and Storeygard, 2013; Castells-Quintana, 2014; Gollin, Jedwab and Vollrath, forthcoming).

2 Data and Background

2.1 Roads

We combine road information from two sets of sources. First, Nelson and Deichmann (2004) provides road locations for all of sub-Saharan 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 (16 sample countries), North/West (18) and North/East (5). We include all of mainland sub-Saharan Africa except for South Africa, Swaziland, and Lesotho, a total of 39 countries and 833 country-years. Figures 1 and 2 show the countries and years, respectively, covered by each region. While specific road categories vary somewhat across maps, the distinction between highways, other paved roads, improved (laterite or gravel) roads, and earthen roads is nearly universal.[7]

[6] Faber (2014) and Bird and Straub (2015) each look at selected types of heterogeneity.

[7] Several Michelin maps distinguish paved roads with one lane only from paved roads with two lanes or more, but not consistently across years. They also distinguish fully and partially improved roads,

The Michelin maps report highways and intercity paved and improved roads comprehensively, but their coverage of earthen roads is less complete, with some changes clearly due to coverage changes as opposed to new roads. Based on the assumption that roads change quality but rarely move or disappear, we thus use the Michelin maps to code each segment from the Nelson and Deichmann (2004) map as paved or improved in each year that it is paved or improved, 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.

Michelin uses four main sources to create the maps: (i) the previous Michelin map, (ii) new government road censuses/maps, (iii) direct information from its stores across Africa, and (iv) correspondence from road users including truckers and tourists.[8] The latter two sources of information are especially important, and we believe new to this literature.[9] Michelin has been producing road maps since 1910, with its first map for West Africa appearing in 1938. As the second largest tire company in the world, 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 therefore in regular contact with this network. Because inaccurate characterization of road surface can lead 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 most other maps classify roads as primary/secondary or major/minor, which is less informative about road quality. We are also 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 raises several issues. First, this revision process likely means that changing conditions are reflected in the maps with a lag. 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. Finally, we cannot capture the quality of roads within a surface class, so when a severely potholed paved road with a low maximum speed is resurfaced, our data do not reflect this. This work has been especially prevalent since 2000, so we may underestimate recent changes.

but again not consistently across years. This is why we use only the four categories listed.

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

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

2.2 City location and population

We obtained the location and population estimates of cities in 33 countries from *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa*.[10] The Africapolis team generated estimates using various sources including population censuses, demographic studies, administrative counts and electoral counts. Based on an initial list of cities with at least 5,000 inhabitants in the most recent census circa 2000, their final database includes all cities that reached a population of at least 10,000 at some point since 1960.

The Africapolis team also defined agglomerations in circa 2000 using satellite imagery. Thus, if two distinct cities in 1970 ultimately became one, 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 as in Baum-Snow (2007), Rothenberg (2013), and Baum-Snow et al. (2015)

We build on the Africapolis data in three ways. First, we use analogous sources to produce an analogous database for the 6 remaining countries not in the Africapolis samples (Angola, Botswana, Malawi, Mozambique, Zambia and Zimbabwe). The sources used for each of these countries are listed in the Web Data Appendix. Second, we added a small number of cities in Africapolis countries that achieved a population over 10,000 at some point between 1960 and 2010 but did not appear in Africapolis sample. Finally, we added locations for these additions and cities in Africapolis missing locations, and corrected locations that appeared to be incorrect, based on Google Earth, GeoNet, and Wikipedia, aggregating multiple administrative cities into one agglomeration using more recent satellite imagery from Google Earth.[11]

Population figures for all 39 countries are interpolated and extrapolated between raw data years to obtain estimates for each sample year (1960, 1970, 1980, 1990, 2000 and 2010).The resulting sample includes population estimates for nearly all cities with a population of over 10,000 at some point since 1960, in all sample years in which their population exceeded 10,000, and in most sample years in which they did not reach this threshold. While ideally we would prefer to have systematic information on cities over 5000, as used in work on other regions (e.g. De Long and Shleifer, 1993; Acemoglu et al., 2005; Dittmar, 2011), information on these cities are not systematically available for our sample region and period. Our baseline sample thus restricts to city-years with populations over 10,000. In robustness checks below, we use varying amounts of

[10] http://www.africapolis.org

[11] We were unable to obtain coordinates for 19 cities in Sudan, but only one of these had a population over 10,000 in multiple years.

available information on smaller cities.[12]

2.3 Other data

We obtain several physical geographic characteristics following Jedwab and Moradi (forthcoming). Climate data are from Willmott and Matsuura (2009). We calculate average annual precipitation (in mm) over the period 1900 to 1960 for each cell.[13] Elevation is from the Shuttle Radar Topography Mission version 3 (SRTM3 DTED1) 90-meter data (Farr et al., 2007).[14] We calculate for each cell the mean and standard deviation of elevation, in meters, across pixels. We use FAO (2001) to obtain for each cell the shares of of class 1 (the most suitable), class 2, class 3, undetermined, sparsely vegetated and submerged soils. Rivers data are from VMAP/GlobalGIS.

2.4 Trends

Railroads were the dominant long-distance transportation technology in Africa until World War II(Chaleard et al., 2006; Chaves et al., 2014; Jedwab and Moradi, forthcoming). Even as they built road networks at home, colonial governments were unable to use the same paving technologies in African soils and climates. The development of more suitable bitumenization and laterite technologies in the 1930s and 1940s, along with large decreases in the cost of imported cars and trucks, led to substantial but gradual expansion of the paved and improved road networks in the late 1940s and 1950s (Gould, 1960; Wasike, 2001; Gwilliam, 2011; Burgess et al., 2015).

Figures 3–8 map the evolution of the road networks over time since 1960. Figure 9 shows aggregate lengths of highways and paved and improved roads over time, and Figure 10 shows their cumulative shares, assuming a constant stock of total roads as measured circa 2004. Following independence, in the 1960s and 1970s, the paved network expanded much more rapidly, fueled by massive public investments in many countries to promote trade and industrialization (e.g. 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, the pace of road transformation decreased markedly, as less funding was available for infrastructure investment.[15] Although investment

---

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

[13] Available at http://climate.geog.udel.edu/~climate/htmlpages/archive.html

[14] Available at http://www2.jpl.nasa.gov/srtm/

[15] It is also possible that Michelin started receiving less updated information.

has increased again since the mid-2000s decade, this is not reflected in our data. We believe this is because investment has been directed primarily towards restoring and rebuilding existing paved roads. For example, in Kenya, the government has invested a lot of money rebuilding the main paved road from Mombasa to Nairobi to its original 1960s standard (Burgess et al., 2015).

The share of paved roads in the overall 39-country network remains small, at 13% compared to the United States (over 99%), China (54%) or India (49%).[16]

Figures 11–16 map the evolution of cities over 10,000 over time since 1960. The sheer number of such cities has increased rapidly, from 431 in 1960 to 2,819 in 2010. In 1960, a large fraction of these cities were regional administrative centers.

City population is of interest as a measure of local economic development, and in its own right (De Long and Shleifer, 1993; Acemoglu et al., 2005; Dittmar, 2011). No subnational GDP or wage data exist for most countries in the sample. While night lights have been used as a measure of city population (e.g. Storeygard 2015), they are not available in digital form before 1992. Even total population (and therefore urbanization rate) is not systematically available for subnational regions for the full temporal and spatial scope used here.[17] As such, we believe city population is the best measure of local economic development for Sub-Saharan Africa from 1960 to date.[18]

3 Empirical Analysis

Our unit of analysis is a 0.1 by 0.1 degree grid square, which covers approximately 123 square kilometers at the equator, decreasing with the cosine of latitude.[19] Using these units simplifies computation compared to the full vector road network, and avoids problems due to missing topological information, concerning which segments connect to each other and which do not, in most vector roads datasets. We assign to each grid

[16]These aggregate numbers are all problematic, due to the unclear nature of the denominator. When summed to the country level, our data are broadly in line with Canning (1998), the most reliable existing source of aggregate data. We thus believe that our measure of highways, paved, etc. roads capture relatively well the situation on the ground, even for the most recent period. The World Development Indicators' most recent estimate for all of sub-Saharan Africa is also 13%.

[17]Henderson et al. (2015) use information or urban and total populations for relatively coarse subnational units of 89 censuses in 29 countries. These data are not consistently available back to the 1960s fir all countries.

[18]While in principal we could use household asset ownership and child mortality from the Demographic and Health Surveys as measure of local economic development as in Young (2012, 2013), these data do not exist for periods before the late 1980s, have limited geographic information before the late 1990s, and exclude many small cities.

[19]These are the same units as Jedwab and Moradi (forthcoming). Because the maximum absolute latitude in our sample is 29 degrees, in practice this creates little variation in unit size.

square a speed of travel for the fastest road segment type (highway, paved, improved, earthen) falling in the grid square, or a baseline speed if no roads are present. Table 1 reports these speeds.[20] We assume 80, 60, 40, 12, and 6 km/h on highways, paved roads, improved roads, earthen roads, and areas with no roads, respectively. The precise values are illustrative; results are insensitive to a scale factor. Like most of this literature, we do not attempt to model congestion.

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

Following Donaldson and Hornbeck (forthcoming), we define origin cell o's market access (MA) in year t, as $MA_{ot} = \sum_{d \neq o} P_{dt} \tau_{odt}^{-\sigma}$, where P is population, d indexes destination cells, τ_{odt} is the time required to travel between cells o and d, and σ is the trade elasticity. Following Donaldson (forthcoming), we use $\sigma = 3.8$ at baseline and consider alternative choices.

We are interested in changes in MA between $t-10$ and t due to changes in roads. Our primary measure of change in MA is thus

$$\Delta \ln MA_{ot} = \ln(\sum_{d \neq o} P_{d,t-10} x_{odt}^{-\sigma}) - \ln(\sum_{d \neq o} P_{d,t-10} x_{od,t-10}^{-\sigma}). \quad (1)$$

In other words, we hold population constant at the initial level, $P_{d,t-10}$, so that we are considering only the variation in MA due to changes in roads between $t-10$ and t. We consider up to 3 lags of market access change to look for changing impacts over time. Indeed, we do not not expect the effect of road investments to be instantaneous.

Our baseline specification is thus:

$$\Delta \ln P_{ot} = \sum_{k=0}^{2} \Delta \ln MA_{o,t-10k} + X_{ot} + \epsilon_{ot} \quad (2)$$

Baseline controls X_{ot} are country-by-year fixed effects to account for arbitrarily

[20] Future work will explore robustness to using alternative speed assumptions from elsewhere in the literature, including Alder (2015), which are broadly similar but emphasize highways and place lower relative speeds on other roads.

[21] For roads in 1960, we assign roads from the earliest available map year (1961 for Central/South, 1965 for Northwest, and 1966 for Northeast). In effect, this assumes no road building between 1960 and the first map. This may underestimate road building between 1960 and 1970. We will show that our results are robust to dropping the 1960s.

shaped time trends that may differ by country, and fourth order polynomials in latitude and longitude by year, to account for potentially time-varying general spatial patterns of growth within countries unrelated to roads. Errors ϵ_{ot} are clustered at the cell level at baseline.

Our primary (intensive margin) sample consists of cell-years with 10,000 or more urban residents. This sample has the advantage of being consistently collected across all countries, but it is not balanced, and this lack of balance raises selection concerns. Places that entered the sample earlier may have been different from other cities in ways that are correlated with road building. As an alternative, in robustness checks we consider the balanced ever-populated sample, which consists of all cells that ever have more than 10,000 urban residents. While in many cases we have population estimates for years when the population is below 10,000, we must make assumptions about the urban populations of other (pre-)urban cells. To do this, we index time for each cell with respect to the first year it crossed 10,000 urban residents. We then calculate sample statistics for each year relative to crossing. At baseline, we assign to all cells with a population that is missing or otherwise under 10,000 to the median population of cells with a known population under 10,000 for that relative year. These medians are: 8,070 for ten years prior to first reaching 10,000, 5,387 twenty years prior, 3,903 thirty years prior, 2,912 forty years prior and 2,100 fifty years prior. In robustness checks we also use the mean, and alternatively, use known populations below 10,000 where available.

Table 2 contains descriptive statistics for our main variables of interest in a sample with two lags. The average city-decade saw small increase in market access of roughly 10%, but this statistic has a wide variance. Lagged changes are larger on average, due to more rapid road building early in the sample.

Our chief identification concerns are reverse causality, omitted variables, and measurement error. In Section 4 below, we discuss these concerns in detail, and we report the results of several alternative identification strategies.

4 Results: Average Effects

4.1 Baseline Results

Table 3 reports estimates of Equation (2), along with variants adding and removing lags and leads, on the intensive margin sample. In this and all subsequent tables, values of the dependent variable are divided by 100, so that coefficient can be interpreted as elasticities multiplied by 100: the percentage change in population associated with a one log point increase in market access. In Column 1, only contemporaneous changes in market access are included. These have essentially no effect. 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 a modest impact on city population, with an elasticity of 1%. A 10% increase in market access due to roads in each decade is associated with a 0.1% increase in population. The overall effect, across these three decades, is thus a little over 3%. Once these prior effects are controlled for, the contemporaneous effect is of similar magnitude, though more noisily estimated. In column 4, the prior decade, 30 years before the measured population change, has little effect.[22] We thus restrict the remaining specifications to two lags.

In columns 5 and 6, we investigates reverse causality by adding one and two leads, respectively. If roads are being built in anticipation of fast growth in a region, we would expect these leads to enter positively. Alternatively, if roads are built because of anticipated stagnation, we would expect them to enter negatively. In practice, neither specification shows any effect of leads. The addition of leads does however increase standard errors on the other coefficients due to the smaller number of observations (because we must drop the most recent decade). The last row of coefficients in Table 3 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 leads, with an elasticity of 3 to 4%.

4.2 Identification strategies

While the exclusion of contemporaneous population changes avoids the most direct forms of reverse causality, it is still possible that roads near a city are being built in part because of future expected growth in that city, or in nearby cities that are growing in tandem. Table 4 reports the results of several specifications intended to disentangle the causal effect of market access due to roads on city growth.

Column 1 repeats the baseline result from Table 3, and column 2 adds a control for the contemporaneous change in time cost (the inverse of the speed) assigned to the cell itself. Road changes within the cell itself are the most likely to be endogenous to city growth - any road built between that city and anywhere will enter that cell. Once we control for this, identification is then coming from changing in roads in other cells.

[22]The specification with two lags also has lower Aikike and Bayesian information criteria (AIC and BIC) than the one with 3 lags.

Standard errors increase, but point estimates are virtually unchanged, and the first and second lags remain significant at the 5 and 10% level, respectively.

Instead of controlling for local roads, Columns 3–6 instrument for changes in overall market access caused by roads built far away

The idea here is that local roads might be built because of a city's current or expected growth, or that there is an unobservable driving both city growth and local road building.

Formally, the instrument measures changes in market access due only to roads built outside a radius $j \in 5, 10, 15, 20$ cells from city o:

$$\Delta_j^{out} \ln MA_{ot} = \ln \left(\sum_{d \neq o, \delta(d,o) > j} P_{d,t-10} x_{odt}^{-\sigma} + \sum_{\delta(d,o) < j} P_{d,t-10} x_{od,t-10}^{-\sigma} \right) - \ln \left(\sum_{d \neq o} P_{d,t-10} x_{od,t-10}^{-\sigma} \right). \tag{3}$$

Effects are again substantially larger than the OLS, with elasticities between 10% and 20%, increasing with the excluded radius. We discount the results using a 20 cell exclusion zone because the instrument is somewhat weaker, with a first stage F-statistic of 12.2. Compared to the upper panel, these larger effects are more precisely estimated, and consistently suggest effects both 10 and 20 years after road changes. As expected, the instrument is stronger at lower radii, because it includes road changes more directly connected to the city.

Column 7 uses an analogous strategy to columns 3–6. However, instead of defining "near" and "far" in terms of Euclidean distance, it defines "near" as in the same country as cell o, and "far" as in a different country. For example, in order to predict city growth in Benin, we instrument with market access changes due to roads built in other countries, most importantly neighboring countries like Nigeria, Niger and Togo. This will be effective in removing endogeneity to the extent that countries do not take into account foreign cities in their road-building decisions.

The point estimate using foreign changes to instrument for overall changes is larger than the OLS estimate, but smaller than most of the estimates using the radius-based instruments, and with a comparable standard error, so that it is not significant at conventional levels. We believe this is because the role of borders is highly heterogeneous. If we cluster errors by country, it is significant at 9 percent.

The fact that the IV estimates are larger than the OLS is consistent to the recent literature. While the initial cause for concern was roads built to cities expected to grow faster, in practice, roads appear to be more likely to have been built toward lagging cities.

The magnitude of the effects we find is mostly somewhat smaller than the 0.25 to 0.3 reported for total population in US counties by Donaldson and Hornbeck (forthcoming), the most similar specification to ours in the literature. This could be because there are likely to be substantially more restrictions against trade and especially migration in this context, especially between countries, than within the United States. Donaldson and Hornbeck (forthcoming) also report estimated discrete effects of railroad construction on agricultural land prices, so that it is a cross-walk to the rest of the infrastructure literature. As noted by Redding and Turner (2015), these are substantially larger than the effects of roads and railroads on land prices and wages elsewhere in the literature, by a factor of two or more in some cases (Chandra and Thompson, 2000; Michaels, 2008; Haines and Margo, 2008; Bogart, 2009). This suggests that our results are broadly similar to other contexts.

4.3 Robustness

Tables 5–7 explore the robustness of results to changing specifications, controls, samples, and other assumptions. Each row represents a separate regression with contemporaneous market access and two lags, latitude and longitude polynomials, and country-year fixed effects, as in our baseline specification, unless otherwise indicated.

Specification and controls Table 5 considers alternative specifications and controls, with respect to the baseline OLS intensive margin sample from Table 3, column 3 (3.42), repeated here in row 1. In rows 2 and 3, clustering by country or country-year increases standard errors, but the lagged effects are still significant at 1%. Rows 4 and 5 split change in market access into 5-year periods before the present: 0–5, 5–10, 10–15, 15–20, 20–25, and 25–30 years. Because population data are only available every ten years, the population component of market access is lagged 10 years for both the 0–5 and 5–10 periods (and the population component of market access is lagged 20 years for both the 10–15 and 15–20 periods), so that these periods only index the roads data. Effects are substantially stronger in the 10–15 and 20–25 year periods than their immediate predecessors. In row 6, rather than lagging population 10 years as in Equation (1), we hold population constant at its 1960 values in calculating market access. Thus no population growth after 1960 is driving variation on the right hand side, only road building. This makes reverse causality from neighbors' population growth to own population growth less likely. Effects are very similar to baseline. Row 7, in contrast, harnesses all variation in both roads and population. In other words, the "after" portion of Equation (1) uses $P_{d,t}$ instead of $P_{d,t-10}$. This specification is more at risk of endogeneity because the same unobserved shocks could be driving population growth in city o and nearby cities d. The contemporaneous effect is larger and more precisely estimated than at baseline. However, the lagged estimates are quite similar. In row 8, market access is calculated only to domestic cities. This is equivalent to assuming that the cost of crossing a border is infinite, as opposed to the baseline in which this cost is zero. As results are very similar, we do not pursue a more realistic assumption of intermediate border costs further.

Rows 9–16 of Table 5 add controls to the baseline. Conceptually, changes in market access can be decomposed into those due to changes in transport costs or changes in population. In practice, a formal decomposition is not possible because of the nature of the aggregation of the market access measure: access to each destination city has its own roads component and its own population component. However, an approximation is possible. Row 9 adds a measure of contemporaneous change in market access due only to population, not roads:

$$\Delta \ln MA \text{ fixing Roads}_{t-10} = \ln(\sum_{d \neq o} P_{d,t} x_{od,t-10}^{-\sigma}) - \ln(\sum_{d \neq o} P_{d,t-10} x_{od,t-10}^{-\sigma}). \tag{4}$$

Because there are not new maps every year, we linearly interpolate road costs (the inverse of speeds) between map years. It is possible that this systematically assigns road improvements to take effect earlier or later than they in fact did. For example, if a grid square's best road is unpaved in 1969 but paved in 1971, we assign a speed of 12 km/h in 1969, 60 km/h in 1971, and 20 km/h in 1970.[23] If the paved road were in fact finished in late 1969, this would be an overestimate for 1970. If conversely it was not open to traffic until 1971, this would be an underestimate. Rows 10 and 11 undo this interpolation, rounding up and down, respectively, to the nearest actual road cost and re-running the Table 4, column 2 specification. To the extent that this over- or under-estimation is systematic in one or more decades, this bounds its effects.

All three of these controls have little effect. Row 12 adds dummies for increases and decreases in road costs. This reduces effects somewhat. Row 13 is an auto-distributed lag model, which adds two lags of the dependent variable alongside the two lags of the independent variable.[24] This increases effects slightly. Row 14 includes only year fixed effects instead of country-year fixed effects. The main difference is that, like in row 7,

[23] $20 = 1/(\frac{1}{2}[\frac{1}{12} + \frac{1}{60}])$

[24] In this case, the long-term effect of a variable X on a variable Y is equal to the sum of the effects of X and its lags on Y, divided by (1 minus the sum of the effects of each lag of Y on Y), provided Y and X are both stationary, which we confirm using various tests (not shown; see e.g. Greene (2008) for a demonstration).

the contemporaneous effect is larger and now significant.

Row 15 adds the following geographic controls: an indicator for being on the coast, distance to the coast, a river dummy, distance to a river, mean altitude, standard deviation of altitude (a form of terrain ruggedness), mean precipitation, shares of class 1, class 2, class 3, undetermined, and sparsely vegetated soils, and water in the cell. The idea is that these features, which are spatially correlated, might be driving both road building and city growth. For example, flat terrain is better for both food production that supports city growth and road building. However, they have essentially no effect on results.

Sample Table 6 varies the sample. Row 1 changes the sample weights to place greater emphasis on larger cities at time $t-10$. In the absence of weights, to the extent that any random noise is additive, it may affect small cities more. Weighting decreases the effect of this noise, but it has essentially no effect. Rows 2–4 and 5–7 drop the cell-years with the largest increases in market access and population, respectively, to ensure that outliers are not driving results. As expected, dropping the top cell-years in terms of market access growth increases effects, and dropping the fastest growing cell-years in terms of population decreases effects, though it also makes estimates of the contemporaneous effect more precise. Row 8 drops the cell that contain the 1960 capital and the cells with the largest two cities (which may include the capital) in 1960 for each country, and row 9 drops these and their equivalents from 2010 (which may be the same as those in 1960), to limited effect. Thus results are not being driven by the largest cities. Remarkably, dropping even more large cities, the top 5 or even 15, as in Rows 10 and 11, or the top decile in the continent or by country, as in Rows 12 and 13, still has little effect on baseline results. These results are akin to the identification strategies of Faber (2014) and Michaels (2008), in that they do not rely at all on the largest cities, whose growth is more likely to have driven the placement of road construction. Instead, they rely on smaller cities, which were more likely to be connected incidentally. Row 14 drops the same three cities as Row 9, while also controlling for the distance to each of them, which also has little effect.

Because not all South African cities re included in calculating market access, it is possible that access to South African cities is biasing results. This is most likely to be true in the four sample countries that border South Africa: Botswana, Mozambique, Namibia, and Zimbabwe. In row 15, removing these countries has little effect, if anything increasing the estimates. Analogously, rows 16 and 17 drop countries nearest to North Africa and the Arabian Peninsula, respectively, with little effect.

As noted above, roads data for the 1960s are potentially less reliable because no maps are available before 1965 for the Northwest and Northeast regions. In row 18, dropping the 1980s, the one period that use road changes from the 1960s, changes the 30-year effect little. It does however shift more of the effect from the contemporaneous period to the prior ten years. Conversely, there are fewer changes in roads documented in the 1990s and 2000s, and this may be partially due to poorer documentation. Dropping the 2000s, and the 1990s and 2000s, in rows 19 and 20, respectively, also has little overall effect, with some redistribution of effects between decades.

Other assumptions Table 7 varies other assumptions used in the empirical analysis. In the baseline analysis, cells are considered contiguous if they share an edge or a vertex (queen contiguity). Row 1 assumes that cells that share only a vertex are not contiguous (rook contiguity), so that travel between them must go through a third cell. This has little effect on results. Rows 2–9 use different values of the trade elasticity σ than the baseline 3.8 in calculating market access. Higher values imply that transport costs are a more powerful impediment to trade. The effect of market access on population decreases as $sigma$ increases, but remains precisely estimated. Importantly, the distribution of $\Delta \ln MA$ compresses substantially as σ increases. Thus, the effect of a standard deviation increase in market access never deviates by more than 28% from the $\sigma = 3.8$ magnitude within the range $\sigma = 1$ to $\sigma = 12$.[25]

The remainder of Table 7 expands the sample with various assumptions about the populations of cities in the years when their population drops below 10,000. varies the way that populations under 10,000 are constructed in the ever-populated sample. Row 10 uses all cell-years with non-zero urban population values. Rows 11 and 12 do this but only for the one and two years, respectively, prior to crossing the 10,000 threshold. The idea is that the longer the period a city has had its population recorded, despite it not being above 10,000, the more likely it is to be special in some unrecorded way, such as a regional capital or a resource base.

To fill in populations further in rows 13–15, we first index time for each cell with respect to the first year it crossed 10,000 urban residents. We then calculate sample statistics for each year relative to crossing. Rows 13 and 14 only assign the sample mean and median (respectively) to those cities with a missing population estimate, but not those cities with a population estimate under 10,000. Finally, Row 15 assigns the relative-year-specific mean to all missing city-year populations. In each case, the overall 30-year effect remains between about 3 and 5%.

[25]Donaldson and Hornbeck (forthcoming) find a similar pattern when varying σ in their context.

5 Results: Heterogeneous Effects

Table 8 explores heterogeneity of the OLS results with respect to city size and location.[26] In the first column, results are stronger for cities above the median population (21,116) in the initial year.[27] However in columns 2 and 3, it is stronger for cities with worse roads in their cell and lower market access. Consistently with this, in columns 5 and 6, it is stronger for cities that are more remote, in the sense that they are farther from the three largest cities, or closer to the border, respectively. Somewhat surprisingly in this context, effects are larger near the coast in column 4. When effects both above and below the median are both significantly different from zero at at least 10% (columns 1, 4, 5 and 6), they differ by a factor of between 2 and 3, but generally remain within the 0.02–0.05 elasticity range, and differences are only significant at the 10 percent level in columns 1, 3 and 4.

Table 9 reports analogous results with the outside 10 cells instrument. The instruments are notably weaker in this context with six instruments and six endogenous variables. The sign of the differential differs from its OLS analog in columns 3, 4 and 5, but only in column 4 is the difference significant, even at 10 percent. While in column 6, the differential between places near and far from a border is large and significant, the instruments are more likely to be weak.

6 Conclusion

We find that road construction in Africa since 1960 has accelerated city growth, much like earlier construction of other transport infrastructure, albeit somewhat less so. The effect is concentrated 10–30 years after road construction, closer to coasts and borders, and farther from a country's largest cities. Several mechanisms could be driving this. Most theoretical and empirical work has focused on reductions in the cost of transporting goods. However, Bryan et al. (2014) show that reduced intercity transport costs also encourage the flow of information and labor. Future work will be needed to disentangle these channels.

References

Acemoglu, Daron, Simon Johnson, and James Robinson (2005) "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic*

[26] Heterogeneity of IV results will be explored in a future draft.

[27] Results using a decade-specific median are very similar and available upon request.

Review, Vol. 95, No. 3, pp. 546–579.

African Development Bank and United Nations Economic Commission For Africa (2003) "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.

Ahlfeldt, Gabriel, Steve Redding, Daniel Sturm, and Niko Wolf (forthcoming) "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*.

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

Atkin, David and Dave Donaldson (2015) "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," working paper, MIT.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012) "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Papers 17897, National Bureau of Economic Research, Inc.

Baum-Snow, Nathaniel (2007) "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, Vol. 122, No. 2, pp. 775–805.

Baum-Snow, Nathaniel, Loren Brandt, Vernon Henderson, Matthew Turner, and Qinghua Zhang (2015) "Roads, Railroads and Decentralization of Chinese Cities," mimeo, Brown University.

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

Bird, Julia and Stéphane Straub (2015) "The Brasília Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil."

Bogart, Dan (2009) "Turnpike trusts and property income: new evidence on the effects of transport improvements and legislation in eighteenth-century England," *Economic History Review*, Vol. 62, No. 1, pp. 128–152.

Bruckner, Markus (2012) "Economic Growth, Size of the Agricultural Sector, and Urbanization in Africa," *Journal of Urban Economics*, Vol. 71, pp. 26–36.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2014) "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, Vol. 82, No. 5, pp. 1671–1748.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel (2015) "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, Vol. 105, No. 6, pp. 1817–51.

Canning, David (1998) "A Database of World Stocks of Infrastructure: 1950-1995," *The World Bank Economic Review*, Vol. 12, pp. 529–548.

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

Castells-Quintana, David (2014) "Malthus living in a slum: Urban concentration, infrastructures and economic growth."

Chaleard, Jean-Louis, Chantal Chanson-Jabeur, and Chantal Beranger (2006) *Le chemin de fer en Afrique*, Paris: Karthala Editions.

Chandra, Amitabh and Eric Thompson (2000) "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, Vol. 30, No. 4, pp. 457–490.

Chaves, Isaias, Stanley L. Engerman, and James A. Robinson (2014) "Reinventing the Wheel: The Economic Benefits of Wheeled Transportation in Early Colonial British West Africa," in Emmanuel Akyeampong, Robert H. Bates, Nathan Nunn, and James Robinson eds. *Africa's Development in Historical Perspective*: Cambridge University Press, pp. 321–365.

Coşar, A. Kerem and Banu Demir (2015) "Domestic Road Infrastructure and International Trade: Evidence from Turkey," mimeo, University of Chicago.

De Long, J Bradford and Andrei Shleifer (1993) "Princes and Merchants: European City Growth before the Industrial Revolution," *Journal of Law and Economics*, Vol. 36, No. 2, pp. 671–702.

Dittmar, Jeremiah E. (2011) "Information Technology and Economic Change: The Impact of The Printing Press," *The Quarterly Journal of Economics*, Vol. 126, No. 3, pp. 1133–1172.

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

Donaldson, Dave and Richard Hornbeck (forthcoming) "Railroads and American Economic Growth: A 'Market Access' Approach," *Quarterly Journal of Economics.*

Faber, Benjamin (2014) "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, Vol. 81, No. 3, pp. 1046–1070.

FAO (2001) "Permanent Crop and Arable Land of Sub-Saharan Africa."

Farr, Tom G., Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas Alsdorf (2007) "The Shuttle Radar Topography Mission," *Reviews of Geophysics*, Vol. 45, No. 2.

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

Feyrer, James (2009) "Distance, Trade, and Income The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," NBER Working Papers 15557, National Bureau of Economic Research, Inc.

Gertler, Paul J., Tadeja Gracner, Marco Gonzalez-Navarro, and Alex Rothenberg (2014) "Road Quality and Local Economic Activity: Evidence from Indonesia's Highways," mimeo, Berkeley.

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

Gould, Peter R. (1960) *The Development of the Transportation Pattern in Ghana*, Chicago: Northwestern University Press.

Greene, William H. (2008) *Econometric Analysis*, Saddle Hill, NJ: Prentice Hall, 6th edition.

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

Haines, M.R. and R.A. Margo (2008) "Railroads and local economic development: the United States in the 1850s," in J.L. Rosenbloom ed. *Quantitative Economic History: The Good of Counting*: Routledge, pp. 78–99.

Hanson, Gordon (1998) "Regional adjustment to trade liberalization," *Regional Science and Urban Economics*, Vol. 28, No. 4, pp. 419–444.

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

Henderson, Vernon, Adam Storeygard, and Uwe Deichmann (2015) "Is climate change driving urbanization in Africa?" mimeo, Tufts University.

Jedwab, Rémi, Edward Kerby, and Alexander Moradi (forthcoming) "History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya ," *Economic Journal*.

Jedwab, Rémi and Alexander Moradi (forthcoming) "The Permanent Economic Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics and Statistics*.

Limão, Nuno and Anthony Venables (2001) "Infrastructure, geographical disadvantage, transport costs, and trade," *The World Bank Economic Review*, Vol. 15, No. 3, pp. 451–479.

Michaels, Guy (2008) "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, Vol. 90, No. 4, pp. 683–701.

Nelson, Andrew and Uwe Deichmann (2004) "African Population database documentation,"Technical report, United Nations Environment Programme and CIESIN at Columbia University.

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

Redding, Stephen J. and Daniel M. Sturm (2008) "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, Vol. 98, No. 5, pp. 1766–97.

Redding, Steven J. and Matthew A. Turner (2015) "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.

Rothenberg, Alexander D. (2013) "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways," mimeo, RAND Corporation.

Storeygard, Adam (2015) "Farther on Down the Road: Transport Costs, Trade and Urban Growth in sub-Saharan Africa," mimeo, Tufts University.

United Nations (2015) *World Urbanization Prospects: The 2014 Revision*, New York: United Nations.

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

Willmott, Cort J. and Kenji Matsuura (2009) "Terrestrial Precipitation: 1900-2008 Gridded Monthly Time Series (1900 - 2008) (V 2.01)."

Young, Alwyn (2012) "The African Growth Miracle," *Journal of Political Economy*, Vol. 120, No. 4, pp. 696 – 739.

——— (2013) "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, Vol. 128, No. 4, pp. 1727–1785.

Table 1: Speeds assumed in city-to-city distance calculations

Surface type	Speed (km/hour)
Highway	80
Paved	60
Improved	40
Earthen	12
No known road	6

Table 2: Descriptive statistics

Main Variable:	Mean	Std. Dev.	Min	Max
$\Delta_{t-10}^{t} \ln$ urban pop	0.318	0.209	-1.533	2.343
$\Delta_{t-10}^{t} \ln MA$	0.101	0.511	-3.94	4.76
$\Delta_{t-20}^{t-10} \ln MA$	0.233	0.732	-3.94	6.11
$\Delta_{t-30}^{t-20} \ln MA$	0.367	0.851	-3.69	6.11
ln urban pop_{t-10}	10.247	0.990	9.21	15.90

Table 3: OLS estimates of the effect of market access on urban population, 1960–2010

Dep. variable:	$(\Delta_{t-10}^{t} \ln \text{urban population})/100$					
	(1)	(2)	(3)	(4)	(5)	(6)
$r\Delta_{t-10}^{t} \ln MA$	-0.16	-0.28	0.91	0.21	1.23**	1.69**
	(0.45)	(0.47)	(0.57)	(0.80)	(0.61)	(0.79)
$r\Delta_{t-20}^{t-10} \ln MA$		0.92**	1.07***	1.71***	1.19**	0.24
		(0.37)	(0.39)	(0.51)	(0.49)	(0.60)
$r\Delta_{t-30}^{t-20} \ln MA$			1.19***	1.24***	0.93**	1.18
			(0.37)	(0.43)	(0.45)	(0.78)
$r\Delta_{t-40}^{t-30} \ln MA$				0.51		
				(0.43)		
$r\Delta_{t}^{t+10} \ln MA$					-1.02	-0.83
					(0.84)	(0.98)
$r\Delta_{t+10}^{t+20} \ln MA$						2.60
						(2.35)
Overall effect	-0.16	0.64	3.17***	3.67***	3.34***	3.11**
($t-40$ to t)	(0.45)	(0.73)	(0.97)	(1.39)	(1.11)	(1.36)
Observations	5,906	5,472	4,725	3,630	2,607	1,094
Cells	2,127	2,127	2,126	2,124	1,515	1,094
Adj. R-squared	0.26	0.21	0.19	0.18	0.21	0.22

Notes: Each column is a separate OLS regression of $\Delta \ln$ urban population$_{t-10}^{t}$ on the change in market access measures shown, where t indexes years 1960 to 2010. Change in market access for city i, $\Delta_{t-10}^{t} \ln MA_i = \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ijt}^{\sigma}) - \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ij,t-10}^{\sigma})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and $\sigma = 3.8$ is the trade elasticity. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Robust standard errors, clustered by cell, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 4: Market access and urban population: additional controls and instrumental variables, 1960–2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Control:		*Instrumental variable (IV):*			
		Own cost	Exclude 5	Exclude 10	Exclude 15	Exclude 20	Foreign
Δ_{t-10}^{t} MA	0.91	0.31	2.91***	3.95**	4.14	5.44	1.71
	(0.57)	(0.80)	(1.05)	(1.83)	(3.31)	(4.79)	(3.29)
Δ_{t-20}^{t-10} MA	1.07***	0.99*	3.73***	6.04***	7.24**	9.25**	2.91
	(0.39)	(0.57)	(0.98)	(1.74)	(2.80)	(3.63)	(1.97)
Δ_{t-30}^{t-20} MA	1.19***	1.53***	3.06***	4.49***	7.40***	8.03***	2.86*
	(0.37)	(0.49)	(0.92)	(1.59)	(2.44)	(2.77)	(1.61)
IV F-Stat			154.01	47.49	26.19	17.53	68.91
Overall effect	3.17***	2.83**	9.70***	14.48***	18.77***	22.72***	7.48
($t-30$ to t)	(0.97)	(1.31)	(2.19)	(3.80)	(5.86)	(7.55)	(4.68)

Notes: Each column in each panel is a separate regression of ($\Delta_{t-10}^{t}\ln$ urban population)/100 on the change in market access measures shown, where t indexes years 1960 to 2010, for 4,725 cell-years. Change in market access for city i, $\Delta_{t-10}^{t}\ln MA_i = \ln(\sum_{j\neq i} pop_{j,t-10}\tau_{ijt}^{\sigma}) - \ln(\sum_{j\neq i} pop_{j,t-10}\tau_{ij,t-10}^{\sigma})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and $\sigma = 3.8$ is the trade elasticity. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Column 2 further controls for the cost of traversing the cell. In columns 3–6 measures of market access change that exclude road surface changes within the radius shown instrument for the main market access change measures. In column 7, measures of market access change that exclude road surface changes within the same country instrument for the main market access change measures. Robust standard errors, clustered by cell, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 5: Robustness: alternate measures and controls

	Δ_{t-10}^{t} MA		Δ_{t-20}^{t-10} MA		Δ_{t-30}^{t-20} MA		Sum Effects	
Specification:								
(1) Baseline	0.97	(0.56)	1.22***	(0.40)	1.22***	(0.38)	3.42***	(0.97)
(2) Cluster by country	0.97	(0.49)	1.22***	(0.65)	1.22***	(0.45)	3.42**	(1.30)
(3) Cluster by country-year	0.97	(0.73)	1.22***	(0.60)	1.22***	(0.55)	3.42**	(1.30)
(4) 0-5	0.75	(0.49)	1.13***	(0.43)	0.94***	(0.30)	3.95***	(1.04)
(5) 5-10	0.31	(0.33)	0.40*	(0.23)	0.42*	(0.25)		
(6) ΔMA Fixing Pop$_{1960}$	1.05*	(0.61)	1.05**	(0.43)	1.08***	(0.40)	3.19***	(1.04)
(7) ΔMA incl. Δ_{t-10}^{t} ln pop	1.66***	(0.35)	1.23***	(0.26)	0.81***	(0.23)	3.70***	(0.60)
(8) Domestic Only (Infinite Border Costs)	1.03**	(0.52)	1.29***	(0.39)	1.25***	(0.38)	3.58***	(0.94)
Add controls to baseline:								
(9) ΔMA Fixing Roads$_{t-10}$	0.92*	(0.56)	1.13***	(0.40)	1.13***	(0.37)	3.18***	(0.96)
(10) Δ ln road cost (round up)	0.48	(0.70)	1.36**	(0.56)	1.37***	(0.47)	3.22**	(1.30)
(11) Δ ln road cost (round down)	0.44	(0.80)	1.17**	(0.56)	1.58***	(0.50)	3.18**	(1.32)
(12) $1(\Delta RC > 0)$ & $1(\Delta RC < 0)$	0.18	(0.82)	1.07*	(0.57)	1.11***	(0.38)	2.35*	(1.29)
(13) ADL(2,2)	0.74	(0.62)	0.87*	(0.51)	0.94*	(0.54)	4.20**	(1.67)
(14) No country FE	1.33**	(0.55)	1.05**	(0.41)	1.19***	(0.40)	3.56***	(0.97)
(15) Geographic controls	1.01*	(0.56)	1.27***	(0.40)	1.19***	(0.37)	3.47***	(0.96)

Notes: Each row is a separate OLS regression of (Δ_{t-10}^{t} ln urban population)/100 on the change in market access measures shown, where t indexes years 1960 to 2010. Baseline change in market access for city i, $\Delta_{t-10}^{t} \ln MA_i = \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ijt}^{\sigma}) - \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ij,t-10}^{\sigma})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and $\sigma = 3.8$ is the trade elasticity. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Robust standard errors, clustered by cell except where indicated, are in brackets. In rows 2 and 3, errors are clustered by country and country-year, respectively. Rows 4 and 5 report results for one regression in which decadal road changes are split into the first and last five years of the decade. In row 6, $\Delta_{t-10}^{t} \ln MA_i$ is calculated using $pop_{j,1960}$ in place of $pop_{j,t-10}$. In row 7, $\Delta_{t-10}^{t} \ln MA_i$ is calculated using $pop_{j,t}$ in place of $pop_{j,t-10}$ in the end year. In row 8, market access is calculated only to domestic cities. Row 9 controls for the change in market access due to changes in population of other cities, holding roads constant. Row 10 controls for each cell's own road cost, rounding up to the worse road when a year falls between map years. Row 11 controls for each cell's own road cost, rounding down to the better road when a year falls between map years. Row 12 includes indicator variables for cells decreasing and increasing road costs. Row 13 is an auto-distributed lag model, controlling for two lags of population. Row 14 replaces country-year fixed effects with year fixed effects only. Row 15 adds the following geographic controls: a coastal dummy, distance to the coast (5 km minimum), a river dummy, distance to a river (5 km minimum), mean altitude, standard deviation of altitude, mean precipitation, shares of class 1, class 2, class 3, undetermined, and sparsely vegetated soils and water in the cell. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 6: Robustness: sampling

	Δ_{t-10}^{t} MA		Δ_{t-20}^{t-10} MA		Δ_{t-30}^{t-20} MA		Sum Effects	
(1) Weights ln Pop in $t-10$	0.99*	(0.56)	1.21***	(0.40)	1.25***	(0.38)	3.45***	(0.98)
(2) Drop top 1% $\Delta \ln$ MA	0.25	(0.81)	1.88***	(0.49)	1.53***	(0.47)	3.67***	(1.25)
(3) Drop top 5% $\Delta \ln$ MA	1.17	(1.49)	1.91**	(0.88)	1.77**	(0.71)	4.84**	(2.19)
(4) Drop top 10% $\Delta \ln$ MA	2.45	(2.11)	2.36	(1.45)	1.71*	(1.01)	6.52*	(3.38)
(5) Drop top 1% $\Delta \ln$ Pop	0.80	(0.54)	1.07***	(0.39)	0.91***	(0.33)	2.79***	(0.94)
(6) Drop top 5% $\Delta \ln$ Pop	1.00**	(0.41)	0.83**	(0.34)	0.76***	(0.29)	2.58***	(0.75)
(7) Drop top 10% $\Delta \ln$ Pop	0.79**	(0.38)	0.61*	(0.32)	0.80***	(0.27)	2.21***	(0.69)
(8) Drop Capital & top 2 cities 1960	1.08*	(0.57)	1.22***	(0.41)	1.12***	(0.38)	3.42***	(0.98)
(9) Drop (8) & Capital + top 2 cities 2010	0.89	(0.57)	1.11***	(0.41)	1.07***	(0.37)	3.07***	(0.97)
(10) Drop (9) & top 5 cities 1960	0.89	(0.59)	1.20***	(0.42)	0.98***	(0.37)	3.07***	(1.00)
(11) Drop (9) & top 15 cities 1960	0.99	(0.63)	1.22***	(0.44)	1.11***	(0.38)	3.31***	(1.07)
(12) Drop (9) & top decile continent	1.12*	(0.60)	1.28***	(0.46)	1.05***	(0.40)	3.46***	(1.07)
(13) Drop (9) & top decile country	1.02*	(0.59)	1.27***	(0.45)	1.10***	(0.39)	3.39***	(1.04)
(14) Drop (9) & ctrl dist to cities dropped	0.75	(0.57)	1.00**	(0.41)	1.03***	(0.37)	2.79***	(0.98)
(15) Drop South Africa neighbors	1.09*	(0.56)	1.48***	(0.40)	1.47***	(0.38)	4.04***	(0.98)
(16) Drop North Africa neighbors	0.98	(0.62)	1.12***	(0.40)	1.33***	(0.41)	3.43***	(1.04)
(17) Drop Arabian Peninsula neighbors	1.01*	(0.60)	0.95**	(0.40)	1.05***	(0.36)	3.02***	(0.99)
(18) Drop 1980s	0.29	(0.77)	1.80***	(0.52)	1.25***	(0.44)	3.34***	(1.26)
(19) Drop 2000s	1.15*	(0.60)	1.32***	(0.49)	0.85*	(0.46)	3.33***	(1.09)
(20) Drop 2000s and 1990s	1.66**	(0.78)	0.44	(0.62)	1.16	(0.75)	3.26**	(1.33)

Notes: Each row is a separate OLS regression of (Δ_{t-10}^{t} ln urban population)/100 on the change in market access measures shown, where t indexes years 1960 to 2010. Baseline change in market access for city i, $\Delta_{t-10}^{t} \ln MA_i = \ln(\sum_{j \neq i} pop_{j,t-10}\tau_{ijt}^{\sigma}) - \ln(\sum_{j \neq i} pop_{j,t-10}\tau_{ij,t-10}^{\sigma})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and $\sigma = 3.8$ is the trade elasticity. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Robust standard errors, clustered by cell, are in brackets. Row 1 weights the sample by population in $t-10$. Rows 2–4 drop the 1, 5 and 10 percent of cells with the fastest growth in $\ln MA$, and rows 5–7 do the same for $\ln pop$. Row 8 drops cells containing the capital and two largest cities in each country, as reported in 1960. Starting from the row 8 sample, row 9 additionally drops cells containing the equivalent cities for 2010. Starting from the row 9 sample, rows 10 and 11 drop the cells containing the 5 and 15 largest cities in 1960, respectively. Starting from the row 9 sample, rows 12 and 13 drop the top decile of cities in the continent and in each country, respectively. Row 14 uses the row 9 sample and controls for the distance to the capital and largest two cities in 1960 and 2010. Rows 15–17 drops the sample countries neighboring South Africa (Botswana, Mozambique, Namibia, and Zimbabwe), North Africa (Mauritania, Mali, Niger, Chad, Sudan), and the Arabian Peninsula (Sudan, Eritrea, Djibouti, Somalia), respectively. Rows 18–20 drop the 1980s, the 2000s, and the 1990s and 2000s, respectively. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 7: Robustness: alternative assumptions

	Δ^{t}_{t-10} MA		Δ^{t-10}_{t-20} MA		Δ^{t-20}_{t-30} MA		Sum Effects	
Market access parameters:								
(1) Rook contiguity	1.30**	-0.54	1.14***	(0.38)	0.99***	(0.36)	3.43***	(0.92)
(2) $\sigma = 1$	12.40*	(7.29)	8.65*	(4.75)	9.55**	(3.98)	30.60***	(11.82)
(3) $\sigma = 2$	2.02	(1.92)	2.28**	(1.14)	2.56**	(1.09)	6.86**	(3.07)
(4) $\sigma = 3$	1.27	(0.83)	1.62***	(0.57)	1.58***	(0.55)	4.48***	(1.42)
(5) $\sigma = 4$	0.89*	(0.52)	1.17***	(0.37)	1.15***	(0.35)	3.21***	(0.90)
(6) $\sigma = 6$	0.51*	(0.30)	0.70***	(0.22)	0.76***	(0.21)	1.97***	(0.53)
(7) $\sigma = 8$	0.36	(0.22)	0.50***	(0.16)	0.57***	(0.15)	1.43***	(0.38)
(8) $\sigma = 10$	0.29*	(0.17)	0.39***	(0.13)	0.46***	(0.12)	1.14***	(0.30)
(9) $\sigma = 12$	0.24*	(0.14)	0.32***	(0.10)	0.38***	(0.10)	0.94***	(0.25)
Everpopulated sample:								
(10) Pop Estimate Available (N = 7307)	1.77***	(0.56)	1.49***	(0.43)	1.30***	(0.36)	4.56***	(0.97)
(11) Only Year Before (N = 6109)	1.54***	(0.57)	1.24***	(0.40)	1.07***	(0.37)	3.85***	(0.95)
(12) Only Two Years Before (N = 6977)	1.74***	(0.54)	1.25***	(0.37)	1.00***	(0.35)	4.00***	(0.91)
(13) Replacing missing with year mean	1.46***	(0.53)	1.29***	(0.43)	1.31***	(0.36)	4.05***	(0.87)
(14) Replacing missing with year median	1.43***	(0.53)	1.31***	(0.42)	1.30***	(0.35)	4.05***	(0.87)
(15) Replacing Pop<10,000 with year mean	0.61*	(0.35)	1.04***	(0.33)	1.31***	(0.29)	2.96***	(0.64)

Notes: Each row is a separate OLS regression of (Δ^{t}_{t-10} ln urban population)/100 on the change in market access measures shown, where t indexes years 1960 to 2010. Baseline change in market access for city i, $\Delta^{t}_{t-10} \ln MA_i = \ln(\sum_{j \neq i} pop_{j,t-10} \tau^{\sigma}_{ijt}) - \ln(\sum_{j \neq i} pop_{j,t-10} \tau^{\sigma}_{ij,t-10})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and σ, the trade elasticity, is set to 3.8 unless otherwise indicated. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Robust standard errors, clustered by cell, are in brackets. In row 1, costs of travelling between cells are calculated assuming rook contiguity instead of queen contiguity. In other words, direct travel is only allowed between cells that share a finite boundary. In rows 2–9, σ is set equal to 1,2,3,4,6,8,10, and 12, respectively. Rows 10–15 define the ever-populated sample and its population values using alternative assumptions. In row 10, all available raw populations are used. In row 11, populations under 10,000 in the year before they cross 10,000 are included, in addition to populations over 10,000. In row 12, populations under 10,000 in the two years before they cross 10,000 are included, in addition to populations over 10,000. In row 13, only missing populations are replaced with the mean for that year (relative to the year of crossing 10,000), and all populations under 10,000 are kept as is. In row 14, only missing populations are replaced with the median for that year (relative to the year of crossing 10,000), and all populations under 10,000 are kept as is. In row 15, all populations missing or under 10,000 are replaced with the mean for that year (relative to the year of crossing 10,000). **, *** mean significance at the ten, five, and one percent level, respectively.

Table 8: OLS estimates: heterogeneous effects of market access on urban population, intensive sample, 1960-2010

Dependent Variable:	Log Growth Rate of City Population (t-10; t)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy for:	High City Size t-10	Paved (incl. High-ways) t-10	High Market Access t-10	Inland Area (Farther from the Coast)	Remote (Farther from Top Three Cities 1960)	Remote (Closer to Border)
ΔMA (t-10;t)	0.54	1.38**	1.43**	1.72**	-0.12	0.81
	(0.66)	(0.61)	(0.63)	(0.83)	(0.61)	(0.59)
ΔMA (t-20;t-10)	0.37	1.34**	2.00***	1.67**	1.10**	1.06**
	(0.52)	(0.57)	(0.57)	(0.68)	(0.48)	(0.47)
ΔMA (t-30;t-20)	1.03**	1.27**	1.36**	2.25***	1.19*	0.51
	(0.49)	(0.50)	(0.57)	(0.79)	(0.63)	(0.48)
ΔMA (t-10;t) x Dummy	0.89	-0.77	-1.59	-1.18	2.49**	0.32
	(1.07)	(1.62)	(1.36)	(1.09)	(1.14)	(1.07)
ΔMA (t-20;t-10) x Dummy	1.81**	-0.33	-1.63**	-0.68	0.22	0.50
	(0.72)	(0.75)	(0.74)	(0.83)	(0.74)	(0.72)
ΔMA (t-30;t-20) x Dummy	0.39	-0.14	-0.27	-1.59*	0.01	1.65**
	(0.77)	(0.71)	(0.71)	(0.87)	(0.75)	(0.74)
Overall effect (t-30;t); Dummy = 0	1.94*	3.99***	4.80***	5.64***	2.17*	2.38**
	(1.14)	(1.24)	(1.22)	(1.58)	(1.16)	(1.05)
Overall effect (t-30;t); Dummy = 1	5.03***	2.75	1.31	2.18*	4.89***	4.85***
	(1.47)	(1.75)	(1.61)	(1.21)	(1.54)	(1.52)
(Dummy = 1) - (Dummy = 0)	3.09*	-1.23	-3.49*	-3.46*	2.72	2.47
	(1.72)	(1.98)	(1.97)	(1.96)	(1.86)	(1.63)

Notes: Each column is a separate OLS regression of $\Delta \ln \text{population}_{t-10}^{t}$ on the change in market access measures shown, where t indexes years 1960 to 2010. Change in market access for city i, $\Delta_{t-10}^{t} \ln MA_i = \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ijt}^{\sigma}) - \ln(\sum_{j \neq i} pop_{j,t-10} \tau_{ij,t-10}^{\sigma})$, where pop_{jt} is the population of city j in year t, τ_{ijt} is the transport cost, in time, between city i and city j in year t, and $\sigma = 3.8$ is the trade elasticity. Overall effect is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln pop_{t-10}$, and fourth order polynomials in longitude and latitude interacted with year fixed effects. Robust standard errors, clustered by cell, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 9: IV estimates: heterogeneous effects of market access on urban population, intensive sample, 1960-2010

Dependent Variable:	Log Growth Rate of City Population (t-10; t)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy for:	High City Size t-10	Paved (incl. High-ways) t-10	High Market Access t-10	Inland Area (Farther from the Coast)	Remote (Farther from Top Three Cities 1960)	Remote (Closer to Border)
IV: Exclude 10						
Overall effect (t-30;t); Dummy = 0	15.72***	16.80***	15.84***	6.67	18.92***	4.06
Overall effect (t-30;t); Dummy = 1	16.94***	12.88**	26.01	20.40***	14.68***	18.26***
(Dummy = 1) - (Dummy = 0)	1.22	-3.92	10.18	13.73*	-4.23	14.21***
First-stage F-stat 22.43	28.02	0.40	13.10	15.10	7.75	7.75

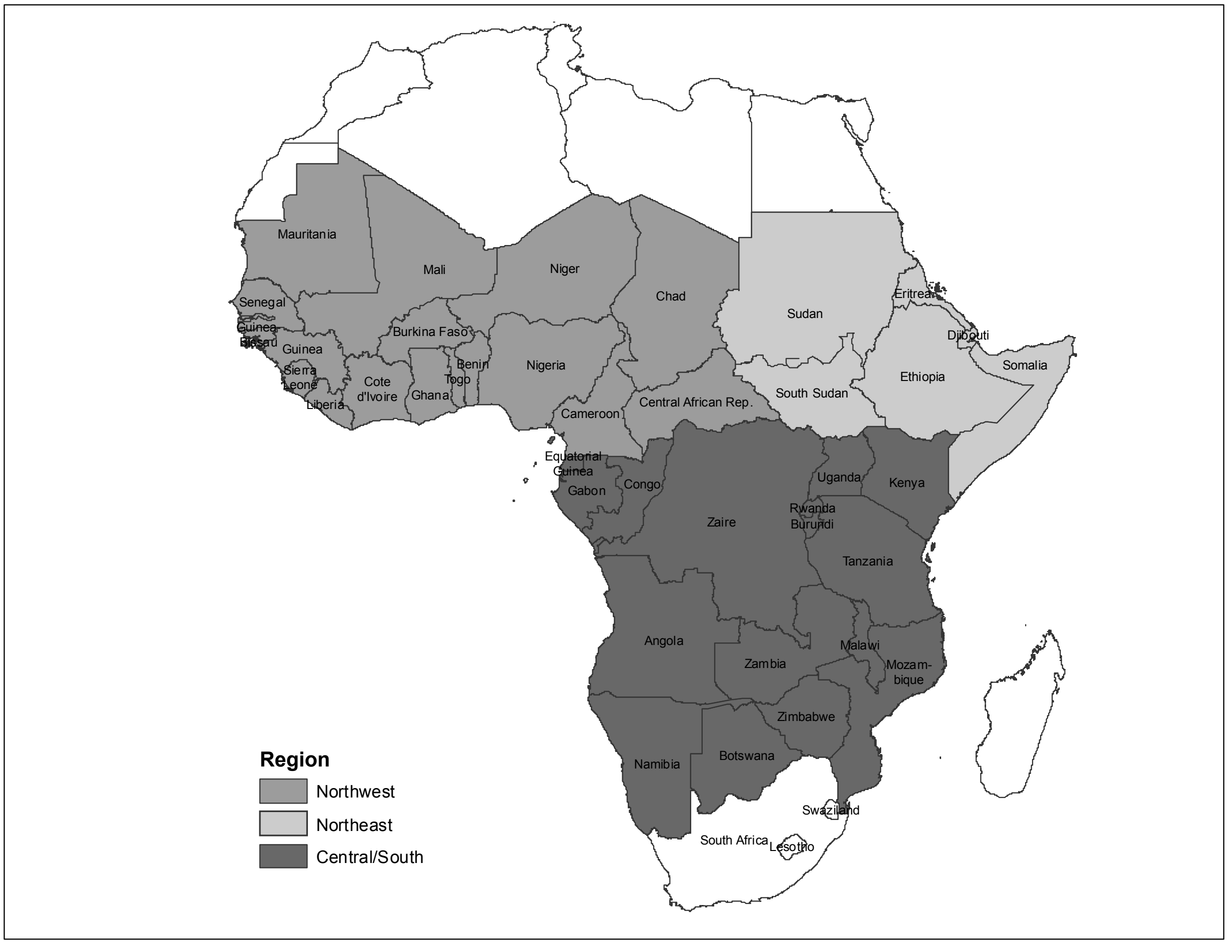

Figure 1: **Map regions** All analysis treats Sudan as one country, as it was for the full population sample period. There are 18 countries in the Northwest, 5 in the Northeast, and 16 in Central/South. South Africa, Swaziland and Lesotho contribute roads (from Central/South maps) and their largest 20, 1, and 1 cities, respectively, to the calculation of market potential for cities in the 39 sample countries, but their cities do not enter the sample.

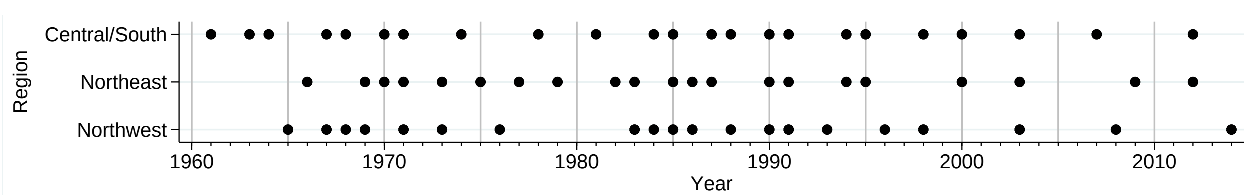

Figure 2: **Map years** There are 20 maps for the Northwest, 21 for the Northeast, and 23 for Central/South.

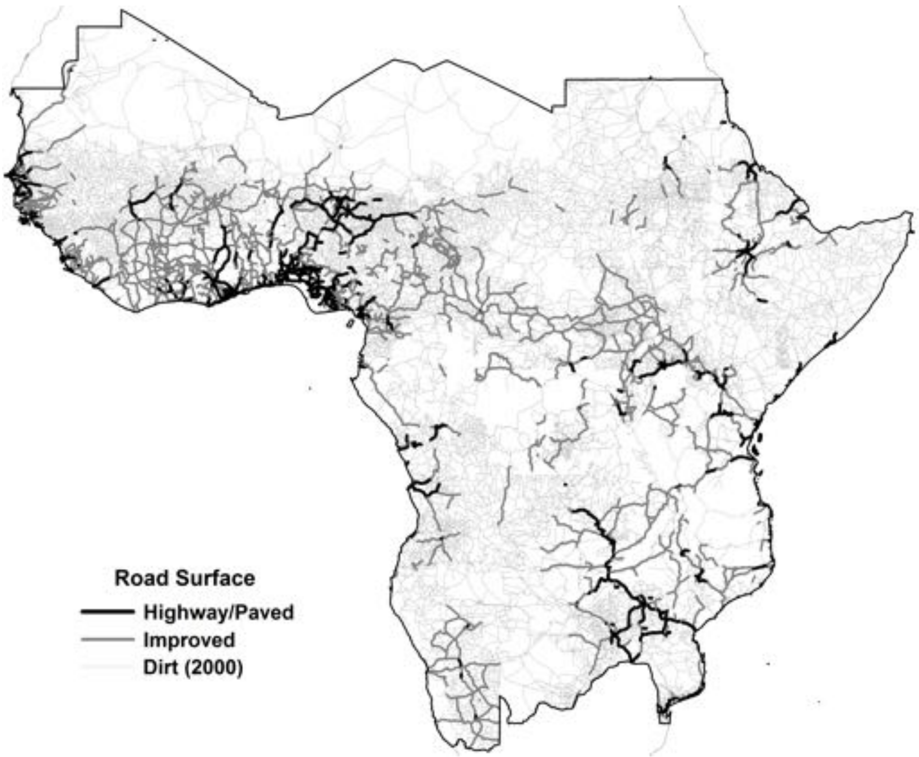

Figure 3: **Roads in 1960**

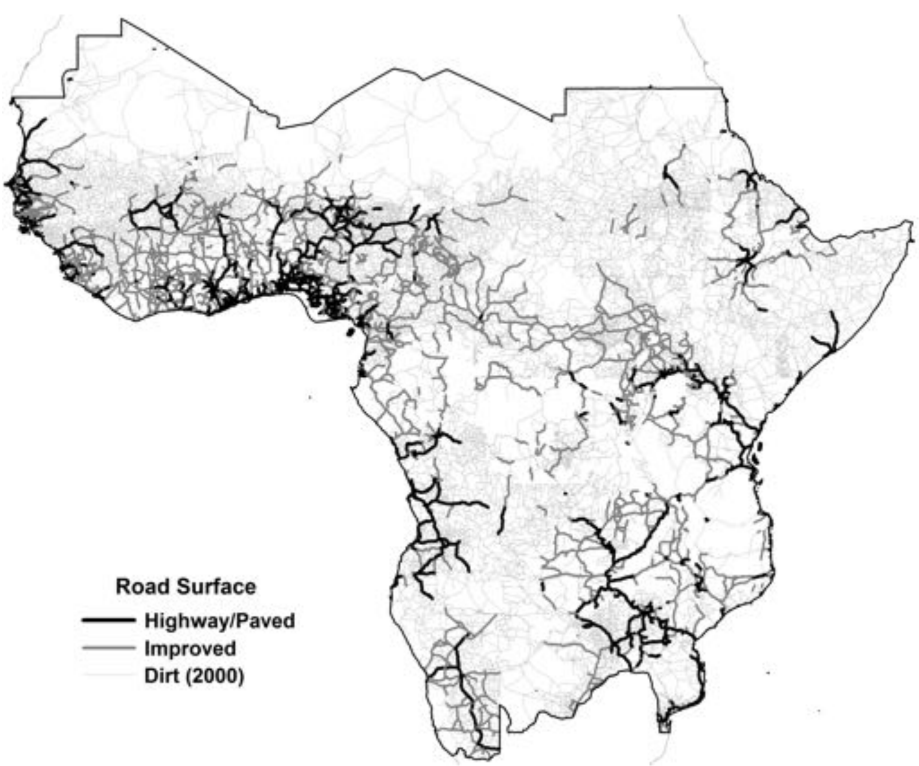

Figure 4: **Roads in 1970**

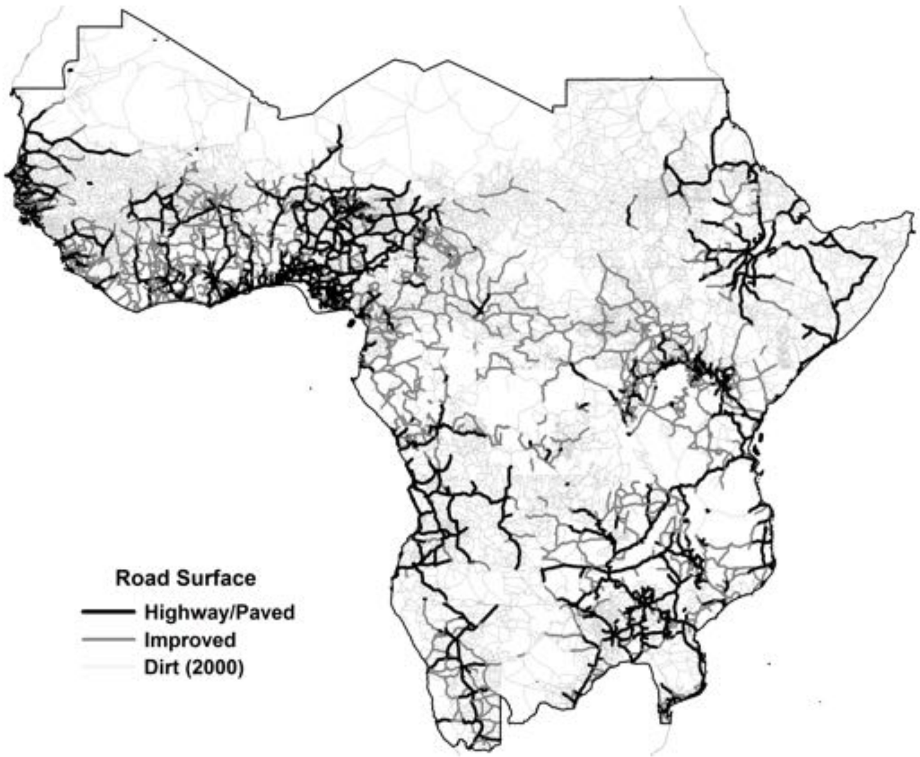

Figure 5: **Roads in 1980**

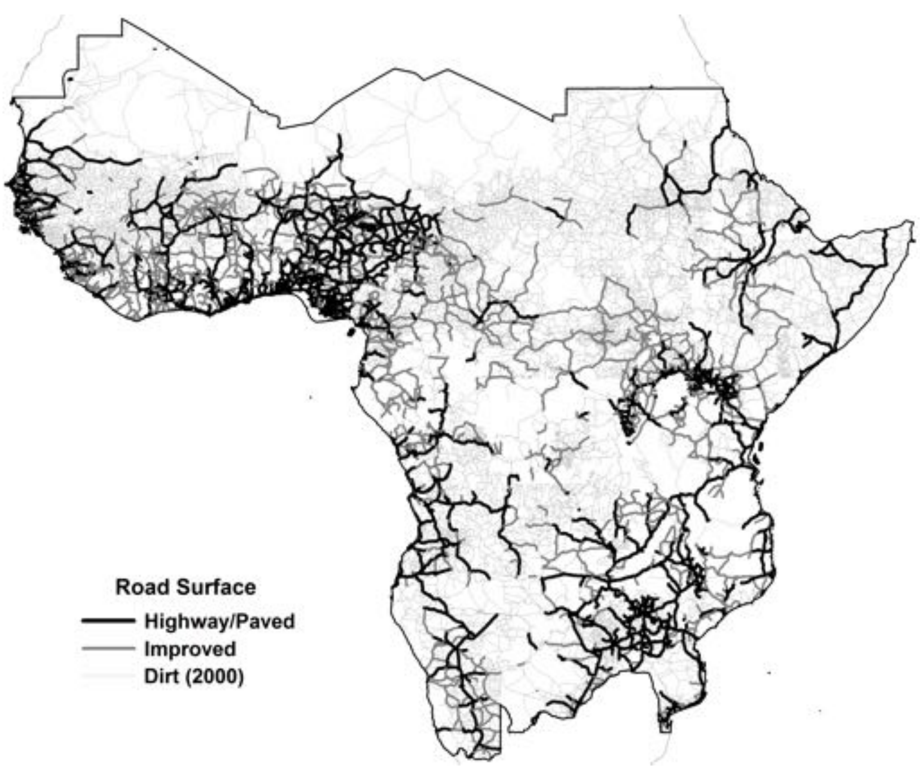

Figure 6: **Roads in 1990**

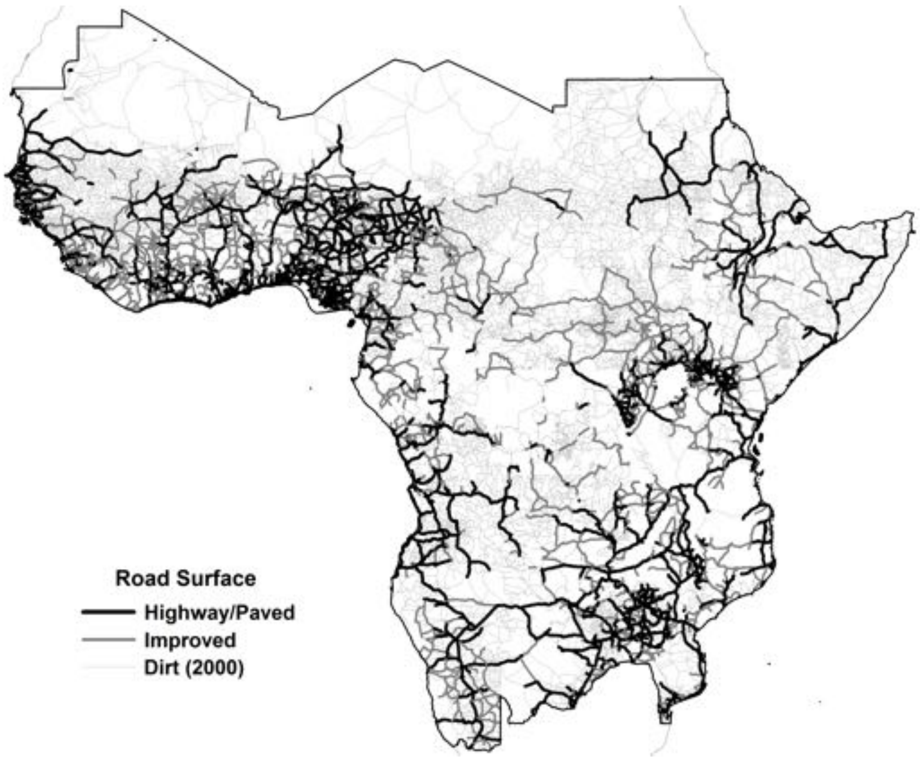

Figure 7: **Roads in 2000**

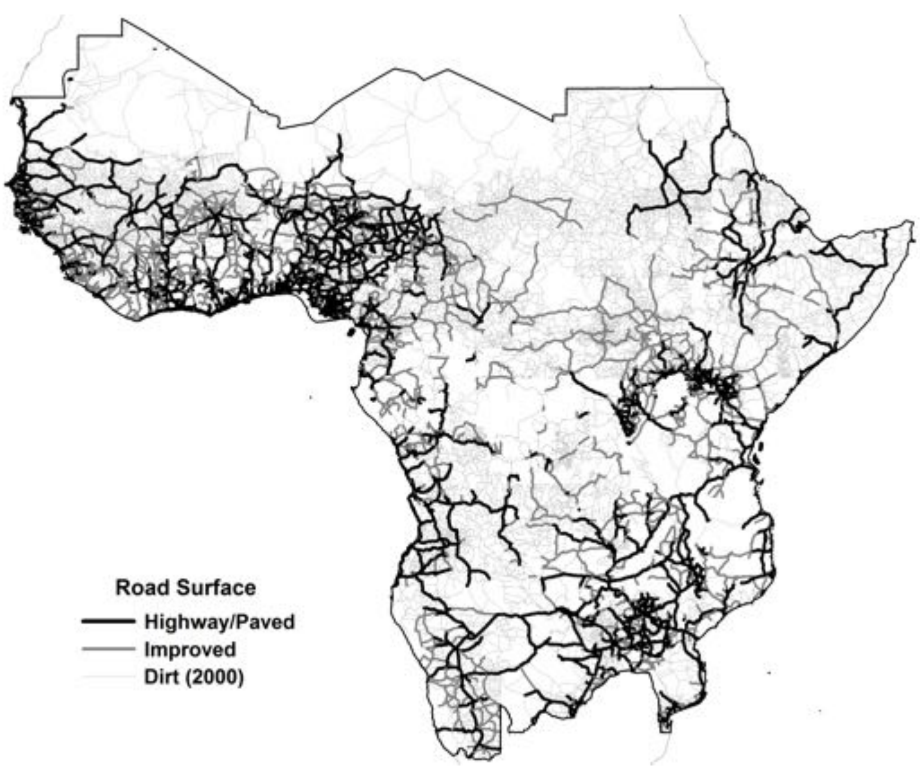

Figure 8: **Roads in 2010**

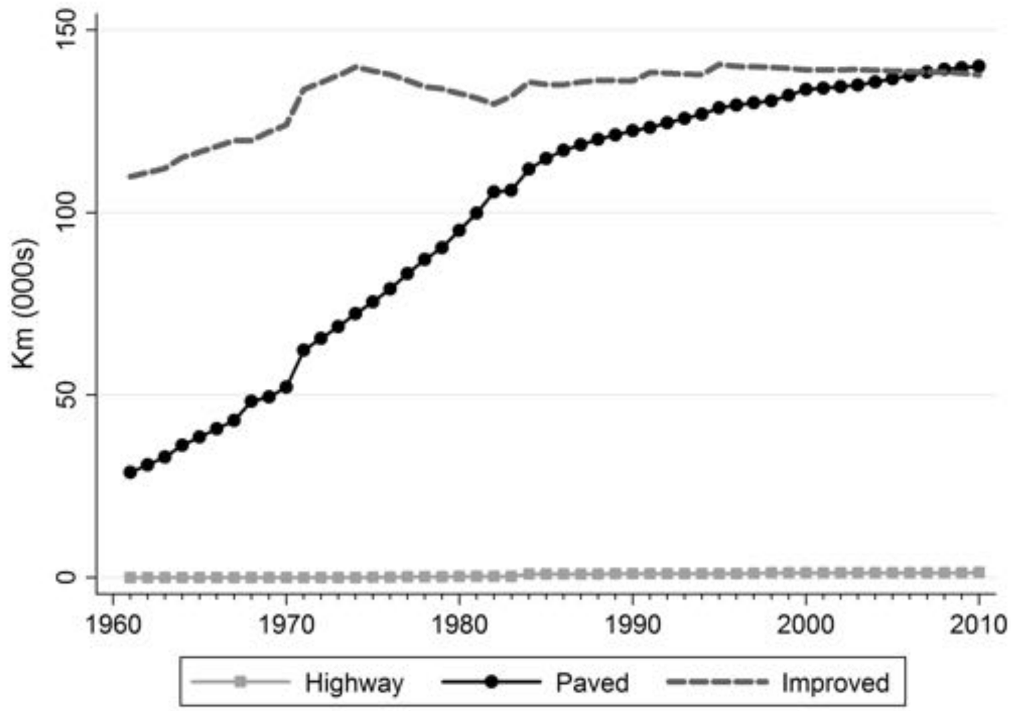

Figure 9: **Length of roads in sample countries, 1961-2010**

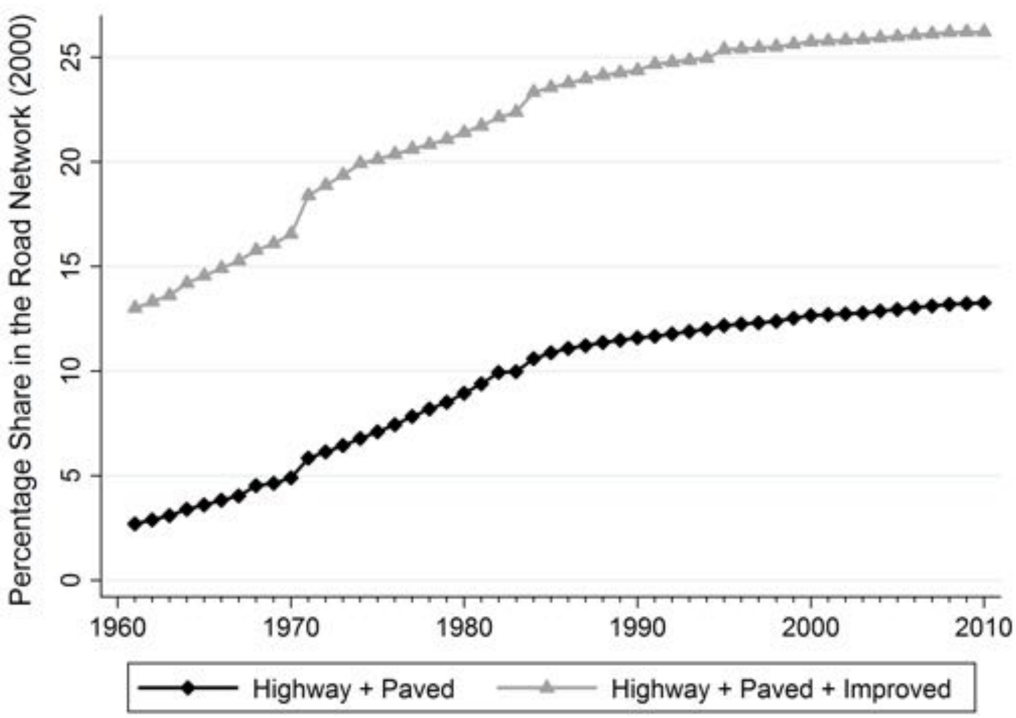

Figure 10: **Fraction of total sample road length in selected classes, 1961-2010**

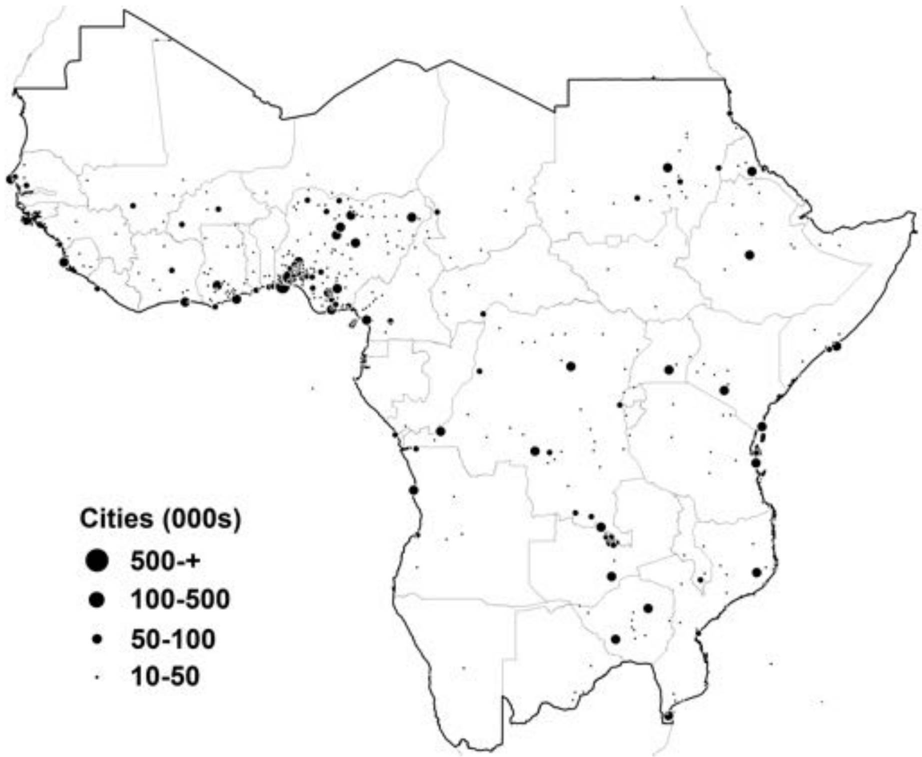

Figure 11: **Cities in 1960**

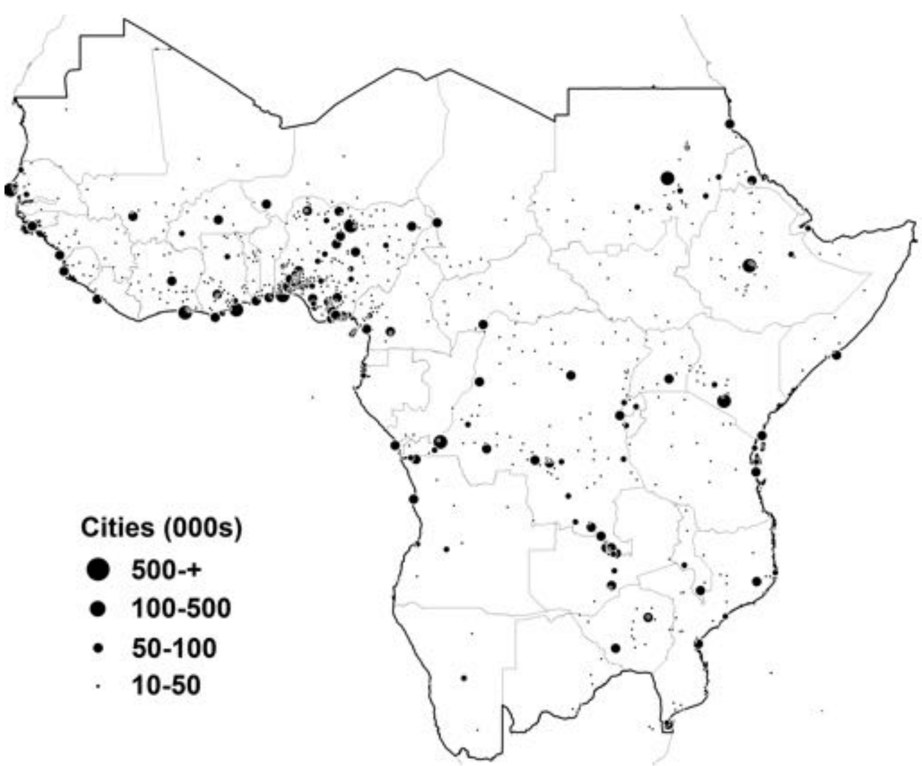

Figure 12: **Cities in 1970**

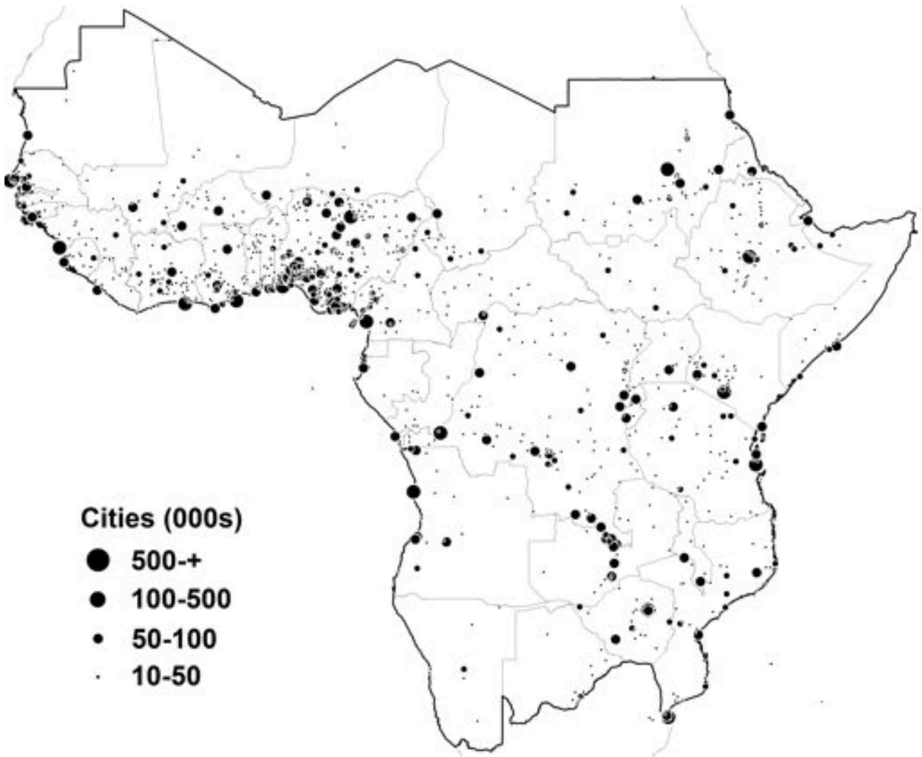

Figure 13: **Cities in 1980**

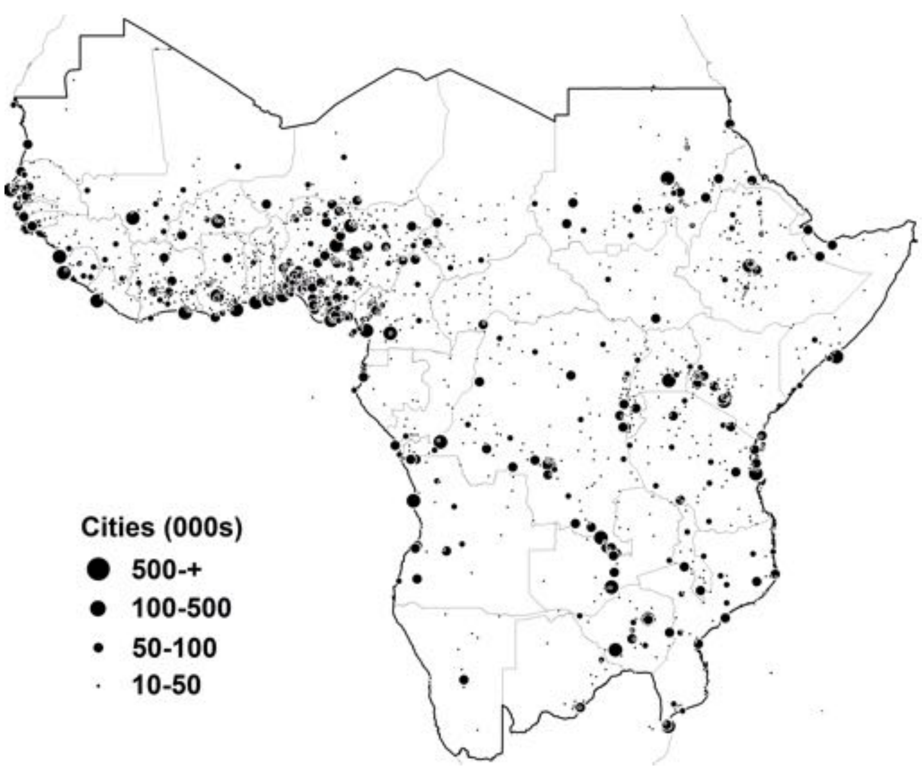

Figure 14: **Cities in 1990**

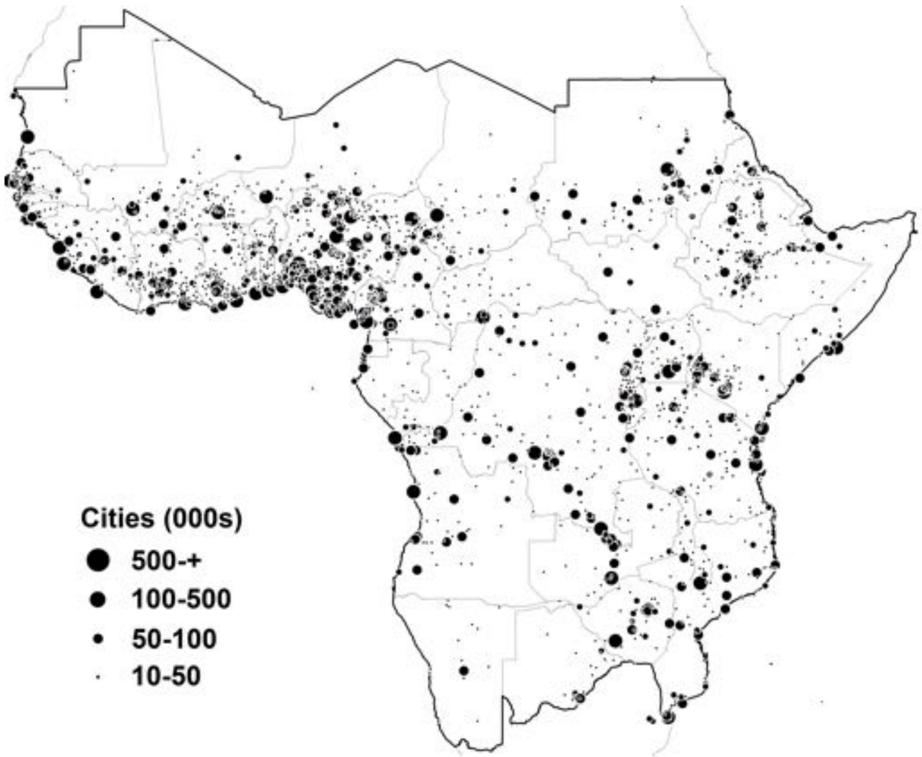

Figure 15: **Cities in 2000**

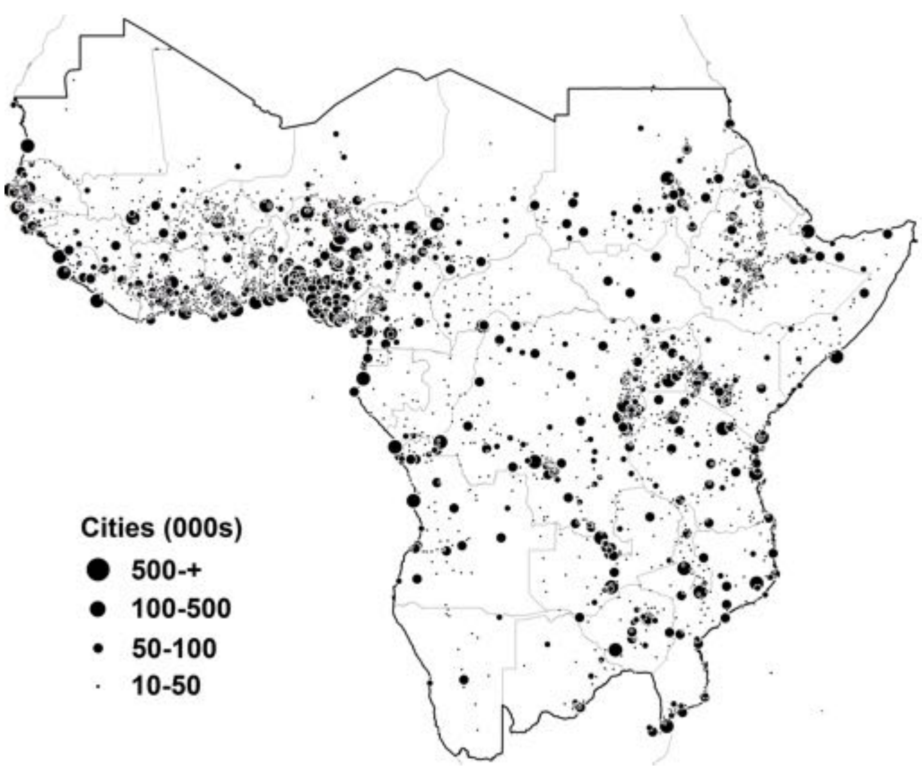

Figure 16: **Cities in 2010**