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

Remi Jedwab and Adam Storeygard*

July 20, 2016

Abstract

Previous work on the effects of transportation investments has focused on estimating average impacts in middle and high income countries. Less is known about the heterogeneity of the impact of these investments, depending on the context in which they take place. We shed light on this by studying how the effects of transportation investments vary as a function of the characteristics of the places they connect. Using new data on roads and cities spanning over 50 years in 39 African countries, we document the effect of road construction on city population growth, through the channel of increasing market access. First, using changes in market access due to distant road construction as a source of exogenous variation in market access, we estimate a 30-year elasticity of city population with respect to market access of 0.05 to 0.20, larger than an OLS estimate of 0.035, but smaller than estimates for other contexts. Second, relying on the same identification strategy, we find suggestive evidence that it is stronger for small and remote cities, and cities that experience a large change in their market access, especially to other cities of the same country or international ports, and weaker for cities in areas with a comparative advantage in agriculture and areas most likely to be favored by ethnic patronage. The observed heterogeneity in the effects confirms the importance of understanding local context when evaluating the impact of transport investment.

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

Keywords: Transportation Infrastructure; Heterogeneity; External Validity; Trade Costs; Roads; Urbanization; Cities; Africa; Market Access; Highways

---

*Remi Jedwab: Department of Economics, George Washington University, jedwab@gwu.edu. Adam Storeygard: Department of Economics, Tufts University, adam.storeygard@tufts.edu. We thank Treb Allen, Costas Arkolakis, Nathaniel Baum-Snow, Gharad Bryan, Victor Couture, Dave Donaldson, Gilles Duranton, Benjamin Faber, Edward Glaeser, Vernon Henderson, Melanie Morten, Mark Rosenzweig, Harris Selod, Anthony Venables, Leonard Wantchekon and seminar audiences at the African School of Economics, Berkeley Haas, Boston Fed, CSAE (Oxford), George Mason, Harvard, IGC (LSE), Merced, NEUDC (Brown), North Dakota, Sciences Po, Stanford, World Bank, the World Bank Secondary Towns Conference, and the World Bank-GWU Urbanization Conference for helpful comments, and Yasmin Abisourour, Karen Chen, Rose Choi, Yury Higuchi, Erin McDevitt, and Emily Ryon for research assistance. We are grateful to Uwe Deichmann and Siobhan Murray for sharing their roads data, François Moriconi-Ebrard for 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, the Institute for International Economic Policy at GWU, and the Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development, for financial assistance.

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/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%2Dand-afdb-success-story-12528/; http://www.afdb.org/en/news-and-events/article/

It is thus imperative 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 population growth during that period, as a result of increased market access to other cities. Using several 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 over the course of the 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 may be 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, 2016; Storeygard, 2016; Jedwab and Moradi, 2016; Donaldson, forthcoming).[5]

More generally, a large literature has looked at how market access affects the growth of neighborhoods (Ahlfeldt et al., 2015), cities (Redding and Sturm, 2008), regions (Hanson, 1998), and countries (Feyrer, 2009). Another large literature has looked at the effect of large highway projects on a variety of outcomes (Baum-Snow, 2007; Rothenberg, 2013; Baum-Snow et al., 2015; Coşar and Demir, 2016). 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., 2015)

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 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

---

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).

experienced in the past 25 years, building over 35,000 km of limited access highways (Baum-Snow et al., 2015, 2016; Faber, 2014). To the extent that gradual evolution is more likely in the future of most developing regions, this is a distinct and instructive context. Second, building on Donaldson and Hornbeck (2016), we use a novel identification strategy, relying on the variation in market access induced by roads built far away, and even 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 the earlier railroad revolution (Jedwab and Moradi, 2016; Jedwab et al., forthcominga) 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 empirical work on urbanization in Africa is primarily cross-country in nature and does not consider variation across cities within countries (Fay and Opal, 2000; Bruckner, 2012; Henderson, Roberts and Storeygard, 2013; Castells-Quintana, 2014; Gollin, Jedwab and Vollrath, 2016; Jedwab, Christiaensen and Gindelsky, forthcomingb).[7]

2. Data and Background

We focus on mainland sub-Saharan Africa, for which we create a new dataset on roads and cities at a precise spatial level over fifty years: 199,814 cells of 0.1x0.1 degrees (11x11km) for 42 countries for the following years: 1960, 1970, 1980, 1990, 2000 and 2010.[8] In our econometric analysis, we will focus on the 2,789 cells that reached an urban population of at least 10,000 at some point since 1960 for 39 of these countries.

2.1. Roads, 1960-2010

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

[6]Faber (2014) finds evidence that negative effects of roads are mitigated in larger and more remote places consistent with a core-periphery model.

[7]A non-transport exceptions is Henderson et al. (2016)

[8]Our unit of analysis covers approximately 123 square kilometers at the equator, decreasing with the cosine of latitude. These are the same units as in Jedwab and Moradi (2016). Because the maximum absolute latitude in our sample is 29 degrees, in practice this creates little variation in unit size.

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). We record road information for all of mainland sub-Saharan Africa, a total of 199,814 cells in 42 countries and 902 country-years. Figures 1 and 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; only 9 maps follow a gap of longer than 3 years. 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.[9]

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 sources to create the maps: (i) the previous Michelin map, (ii) new government road censuses/maps, (iii) direct information from its tire stores across Africa, and (iv) correspondence from road users including truckers.[10] The latter two sources of information are especially important, and we believe new to this literature.[11] Michelin has been producing road 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 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 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.

[9]Several Michelin maps distinguish paved roads with one lane only from paved roads with two lanes or more, but not consistently across years. Many also distinguish fully and partially improved roads, but again not consistently across years. This is why we use only the four categories listed.

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

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

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

2.2. City Location and Population, 1960-2010

We obtained location and population estimates of cities in 33 countries from *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa*.[12] The Africapolis team generated estimates using various sources including population censuses, "non-native" population censuses, demographic studies, administrative counts, electoral counts, and statistical abstracts. Based on an initial list of cities with at least 5,000 inhabitants in the most recent census circa 2000, their final database nominally includes all cities that reached a population of at least 10,000 at some point since 1960. The Africapolis team also defined agglomerations in circa 2000 using satellite imagery. 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) 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 6 countries not in the Africapolis samples (Angola, Botswana, Malawi, Mozambique, Zambia and Zimbabwe). 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 (in 2010) cities in South Africa, Lesotho and Swaziland, to minimize bias in measures for cities near them. 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

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

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.[13] The sources used for all countries are listed in Web Appendix Table **??**.

Population figures for all 42 countries are exponentially interpolated and extrapolated between raw data years to obtain estimates for each year (1960, 1970, 1980, 1990, 2000 and 2010). The resulting sample includes population estimates for all cities with a population of over 10,000 at some point since 1960, in all sample years in which their population exceeded 10,000, and in most sample years in which they did not reach 10,000. While ideally we would prefer to have systematic information on smaller cities, such information are not systematically available for our sample region and period.

2.3. Other data

We use several physical geographic characteristics as control variables following Jedwab and Moradi (2016). Climate data are from Willmott and Matsuura (2009). We calculate average annual precipitation (in mm) over the period 1900 to 1960 for each cell.[14] Elevation is from the Shuttle Radar Topography Mission version 3 (SRTM3 DTED1) 90-meter data (Farr et al., 2007).[15] 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.[16].

We also use several non-physical geographic characteristics as control variables. For each country, we know its capital, largest, and second largest cities in both 1960 and 2010. We also use Law (2016) and Wikipedia (2016) to know which first-level administrative unit, or "province", each cell belonged to in 1960 and in 2010, and whether the cell contained a "provincial" capital in 1960 and in 2010. From Jedwab and Moradi (2016), we also know the location of railroad lines, and when each line was built.

When investigating the heterogeneous effects of market access, we will also employ data on: (i) The imputed total population of each cell in 1960 using GIS data on the distribution of population in Africa from Nelson and Deichmann (2004), (ii) Land

[13]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 and would thus have entered our estimation sample.

[14]Available at http://climate.geog.udel.edu/~climate/html_pages/archive.html

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

[16]Available at http://www.agiweb.org/pubs/globalgis/metadata_qr/perennial_rivers_qk_ref.html

suitability for crops from IIASA and FAO (2012).[17], (iii) The locality of birth, locality of origin and ethnicity for the 189 heads of state of the 39 sample countries between 1960 and 2010. We used as our main sources the English and French versions of Wikipedia (2016), and verified when possible the information using the appendices and/or raw data from Fearon et al. (2007), Hodler and Raschky (2014), Burgess et al. (2015) and Francois et al. (2015), who all focus on selected countries and periods among our sample. To our knowledge, we are the first paper to collect this data for most sub-Saharan African countries from independence to date. Using GIS files of the Murdock (1959) ethnic map and the GREG ethnic map compiled by Weidmann et al. (2010), we know the main ethnic group in each cell according to each of the two group classifications. For each head of state, we then compared the exact location of the locality of birth and the location of the ethnic group to see if the president was indeed born in her ethnic homeland. We find that a few presidents were born abroad or in the capital city. In those cases, the locality of origin differs from the locality of birth, as we assign it the coordinates of the centroid of the ethnic homeland in either the Murdock (1959) map or the GREG map, and (iv) The location of the 44 main international ports in sub-Saharan Africa circa 2010 from Ocean Shipping Consultants, Ltd. (2009).

2.4. Aggregate Patterns in Road Building and Urban Growth

Railroads were the dominant intercity transportation technology in Africa before World War II (Chaleard et al., 2006; Chaves et al., 2014; Jedwab and Moradi, 2016). 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 decreases in the cost of imported cars and trucks, led to gradual expansion of the paved and improved road networks post-World War II (Gould, 1960; Wasike, 2001; Gwilliam, 2011). However, by 1960, most of today's road network was still unpaved, as can be seen when comparing Figures 3a and 3b that show the road network in 1960 and 2010.

Figure 4a shows aggregate lengths of highways and paved and improved roads over time, and Figure 4b shows their cumulative shares, assuming a constant stock of total roads as measured circa 2004. Initially, in 1960, less than 5% of today's network was paved. Following the independence of most African countries in the early 1960s and into the 1970s, the paved network expanded much more rapidly, fueled by massive public investments in many countries to promote trade and industrialization (e.g.

[17] Available at http://www.fao.org/nr/gaez/en/

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 following the adoption of structural adjustment programs by most African countries (Konadu-Agyemang and Panford, 2006; Gwilliam, 2011). Although investment may have increased again since the mid-2000s decade, 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, p.281-300), roads have deteriorated badly in most African countries in the 1980s, as road maintenance agencies were systematically underfunded. 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).

At 13% in 2010, the share of paved roads in the overall 39-country network remains small compared to the United States (over 99%), China (54%) or India (49%).[18]

Figures 6a–6b then map the evolution of cities over 10,000 over time between 1960 and 2010. 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 trading centers or regional administrative centers established by colonial administrations (Bairoch, 1988; Coquery-Vidrovitch, 2005). Figure 5a shows that the urban population of the 39 countries, here defined as the total population of all cities over 10,000, has increased from less than 25 millions in 1960 to almost 250 millions in 2010. The urbanization rate then increased from only 9% in 1960 to 28% in 2010, as can be seen in figure 5b.

City population is of interest as a measure of local economic development, and in its own right (see for example De Long and Shleifer, 1993 and Acemoglu et al., 2005). 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 2016), 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.[19] As such, we believe city population is the best available measure of local economic development for Sub-Saharan Africa from 1960 to date.[20]

[18]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%.

[19]Henderson et al. (2016) use information on total populations for the subnational units of 89 censuses in 29 countries. These data are not consistently available back to the 1960s for all countries.

[20]While we could use household asset ownership and child mortality from the Demographic and Health

3. Empirical Analysis

As explained above, our unit of analysis is a 0.1 by 0.1 degree grid square. Using these units simplifies computation compared to the full vector road network, and avoids problems due to missing topological information, concerning which segments connect to each other and which do not, in most vector roads datasets. We assign to each grid 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. 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.[21]

The fastest time required to travel from each cell to all cells containing cities across the 199,813 other cells 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. When a map is not available for a given year, we interpolate speeds between the closest map years before and after.[22]

Following Donaldson and Hornbeck (2016), we define origin cell o's market access (MA) in year t, as $MA_{ot} = \sum_{d \neq o} P_{dt}\tau_{odt}^{-\sigma}$, where P is urban population, d indexes destination cells, τ_{odt} is the time required to travel between cells o and d, and σ is the trade elasticity measuring how trade falls as trade costs increase. 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. Our primary measure of change in MA is thus

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

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:

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

Surveys as measure of economic development as in Young (2013), these data do not exist for periods before the 1980s, have limited geographic information before the 1990s, and exclude many small cities.

[21] As shown below, results are robust to using alternative speed assumptions from Alder (2015) and Shiferaw et al. (2013), which are broadly similar but emphasize highways (see Table 1).

[22] 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. Results will hold when dropping that decade.

Baseline controls X_{ot} are country-by-year fixed effects to account for arbitrarily shaped time trends that may differ by country, lagged log population to account for any divergence (convergence) if large cities grow faster (slower) than small cities, depending on local increasing returns, and third 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. With no lags, this specification is analogous to the baseline of Donaldson and Hornbeck (2016).

Our primary sample consists of 2,789 x 6 = 16,734 cell-years with more than 10,000 at one point since 1960. Since the model is estimated in first-difference form, we do not include cell fixed effects. We also lose one round of data, and the cell-decades for which the urban population is nil (or below 10,000) in the initial year are mechanically dropped from our analysis (as the log of 0 is missing). The sample then consists 5,906 cell-decades. Table 2 contains descriptive statistics for our main variables of interest in the sample with two lags (4,725 cell-periods). 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, functional form, and measurement error. In Sections 4. and 4.3. below, we discuss these concerns in detail, and we report the results of several alternative identification strategies.

4. Results: Average Effects

4.1. OLS 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 a doubling in market access. In Column 1, only contemporaneous changes in market access are included. These have a modest impact on city population, with an elasticity of 1.34%. Columns 2–4 add lagged changes in market access from previous decades. Changes in market access in the decade prior to the population change in question and in the decade prior to that each appear to have broadly similar but somewhat smaller effects. The overall effect of a 100% increase in market access in each decade, across these three decades, is thus over 3%. In column 4, the prior decade, 30 years before the measured population change, has a smaller effect that is imprecisely measured. [23]

In column 5, we investigate reverse causality by adding a lead to the column 3 specification. 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. The lead is insiginificant and has little impact on the other coefficients. 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 the lead, with an elasticity of 3 to 4%.

4.2. Main Identification Strategies

Our chief identification concerns are omitted variables and reverse causality. Our measure of market access changes due to both changes in the population of city trading partners and changes in the road infrastructure connecting them. Unmeasured factors increasing a city's population could also be increasing 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 create problems.

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, column 3, and column 2 adds controls for the contemporaneous and lagged changes in time cost (the inverse of the road 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 else will enter that cell. Once we control for this, identification is then coming from changing in roads in other cells. Estimates are virtually unchanged. However, this strategy does not account for endogenous changes in neighboring cells.

Instead of controlling for local roads, Columns 3–6 instrument for changes in overall market access due only to roads built far away, excluding changes due to the building of roads nearby as well as nearby city growth. Formally, the instrument measures changes in market access due only to roads built outside a radius $j \in 5, 10, 15, 20$ cells

[23]The specification with two lags also has lower Aikike and Bayesian information criteria (AIC and BIC) than the one with 3 lags. We thus restrict the remaining specifications to two lags.

(approximately 56, 111, 167, and 222 km) from city o:

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

Effects are larger than the OLS, with 30-year elasticities between 8.8% and 17.7%, spread roughly evenly across the three decades and increasing with the excluded radius. We discount the overall effect using a 20 cell exclusion zone because the instrument is somewhat weaker, with a first stage F-statistic of 6.9. As expected, the instrument is stronger at lower radii, because it includes road changes more directly connected to the city. However, roads changes that are too close to the city may also be exogenous, so the exclusion restriction may be violated. Thus there is a tradeoff between the strength and exogeneity of the instruments. In the rest of the analysis, we will not value one instrument over another and always report the effects using the 5, 10 and 15 cell radii.

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. For example, if Nigeria upgrades its main road between Abuja and Lagos, this will directly affect Cotonou in Benin since Cotonou will now be better connected to cities in the North of Nigeria. Obviously, this strategy will be effective in removing endogeneity to the extent that countries do not disproportionately take into account specific foreign cities in their road-building decisions (remember that country-year fixed effects are included).

The point estimate using foreign changes to instrument for overall changes is slightly larger than the OLS estimate, but smaller than the estimates using the radius-based instruments, and also less significant (p-value of 0.053). Alternative specifications based on this instrument in Tables **??**–**??** are also more variable than those based on the other instruments. We believe this is because the role of borders is highly heterogeneous, which may increase the size of standard errors relative to the size of the coefficient. In column 8, an analogous instrument that excludes a country's direct neighbors along with itself is weak (IV F-stat of 4.03). It is partly mechanical. Given a trade elasticity of 3.8, road changes farther away from a city predict relatively less well its overall changes in market access. The overall effect remains positive and significant (at 10%), and is somewhat in line with the overall effect when using a 5 cell exclusion zone.

The fact that the IV estimates are larger than the OLS is consistent with the recent literature (Redding and Turner, 2015). 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. Alternatively, this downward bias may be the result of measurement error in the market access measure. Finally, the higher IV could simply reflect heterogeneity in the overall effect. The cities most likely to be impacted by road changes far away (i.e. the instrument) are those for which nearby destinations matter little. This is likely to be true of relatively remote cities. The different IVs may thus reflect various local average treatment effects (LATE), hence the need to always compare them.

The magnitude of the effects we find is somewhat smaller than the 0.25 to 0.3 reported for total population in US 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 substantially higher costs of trade and especially migration in this context, especially between countries and perhaps across ethnic territories, in part because of limited land markets. In that sense our context may be closer to modern China with its restrictive Hukou system (Bosker et al., 2015). Second, there was much lower economic growth overall in our context. Donaldson and Hornbeck (2016) study the U.S. in 1870-1890, a period during which it was experiencing its Second Industrial Revolution and receiving massive inflows of immigrants.

Donaldson and Hornbeck (2016) 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 5–6 explore the robustness of results to changing specifications and samples. Every row reports the joint 30-year effect of changed market access (contemporaneous and two lags) for five different identification strategies: OLS, IV excluding 5, 10, and 15 cell radii, and IV excluding domestic changes. Each controls for latitude and longitude polynomials interacted with year fixed effects, country-year fixed effects and lagged log population, as in our baseline specification, unless otherwise indicated.

Specification: Controls. Table 5 adds and removes several controls. Row 1 shows our baseline results. Row 2 adds the following geographic and non-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, several indicators if the cell contains the national capital, the largest city, the second largest city or a regional capital, in 1960 and in 2010, distances to the capital city, the largest city, and the second largest city, in 1960 and in 2010, and an indicator if the city has been for more than 5 years within 150 Km from the place of origin of the president of the same country as the city between t-30 and t. 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.

Row 3 adds 287 decade-specific fixed effects for every 1960 province. The coefficients remain high but they are less precisely estimated given the high number of fixed effects. Then, in the baseline regressions (row 1) we find that the effect of lagged log population is not significantly different from 0, which implies that larger cities have, on average, not been growing faster than smaller cities in our sample, consistent with net agglomeration effects being nil and Gibrat's law being verified. In row 4 we interact lagged log population with decade fixed effects to account for the fact that the importance of these local increasing returns may have been changing over time. Results are unchanged. In row 5, we control for log market access in 1960, to account for mean reversion as governments may have been specifically targeting places that were relatively less connected initially. Our instruments are then weaker for some specifications (especially for "Excl. 15"), but the point estimates remain large in general.[24]

Specification: Functional Form. Row 6 is an auto-distributed lag model, which adds two lags of the dependent variable alongside the two lags of the independent variable.[25] This has little effect except on the foreign instrument specification, which decreases in magnitude and is no longer significant. Row 6 removes the country fixed effects, so that the only remaining fixed effects are for each decade. This does increase the OLS and foreign IV coefficients. In row 8 we cluster standard errors at the country level to account

[24]Point estimates are similar but unsurprisingly not significant if we add 1,147 decade-specific fixed effects for every 2010 province (not shown, but available upon request). Results then hold if we: (i) interact lagged log population with country-decade fixed effects, as the importance of local increasing returns may have been changing differentially across countries over time, (ii) control for railroad market access (using the speed of "Paved Roads" for railroads and the speed of "No Roads" where there are no railroads, see Table 1) in 1960, and (iii) control for road + railroad market access in 1960 (not shown).

[25]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).

for spatial autocorrelation within countries over time. Results remain the same.[26]

Specification: Market Access Measures. In row 9, all populations used in calculating $\Delta lnMA_{ot}$ are from year $t - 10$; P_{dt} in Equation 1 is replaced with $P_{d,t-10}$. In row 10, this strategy is carried to its extreme: all populations used in calculating $\Delta lnMA_{ot}$ are from 1960. Both of these rows purge population change from the change in market access variable, restricting its variation to be induced by changes in roads alone. Most columns change little from baseline. The main exception is the last column of row 10, where the effect doubles using the foreign IV. Rows 11 and 12 replace the speeds assumed at baseline with those used by Alder (2015) and Shiferaw et al. (2013) (see Table 1). Magnitudes increase modestly. Lastly, our baseline analysis assumes $\sigma = 3.8$ in defining market access, following Donaldson (forthcoming). Assigning larger (smaller) values of σ mechanically increases (decreases) the coefficients on market access, as it shrinks (widens) the variation in market access, without altering the variation in city growth on the left-hand side. This can be seen in rows 13 and 14 where we use $\sigma = 2$ and $\sigma = 8$ respectively. However, as shown in Figure A.1, the effect of a one standard deviation increase in $\Delta^{t}_{t-10} \ln MA$, $\Delta^{t-10}_{t-20} \ln MA$, and $\Delta^{t-20}_{t-30} \ln MA$ is quite stable for values of $\sigma \geq 2$.[27] Our results thus do not depend on the choice of a specific trade elasticity.[28]

We also find rather similar effects (not shown, but available upon request) if we: (i) close borders to travel (or assign border crossing costs of 1, 4 or 24 hours), (ii) assume that cells that share only a vertex are not contiguous (rook contiguity), so that travel between them must go through a third cell. Indeed, in the baseline analysis, cells are considered contiguous if they share an edge or a vertex (queen contiguity), and (iii) consider railroad and road market access with railroads having the same speed as roads.

Sampling: Outliers, Countries and Decades. Table 6 varies the sample. Row 1 shows the baseline results. Rows 2–3 drop the cell-years with the largest and smallest one percent of increases in market access and population, respectively, to ensure that outliers are not driving results. Dropping the population outliers decreases OLS and radius-IV effects by 20 to 50 percent, while the foreign-IV estimate increase slightly.

Because not all South African cities are included in calculating market access, it is possible that access to South African cities is biasing results. This is most likely to be

[26]Results also hold if we add a fourth-order term to the spatial polynomial or cluster standard errors at the country-year level (not shown, but available upon request).

[27]Donaldson and Hornbeck (2016) find a similar pattern when varying σ in their context.

[28]We also verify that a high trade elasticity weakens our instruments (not shown, but available upon request), since road changes farther away become by construction less relevant to local city growth. Conversely, a low trade elasticity attenuates the negative effect of distance on trade, which leaves us with less variation in market access to exploit within each country-period. Based on these limitations, we believe that choosing an intermediary trade elasticity of 3.8 for the main analysis is the best choice.

true in the four sample countries that border South Africa: Botswana, Mozambique, Namibia, and Zimbabwe. In row 4, removing these countries has little effect.

Roads data for the 1960s are potentially less reliable because no maps are available before 1965 for the Northwest and Northeast regions. In row 5, dropping the 1980s, the one period that uses road changes from the 1960s, changes the 30-year effect little. Conversely, there are fewer changes in roads documented in the 2000s, and this may be due to poorer documentation. Dropping the 2000s in row 6 slightly reduces the effects.[29]

Sampling: Largest Cities. Rows 7-9 drop cells containing various groups of the largest cities, from just the capital and the two largest cities in 1960 and in 2010 (row 7), to the 5 largest cities in each country in 1960 (row 8) or the 10 largest cities in each country in 1970 (row 9).[30] Remarkably, even excluding the ten largest cities in each country has little effect beyond strengthening the foreign instrument. These results are akin to the identification strategies of Michaels (2008) and Faber (2014), 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. Moreover, these results hold if we additionally control for each city's distances to the largest city, second largest city and capital city in 1960 and 2010 (not shown, but available upon request).

Sampling: Population Data Quality. Our main sample has the advantage of applying a consistent population threshold across all countries and using the same years across all countries. This strategy has two important flaws. First, the sample is not balanced. Places that entered the sample earlier may have been different from other cities in ways that are correlated with road building, raising selection concerns. Second, it requires interpolation and extrapolation, sometimes several years away from censuses or other official estimates. This affects both the dependent variable and the independent variables of interest. As long as this interpolation and extrapolation does not systematically overestimate or underestimate either of these, it should bias

[29] Results are generally similar if we (not shown, but available upon request): (i) weight the sample by population in $t-10$. In the absence of weights, to the extent that any random noise is additive, it may affect small cities more. (ii) drop countries nearest to North Africa and the Arabian Peninsula, respectively, (iii) drop country-periods during which the country was still a colony or was affected by a war, respectively (using data from the Integrated Network for Societal Conflict Research of the Center for Systemic Peace), and (v) drop the 1990s and 2000s, as the 1990s may have also been affected by poorer documentation.

[30] Some of the 5 largest localities in each country in 1960 and 10 largest localities in each country in 1970 are not in our sample of cities because they never reached the 10,000 threshold since 1960. Therefore, to obtain a list of the top localities in each country in 1960 and 1970, we had to use various original sources including population censuses, demographic studies and administrative counts circa 1960 and 1970. Due to lack of data on smaller towns then, we were not able to rank localities that were not among the top 5 cities in each country in 1960. But we were able to do so for the 10 largest localities in 1970.

estimates downward. Note that we always add country-decade fixed effects, which should control for measurement issues that are also at the country-decade level.

Rows 10-11 use additional population estimates for cell-years under 10,000 to increase the sample's balance. Row 10 uses all cell-years with non-zero population estimates. The sample size increases to 7,369. Since 2,789 cells had a city at one point since 1960, the full sample should consist of 8,367 city-periods, which means that we are only missing only about 10% of the full sample. Row 11 also uses all cell-years with non-zero population estimates but only for the one year 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. In both rows 10 and 11, the OLS effects are stronger, while the effects for the radius-exclusion instruments are generally slightly smaller.[31]

Rows 12-14 restrict the sample to country-decades with population estimates that are the most likely to be reliable. Row 12 restricts the sample to periods whose beginning and end populations are each based on at least two census populations, as opposed to other sources. Row 13 excludes country-decades for which the initial and final populations are both at least 5 years from a population data source. In each case, effects on the results are modest. Row 14 applies substantially stronger data quality criteria, excluding country-decades for which the initial OR final populations are more than 5 years, respectively, from a population data source. The sample is reduced by more than 50 percent. The point estimates are reduced as well, and while the two weakest instruments are further weakened, the 5-cell and 10-cell instruments remain strong and they suggest effects that are significant, if reduced by up to a third from baseline.

Instrument Exogeneity If the faraway roads identifying changes in market access in our IV strategy are in fact proxying for nearby roads, as a continuation of radial roads being built, then our identification assumption is violated. Table 7 considers this possibility by excluding cities in which a road is built in a given octant with respect to a city (i.e. North-northeast or West-southwest) both in an inner ring, where it is excluded from the instrument, and an outer ring, where it is included, in the same decade. We refer to such an event as *spatial co-investment*. Row 1 show the baseline results. In rows

[31]To fill in population further in rows 10-11, we index time for each cell with respect to the first year it crossed 10,000 urban residents, and calculate sample statistics among non-missing cities for each year relative to crossing. We then assign the year-specific sample median to those cities with a missing population estimate, but not those cities with a population estimate under 10,000. 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. We can then use the full sample of 8,367 cell-years, but results are unchanged (not shown, but available upon request). They are also unchanged if we assign the relative-year-specific median to all city-year populations with no known population over 10,000 (not shown).

2-4, we then drop cities in which roads were built at radius 2–3 cells, near the city, and at radii 5–6 (row 2), 10–11 (row 3), and 15–16 (row 4) cells, just outside the exclusion zones corresponding to the 5-, 10-, and 15-cell instruments. Sample size is reduced by more than half from the baseline: even when co-investment occurs in just one decade, it can effect subsequent decades via lags. The instruments are weaker than at baseline, and overall there is less variation in market access, because cities with nearby road building are among the ones that are most likely to see large changes in market access, and by construction they are the ones with more variation in the instruments. Still, results are remarkably consistent with those in the overall sample. In rows 5-7, we show that these results hold if we use quartants (i.e. Northeast, Southeast, Southwest or Northwest) instead of octants, although we lose even more cell-years then.[32]

4.4. Net Creation vs. Reorganization of Economic Activity

Results thus far have not distinguished between different sources of growth from the perspective of an individual city. Increased market access could induce a city to grow by attracting rural residents, by attracting urban residents from other cities, or by increasing its rate of natural increase.

Table **??** provides some evidence distinguishing between the first two possibilities. Each panel repeats the baseline regression on successively larger units of analysis, created by aggregating individual grid cells into mutually exclusive square blocks. In the first panel, each unit is a 3x3 square of the original units. Because some such 3x3 squares contain multiple cities, the sample size shrinks. By the last panel, the average 9x9 square contains approximately two cities. If all urban growth induced by roads was pure reallocation within such 9x9 grid squares, we would expect no effect on this sample. Effects do on average become smaller and noisier, with weaker instruments, as is expected given the smaller sample size. However, they are broadly of the same magnitude as baseline results, suggesting that the majority of the effect is not due to local reallocation. We cannot distinguish reallocation between cities across larger distances, as aggregation to larger squares produces smaller sample sizes and weaker identification. However, it is worth noting that less than 25% of sub-Saharan African population was urban in 1980 at the beginning of the regression sample, and this

[32]We obtain generally similar results if we drop the observations for which there are: (i) spatial co-losses, i.e. negative changes in both the inner ring and the outer ring, (ii) spatial co-changes, i.e. any positive or negative change in both the inner ring and the outer ring, (iii) spatial co-investments in opposite octants, i.e. positive changes in both the inner ring of one octant and the outer ring of the opposite octant (e.g., North-northeast and South-southwest), in order to capture co-investments in opposite directions, and (iv) co-investments when using different radii for the inner ring (e.g., 4-5).

number had increased only to 34% by 2010, so the pool of rural potential migrants was always 2-3 times as large as the pool of urban potential migrants.

While no direct evidence can help us to distinguish between rural-urban migration and natural increase, we expect that the effect is driven primarily by induced migration. Increased market access in a city affects incomes by increasing labor demand and therefore wages. Increased wages, in turn, decrease both fertility and mortality. However, variation in urban rates of natural increase across African countries in the period under study was driven primarily by variation in birth rates, whereas urban mortality was much more uniform (Jedwab et al., forthcomingb). Thus if anything, increased market access is more likely to decrease the rate of natural increase, working against urban population growth.

5. Results: Heterogeneous Effects

Tables 8–9 explore heterogeneity of results with respect to several factors highlighted in recent literature on economic geography, structural change, and political economy. The goal of this section is to show that transport investments may have very different effects depending on the local context in which they take place, here focusing on Africa. As in Tables 5–6, each row shows 30-year estimates of a variant of equation (2), in which we interact the contemporaneous and lagged changes in market access with the dummy variable shown at left. Likewise, the three instruments are also interacted with the dummy variable. For the OLS and IV5 estimation strategies the table reports the 30-year coefficient for the dummy=0 group, the dummy=1 group, and the difference; for IV10 and IV15, for which instruments are generally weaker, only the difference is reported. The individual effect of the dummy variable is not shown. The first stage F-statistics for the three instruments are reported on the left.

These exercises are very demanding on the data, with six endogenous variables and six instruments per regression. Not surprisingly, the instruments are generally weaker, especially with only 4,725 observations and country-year fixed effects. Many cross effects of the six instruments in the six first stage regressions are small or even nil. This results in "multiple weak instruments", and the IV F-statistic becomes low and standard errors disproportionately increase as a result (Angrist and Pischke, 2009, p.205). In addition, the extent to which the coefficient for the dummy=0 group and the coefficient for the dummy=1 group are significantly different from each other depends on precisely estimated each coefficient is, and thus the number of observations in each group. The dummy variables do not always evenly split our sample of 4,725 cell-years (the shares of the dummy=1 groups are reported on the left), therefore there may be cases where the coefficient of the difference is high, although mechanically imprecisely estimated.

All in all, results shown are illustrative of broadly consistent general patterns but none are robustly significantly different from zero across a wide variety of specifications.

Economic Geography. Table 8 shows variation with respect to various economic geography characteristics. In rows 1 and 2 we show the interaction effects when the dummy variable controls for whether the city was initially (in $t - 30$) larger than the median city size at the continental level or the country level, respectively. Larger cities generally see smaller effects. The core-periphery model predicts that reduced trade costs should actually disproportionately increase the size of the bigger cities. However, we find reduced trade costs lead to a decentralization of economic activity in our context. Note that the effects are also smaller when the size is measured with respect to the continental standard (row 2) rather than the country-specific standard (row 1). This suggests that the largest cities of the continent are the ones that benefit the least from reduced trade costs to other African cities. However, the growth of the largest cities in Africa may be disconnected from their surrounding environment, and depend more on national and international factors, which should lower the value of local connections.[33]

Rows 3-7 consider several dummy variables proxying for economic remoteness as of 1960: above median market access in the country (row 3), presence of a railroad in the cell of the city (row 4), presence of a paved or improved road in the cell (row 5), being nearer to the "top" cities – the largest, second largest and capital cities – in each country (row 6), and being far from the border (row 7).[34] In each case except the last, cities with attributes related to better market access see smaller effects of a marginal improvement. In other words, cities that were better connected initially do not react as strongly to becoming even more connected. This potentially suggests that there may be decreasing marginal returns to transportation investments.[35]

[33] Differentials based on population in 1960, as opposed to $t - 30$, are generally in the same direction but smaller (not shown, but available upon request), perhaps because most cities have an assigned population of zero in 1960, so the group of small (zero-population) cities is always much larger than the group of big (populated) cities. Effects are also similar if we use arbitrary population thresholds (e.g., 20,000) or district urbanization rates in 1960 using as the numerator the population of cities above 10,000 and as the denominator the imputed population of each district in 1960 (district boundaries circa 2000).

[34] In row 3 the dummy variable is defined with respect to market access in 1960, and not in $t - 30$ as for city size (rows 1-2), because having to also control for market access in $t - 30$ would mechanically prevent us from correctly estimating the second lag in market access from $t - 30$ to $t - 20$.

[35] We find rather similar effects if we compare the effects of positive vs. negative, or bigger vs. smaller, changes in market access (not shown, but available upon request). However, the coefficients for the dummy=0 group, and thus the relative effects, are imprecisely estimated due to the fact that few city-years have experienced a negative or small change in their market access. Nonetheless, these results also suggests that only significant changes in market access may have strong economic effects.

The previous results suggest that remoteness raises the returns to transportation investments, when it is defined in terms of initial access to transportation infrastructure and other cities. In row 7, cities farther from a border show little difference from elsewhere. The effect when using the 15 cell exclusion zone is significant but the IV F-stat is low, at 5.3. If anything, the point estimates are positive, which may imply that being "remote" does not increase the marginal returns to transportation investments when remoteness is defined geographically. Many border regions are relatively poor in Africa, so being far from markets may not be a necessary condition for roads to have a strong impact. Another interpretation could be that border regions are less likely to be remote in Africa than in other regions, which would then lower the returns to new transportation investments. Indeed, ten capital cities in the region are within 20 km of an international border, and more than 20 are within 100 km. We will confront these two hypotheses in a future version of the paper.

Physical Geography. Rows 1-4 of Table 9 show variation with respect to various physical geography characteristics. The first three rows cut the sample at the 75th, 25th and 50th percentiles of the distribution of land suitability for all crops, as measured within one grid cell of the city.[36] At all three cutoffs, cities with worse land are more positively affected by increases in market access. In row 1 we find that cities in areas where land suitability is over 75% grow relatively slower when market access increases. The fact some of the differences (for Excl. 5 and Excl. 15) are significant are all the more impressive given that the dummy=0 group represents only 5% of the sample and the coefficients for the dummy=0 group are imprecisely estimated as a result. One interpretation is that these areas have a comparative advantage, and thus specialize, in agriculture, using land and other available resources for the agricultural sector rather than for the urban sector. Conversely, in row 2 we show that cities in areas where land suitability is under 25% grow relatively faster when they are better connected to other cities. This is consistent with a model in which cities endowed with poor agricultural productivity specialize in activities that are more transport-sensitive. All in all, with reduced trade costs, some areas specialize in growing food and other crops for cities in other areas, and whose growth is not less constrained. Interestingly, the specialization effects are weaker when we use a land suitability cut-off of 50%, as the comparative advantage of cities across both sides of the cut-off is less clear. These results are consistent with an Eaton and Kortum (2002) model of trade and economic growth.[37]

[36]Land suitability is defined as the suitability of the most suitable crop in GAEZ. Calculating land suitability within a broader area (up to 3 or 5 cells away) results in differentials that are generally similar, but the 75th percentile instruments are substantially weaker (not shown, but available upon request).

[37]We find generally similar effects as for land suitability when we study the interaction effects with

In row 4, we find that cities farther from the coast than the median city in a country look no different than other cities. The point estimates are positive, but not significant.[38]

Political Geography. Rows 5-8 of Table 9 show variation with respect to various political geography characteristics. Knight (2004) finds that in the U.S. legislators are ceteris paribus more likely to support transportation projects in their own district, no matter how inefficient these spatially misallocated investments are at the country level. Likewise, closer to our context, Burgess et al. (2015) show that Kenyan heads of state build more roads in areas with high representation of co-ethnics. If these "roads to nowhere" are built in specific areas because of their high marginal political returns rather than for their economic returns, there are reasons to think that these additional roads may be less effective in promoting growth Tanzi and Davoodi (1998).

Row 5 compares cities within 150 km of the place of origin of a head of state who was in power for at least five years between $t-30$ and t with all other cities. We use five years because we believe a leader must have been in office for more than just a few years before being able to durably influence the spatial allocation of roads in her country.[39] The effects are clearly negative, thus 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. This is all the more surprising as one would think that these areas would also receive complementary public investments and subsidies, which should actually increase the returns to transportation investments. We indeed find that the individual effect of the ethnic favoritism dummy is indeed positive and significant, at 0.036* (not shown), thus implying that cities around the place of origin of the president grows 3.6 percentage points faster than other cities in the country on average. If anything, our effects should be upward-biased, and thus give conservative estimates of the potentially negative effects of ethnic favoritism on the efficiency of road investments.[40]

rainfall or the presence of arid and/or semi-arid areas (not shown, but available upon request). "Desert cities" disproportionately benefit from reduced trade costs to other cities.

[38]We find no significant difference when the dummy variable proxying for coastal access is defined at the country level, thus comparing coastal and landlocked countries (not shown, but available upon request). We also find no significant difference when using the Euclidean distance to an international port (using the list of ports in 2010) rather than the Euclidean distance to the coast. Indeed, what may matter for city growth is instead the network distance to a port, as we will investigate below.

[39]Burgess et al. (2015) find that the Ministry of Public Works is always the financially most well-endowed ministry in Kenya, and that the head of state always choose a Minister from the same ethnic group. Changes in the allocation of roads can be noticed one year after the selection of a new head of state.

[40]We choose 150 km as the cut-off because we need the dummy=1 group to have enough observations for the effects to be enough precisely estimated. If we choose smaller cut-offs such as 50 or 100 km, our dummy=0 and dummy=1 samples are more unevenly split, and standard errors increase. However, the effects generally hold when doing so (not shown, but available upon request). Results are also somewhat unchanged if we: (i) choose larger distances (e.g., 200 and 250 km), (ii) use other arbitrary thresholds for the length in office (e.g., 2 or 10), (iii) use the place of birth instead of the place of origin (treating

In row 6, rather than defining whether a city was potentially favored by a head of state between $t-30$ and t, we now define whether it was potentially favored in each of the three periods between $t-30$ and $t-20$, $t-20$ and $t-10$, and $t-10$ and t, and interact each lag-specific favoritism dummy variable with the corresponding change in market access variable. Since each period of interest lasts ten years instead of thirty years, we must choose a lower threshold for the length in office, and in this case we choose 2 years. Row 6 thus compares city-decades within 150 km of the birthplace of a head of state who was in power for at least two years in that decade with all other city-decades. When doing so, the IV F-stats increase, but the samples for the dummy=0 and dummy=1 groups also become more unevenly split (the share of the dummy=1 group is now 0.17, instead of 0.34), so standard errors increase relative to the size of the coefficient. Overall, the effects remain generally negative (the p-value for the coefficient of the difference for Excl.10 is 0.103, so the difference is almost significant at 10%).[41]

Conversely, and unlike large cities in Panel A, regional capitals see if anything larger effects of increased market access on their growth, consistent with, for example, complementarity between government services and transport-sensitive activities. However, this differential is only significant when using capitals in 2010 rather than in 1960 (rows 7-8), when regional capital status could have been affected by prior road building. The samples for the dummy=0 and dummy=1 groups are also more unevenly split when using the regional capitals in 1960, since there were fewer of them then, which may explain why the coefficients of the differences are less precisely estimated. If anything, we do not find the same negative effects as for cities that are favored because of their ethnic connections (rows 5-6). Roads built to connect regional capitals to each other may be necessary for nation-building, especially in Africa Herbst (2000). This is additional evidence roads built for political reasons may have different effects depending on the underlying true motivation for these transportation investments.

as missing the heads of state born abroad), and (iv) use the capital city as the place of origin when the country was still a colony, assuming that Europeans favored that city over other cities.

[41]These results are robust to the tests explained in the previous footnote. We then find similar effects when the favoritism dummy is constructed using the ethnicity of the head of state rather than her place of origin. As described in subsection 2.3., we know the main ethnic group in each cell, for both the Murdock and the GREG classifications. We can thus create a dummy if the dominant ethnic group in a city also happens to be the ethnic group of a head of state whose length in office has been at least 5 or 2 years depending on the period considered. The point estimates are in line with the effects in rows 5-6 but they are more imprecisely estimated. Indeed, there is no clear official definition of ethno-linguistic groups in Africa, and we had to make some arbitrary choices to match our raw data on the ethnicity of heads of state with the ethnic classifications available in the Murdock and GREG maps. In addition, some other ethnic groups may coincidentally receive transportation investments simply because they are on the path between say the capital city and the place of origin of the head of state. We can only capture this misallocation when our favoritism dummy is based on the catchment area from the place of origin.

Domestic vs. Foreign Market Access. Table 10 investigates the respective effects of domestic overland, foreign overland and foreign overseas market access. Row 1 shows our baseline results from Table **??**. In Panel B, we compare the effects of domestic overland market access and foreign overland market access, here defined as market access (via roads) to domestic cities and foreign cities only, respectively. In row 2 we find that the effects of domestic market access are high, at 7.5-11 for the IVs, but a bit lower than the effects of overall market access (row 1). In row 3, we show that these results are robust to controlling for foreign market access.[42] In rows 4-5, we show that the effects of foreign market access are smaller, whether we control for domestic market access or not. When the identification only relies on foreign roads, the effect is about 5, which is similar to the estimated LATE of overall market access when we use for the instruments changes in foreign roads only (4.61*, see column (5) in row 1).

In Panel C, we compare the effects of overland market access and overseas market access, here defined as market access (via roads) to domestic and foreign cities in Africa and domestic and foreign African cities with an international port, respectively. Indeed, we may expect stronger effect of transportation investments for cities that are now better connected to international ports, and thus the rest of the world. Here, we identify 44 cities with an international port in 2010 (including 5 ports in South Africa). First, for these cities we verify that there is a strong correlation between their population size in 2010 and a measure of cargo traffic circa 2010 (coefficient of correlation of 0.68).[43] Second, we decide to use their population size in year t as a proxy for port traffic in year t. Third, to construct overseas market access in year t, we only use the population of these 44 cities in year t. Fourth, we can then construct the contemporaneous and lagged changes in overseas market access. In row 6 we show that the effects of overland market access remain similar, or if anything increase, when controlling for overseas market access (the instruments are too weak for Excl. 15 so we may want to discard the associated effect). In row 7 we find strong effects around 9.2-16.5 of overseas market access, and this across all specifications. Row 8 shows that these results remain robust to controlling for changes in overland market access (the p-value of the coefficient for Excl.Dom. is 0.15, so the effect of overseas market access is only significant at 15%). Interestingly, the effects of foreign overseas market access are much larger than the effects of foreign overland market (row 7 vs. row 5). This suggests that being better

[42] As explained in footnote 35, larger changes in market access tend to have disproportionately larger effects than smaller changes in market access. If changes in domestic and foreign market access are partially correlated, maybe because transport investments improve market access to both domestic and foreign cities, the changes in market access are smaller, and the estimated effects are also smaller.

[43] Note that we are currently collecting data on the traffic of the major African ports in 1960.

connected to other cities in other African countries may be less valuable being better connected to the rest of the world. This also potentially suggests that policies aiming at better connecting African countries to each other may have smaller effects than connecting these same countries to other countries outside Africa.[44]

To summarize, we find suggestive evidence that the effect of market access is stronger for small and remote cities, and cities that experience a large change in their market access, especially to other cities of the same country or international ports, and weaker for cities in areas with a comparative advantage in agriculture and areas most likely to be favored by ethnic patronage. This shows how heterogeneous the effects of transportation investments can be depending on the context in which they place.

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 in the 30 years after road construction, and appears to be somewhat larger for smaller and more remote 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.

[44]This may be because other cities in other African countries are poor and/or have the same comparative advantage, so that gains from intracontinental trade are lower than gains from intercontinental trade, or that there are border costs that prevent intra-continental roads from having larger economic effects. While there are also border costs when trading with the rest of the world, these border costs may actually be lower than at border crossings between African countries.

REFERENCES

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

African Development Bank and United Nations Economic Commission For Africa, "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.

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

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

Angrist, Joshua D. and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion.* 2009.

ao, Nuno Lim and Anthony Venables, "Infrastructure, geographical disadvantage, transport costs, and trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

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

Bairoch, Paul, *Cities and Economic Development: Frow the Dawn of History to the Present* 1988.

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

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 05 2007, *122* (2), 775–805.

__ , Loren Brandt, Vernon Henderson, Matthew Turner, and Qinghua Zhang, "Roads, Railroads and Decentralization of Chinese Cities," mimeo, Brown University 2015.

__ , Vernon Henderson, Matthew Turner, Qinghua Zhang, and Loren Brandt, "Highways, Market Access, and Urban Growth in China," mimeo, Brown University 2016.

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

Bogart, Dan, "Turnpike trusts and property income: new evidence on the effects of transport improvements and legislation in eighteenth-century England," *Economic History Review*, 02 2009, *62* (1), 128–152.

Bosker, Maarten, Uwe Deichmann, and Mark Roberts, "Hukou and highways : the impact of China's spatial development policies on urbanization and regional inequality," Policy Research Working Paper Series 7350, World Bank 2015.

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

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 09 2014, *82* (5), 1671–1748.

Burgess, Robin, Remi 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.

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

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

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

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

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

Chaves, Isaias, Stanley L. Engerman, and James A. Robinson, "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, 2014, pp. 321–365.

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

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

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

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

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

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

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

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, "The Shuttle Radar Topography Mission," *Reviews of Geophysics*, June 2007, *45* (2).

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

Fearon, James D., Kimuli Kasara, and David D. Laitin, "Ethnic Minority Rule and Civil War Onset," *American Political Science Review*, 2 2007, *null*, 187–193.

Feyrer, James, "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 December 2009.

Francois, Patrick, Ilia Rainer, and Francesco Trebbi, "How Is Power Shared in Africa?," *Econometrica*, 2015, *83* (2), 465–503.

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.

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

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

Greene, William H., *Econometric Analysis*, 6 ed., Saddle Hill, NJ: Prentice Hall, 2008.

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

Haines, M.R. and R.A. Margo, "Railroads and local economic development: the United States in the

1850s," in J.L. Rosenbloom, ed., *Quantitative Economic History: The Good of Counting*, Routledge, 2008, pp. 78–99.

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

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

Henderson, Vernon, Adam Storeygard, and Uwe Deichmann, "Is climate change driving urbanization in Africa?," mimeo, Tufts University 2016.

Herbst, Jeffrey, *States and Power in Africa: Comparative Lessons in Authority and Control (Princeton Studies in International History and Politics* 2000.

Hodler, Roland and Paul A. Raschky, "Regional Favoritism," *The Quarterly Journal of Economics*, 2014, *129* (2), 995–1033.

IIASA and FAO, "Global Agro-ecological Zones (GAEZ v3.0)," Technical Report 2012.

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

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

__ , Luc Christiaensen, and Marina Gindelsky, "Demography, Urbanization and Development: Rural Push, Urban Pull and... Urban Push?," *Journal of Urban Economics*, forthcoming.

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

Konadu-Agyemang, Kwadwo and Kwamina Panford, *Africa's Development in the Twenty-First Century: Pertinent Socioeconomic and Development Issues* 2006.

Law, Gwillim, *Administrative Divisions of Countries (Statoids)*. 2016.

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

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

Murdock, George, *Africa: Its Peoples and their Culture*. 1959.

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

Ocean Shipping Consultants, Ltd., *Beyond the Bottlenecks: Ports in Africa* 2009.

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 N. Kaplan, ed., *Mergers and Productivity*, University of Chicago Press, January 2000, pp. 51–92.

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

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

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

Shiferaw, Admasu, Mans Soderbom, Eyersusalem Siba, and Getnet Alemu, "Road Infrastructure and Enterprise Dynamics in Ethiopia," Working Paper, The College of William and Mary 2013.

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* 1998.

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

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

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

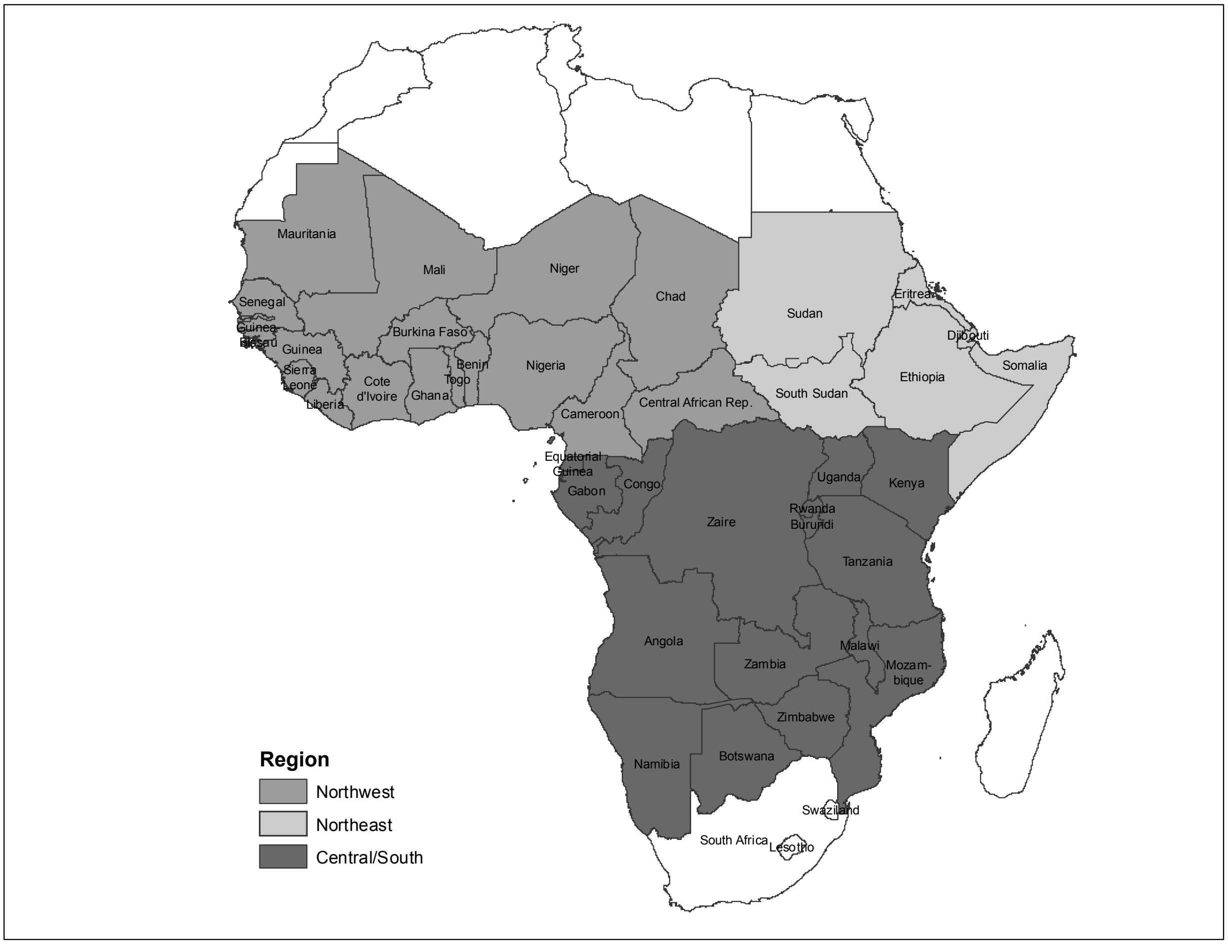
Wikipedia, *Administrative Divisions of Country X; List of Heads of State of Country X; Profile of Head of State Y* 2016.

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

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

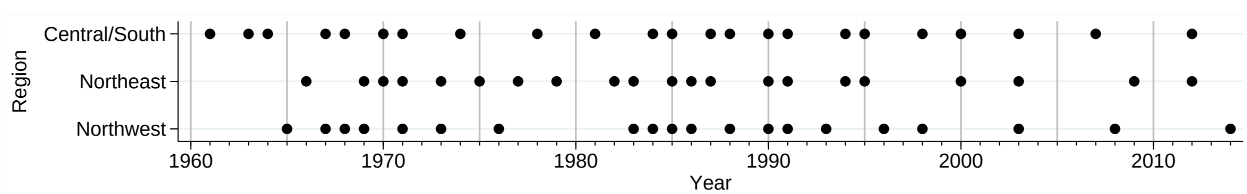
Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

Figure 1: Map Regions and the 39 Countries of the Main Sample



Notes: This figure shows the 39 countries of our main analysis. There are 18 countries in the *Northwest* section, 5 in the *Northeast* section, and 16 in the *Central/South* section of the road maps for sub-Saharan Africa. All analysis treats Sudan as one country, as it was for the full population sample period. South Africa, Lesotho and Swaziland contribute roads (from the *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. See Web Data Appendix for more details on data sources.

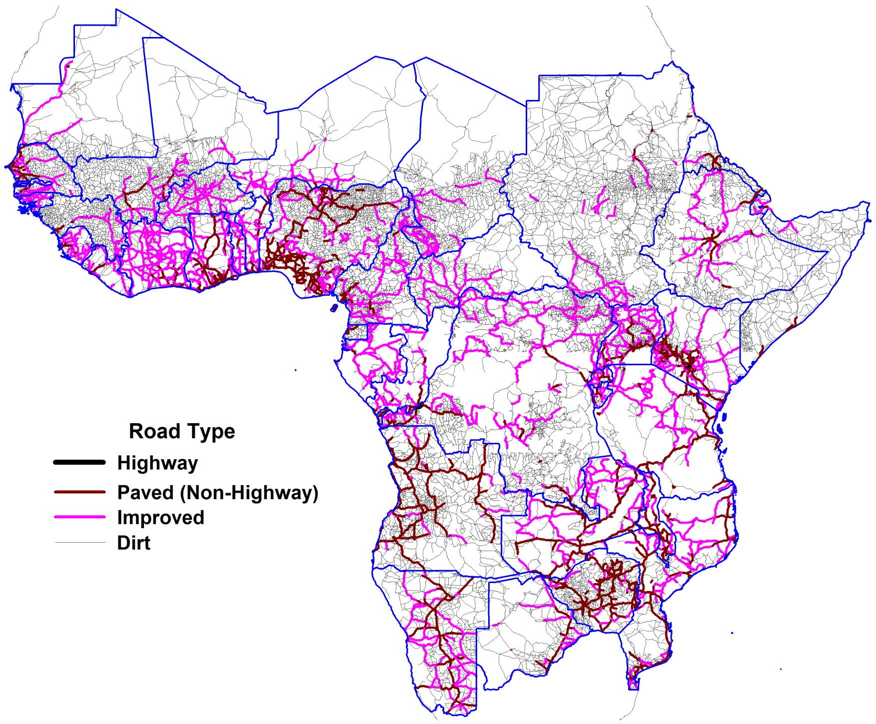
Figure 2: Map Years for Each Map Region



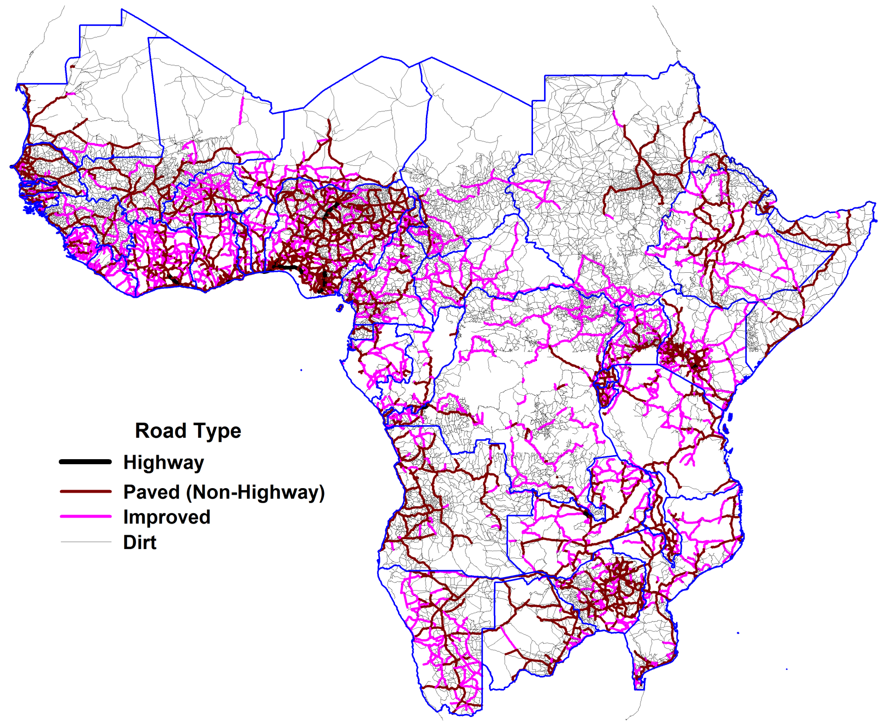
Notes: This figure shows the years for which we have a map for each of the three sections of the road maps for sub-Saharan Africa. There are 20 maps for the *Northwest* section, 21 for the *Northeast* section, and 23 for the *Central/South* section, therefore 64 in total. The average gap between maps across regions is under 2.5 years, and the longest is 7 years. See Web Data Appendix for more details on data sources.

Figure 3: Maps of the Road Network for the 39 sub-Saharan African Countries, 1960-2010

(a) Roads Circa 1960



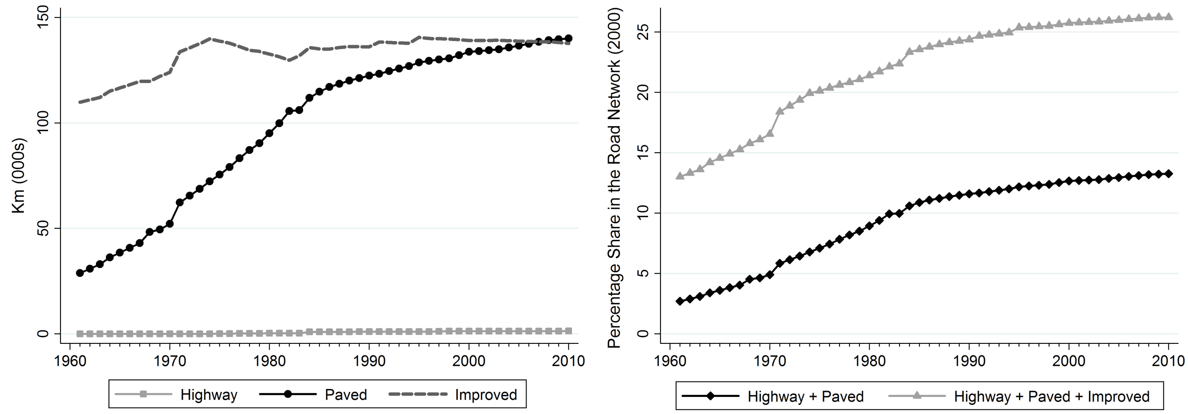
(b) Roads Circa 2010



Notes: This figure shows the evolution of the intercity road network for the 39 sub-Saharan African countries of our main sample between 1960 and 2010. Subfigures 3a and 3b show the roads in 1960 and in 2010 respectively. Roads are classified into four categories: *highways*, *paved roads*, *improved roads*, and *dirt roads*. The dirt roads are defined for the year 2004. Indeed, the road paper maps that we have at our disposal do not cover dirt roads comprehensively. Instead, we use our road paper maps to code each segment from the Nelson and Deichmann (2004) GIS map of existing roads in Africa in 2004 as a highway, a paved road or an improved road in each year that it is a highway, paved or improved. See Web Data Appendix for details on data sources.

Figure 4: Evolution of the Road Network for the 39 Sample Countries, 1960-2010

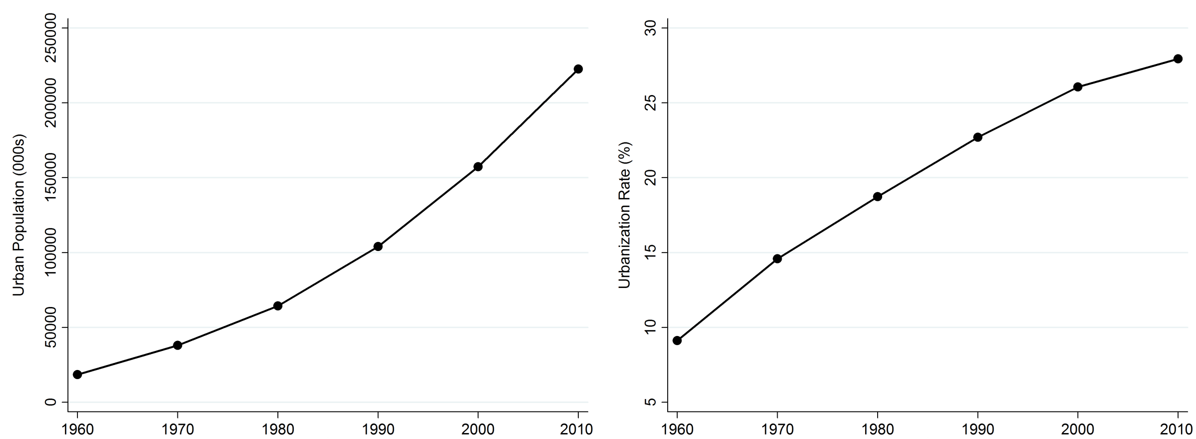
(a) Total Length of Each Type (b) Fraction of Each Type



Notes: This figure shows the aggregate evolution of the intercity road network for the 39 sub-Saharan African countries of our main sample between 1960 and 2010. Subfigure 4a shows the total length (000s Km) of each road type over time: highways, paved roads and improved roads. Subfigure 4b shows the fraction (%) of paved roads broadly defined (incl. highways and paved roads strictly defined) and non-dirt roads (incl. highways, paved roads strictly defined, and improved roads) in the total road network (defined circa 2004). See Web Data Appendix for more details on data sources.

Figure 5: Evolution of Urban Patterns for the 39 Sample Countries, 1960-2010

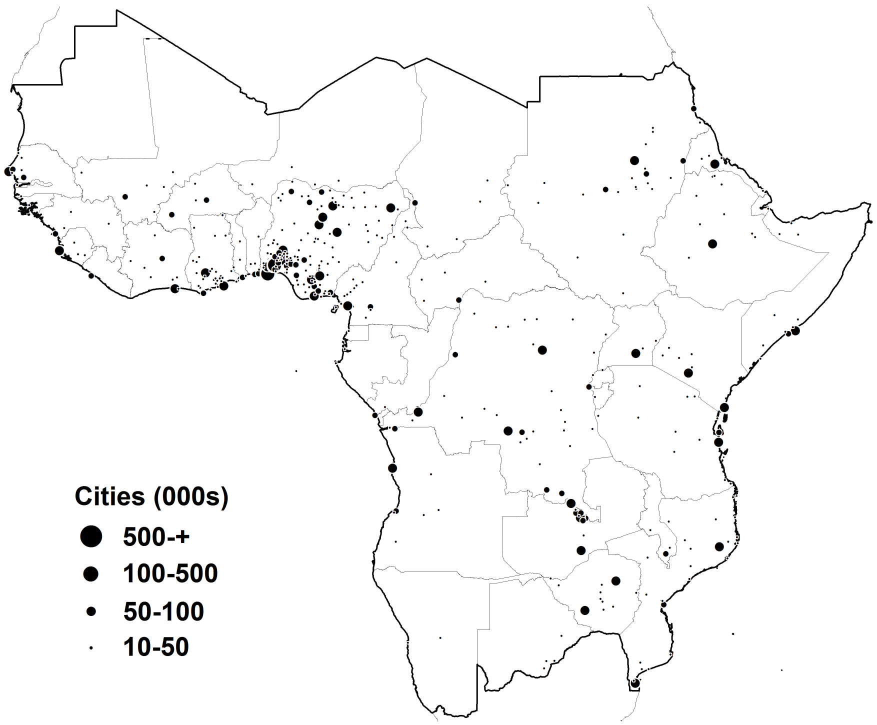
(a) Urban Population (b) Urbanization Rate



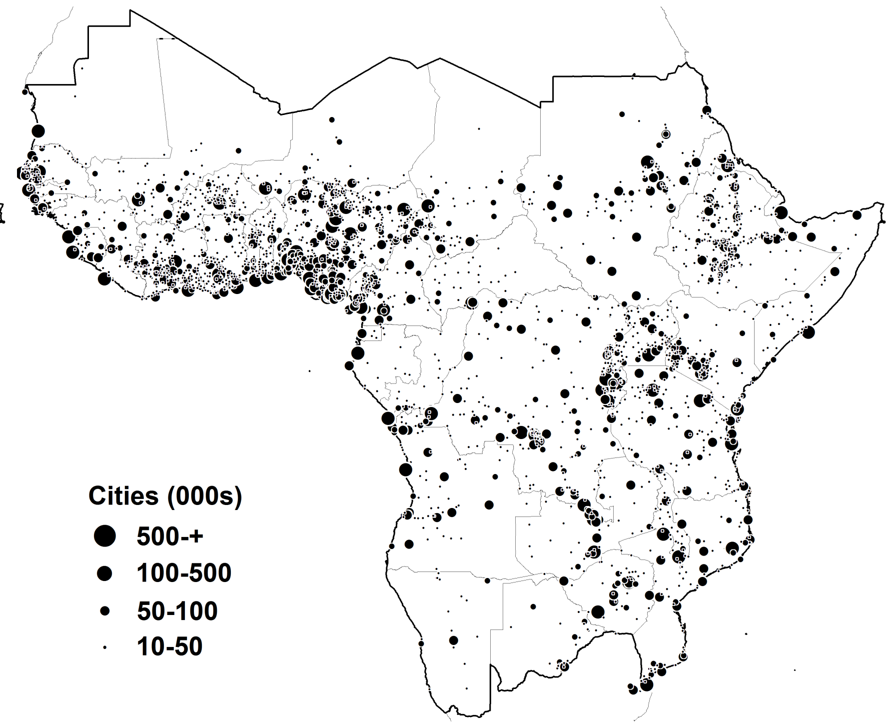
Notes: This figure shows aggregate urban patterns for the 39 sub-Saharan African countries of our main sample between 1960 and 2010. Subfigure 5a shows the urban population (000s), here defined as the total population of localities where population is greater than 10,000 inhabitants in each year. Subfigure 5b shows the urbanization rate (%), here defined as the share of the urban population defined above in the total population in each year. See Web Data Appendix for details on data sources.

Figure 6: Maps of the Urban Network for the 39 sub-Saharan African Countries, 1960-2010

(a) Cities in 1960



(b) Cities in 2010



Notes: This figure shows the evolution of the urban network for the 39 sub-Saharan African countries of our main sample between 1960 and 2010. Subfigures 6a and 6b show the cities in 1960 and in 2010 respectively. Cities are localities where population is greater than 10,000 inhabitants in 1960 (N = 428) and 2010 (N = 2,891). The main sources are *Africapolis I* and *Africapolis II*, and population censuses. See Web Data Appendix for more details on data sources.

Table 1: Speeds (Km / Hour) Assumed in City-to-City Distance Calculations

Category	This Paper	India (Alder, 2015)		Ethiopia (Shiferaw et al., 2013)	
Highway	80	Golden Quadrilateral	75	Post-Rehab Asphalt	70
Paved	60	Conventional Highways	35	Pre-Rehab Asphalt	50
Improved	40	Roads of Lower Quality	25	Pre-Rehab Federal Gravel	25
Earthen	12	Unpaved/No Roads	10	Pre-Rehab Regional Gravel	25
No Road	6	Unpaved/No Roads	10	Pre-Rehab Earth	20

Table 2: Descriptive Statistics for the Main Sample of City-Period Observations

Main Variable:	Obs.	Mean	Std. Dev.	Min	Max
Δ_{t-10}^{t} ln Urban Population	4,725	0.318	0.209	-1.533	2.343
Δ_{t-10}^{t} ln MA (Market Access)	4,725	0.666	0.903	-8.192	10.613
Δ_{t-20}^{t-10} ln MA (Market Access)	4,725	0.909	1.113	-8.192	11.407
Δ_{t-30}^{t-20} ln MA (Market Access)	4,725	1.175	1.300	-8.192	13.288
ln Urban Population$_{t-10}$	4,725	10.247	0.990	9.210	15.902

Table 3: OLS Estimates of the Average Effect of Market Access on Urban Population

Dependent Variable:	(Δ_{t-10}^{t} ln Urban Population)/100				
	(1)	(2)	(3)	(4)	(5)
Δ_{t-10}^{t} ln MA (Market Access)	1.34***	1.27***	1.58***	1.63***	1.50***
	[0.32]	[0.32]	[0.35]	[0.44]	[0.38]
Δ_{t-20}^{t-10} ln MA (Market Access)		1.02***	1.23***	1.55***	1.11***
		[0.24]	[0.26]	[0.34]	[0.30]
Δ_{t-30}^{t-20} ln MA (Market Access)			0.81***	0.89***	0.79***
			[0.23]	[0.29]	[0.27]
Δ_{t-40}^{t-30} ln MA (Market Access)				0.27	
				[0.23]	
Δ_{t}^{t+10} ln MA (Market Access)					0.67
					[0.49]
Overall Effect	1.34***	2.29***	3.62***	4.33***	3.40***
($t-40$ to t)	[0.32]	[0.45]	[0.59]	[0.83]	[0.65]
Observations	5,906	5,472	4,725	3,630	2,607
Adj. R-squared	0.26	0.22	0.19	0.18	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. See the main text for details on how the market access measures are constructed. "Overall Effect" is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln$ urban pop$_{t-10}$, and third order polynomials in longitude and latitude interacted with year fixed effects. Robust 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

Dep. Var.:	$(\Delta_{t-10}^{t}$ ln Urban Population$)/100$							
		Control:	*Columns (3)-(8): Instrumental variable (IV):*					
	OLS	Own Cost	Exclude 5	Exclude 10	Exclude 15	Exclude 20	Foreign	Non-Neighbor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ_{t-10}^{t} ln MA	1.58***	1.52***	2.98***	4.59***	5.75*	7.22*	1.79	8.43*
	[0.35]	[0.39]	[1.00]	[1.76]	[2.95]	[4.09]	[1.89]	[4.60]
Δ_{t-20}^{t-10} ln MA	1.23***	1.24***	3.28***	5.76***	7.34***	9.01***	1.73	-2.49
	[0.26]	[0.29]	[0.87]	[1.59]	[2.46]	[3.13]	[1.43]	[2.66]
Δ_{t-30}^{t-20} ln MA	0.81***	0.83***	2.57***	3.38**	4.60**	4.07**	1.09	2.06
	[0.23]	[0.24]	[0.86]	[1.39]	[1.95]	[1.92]	[1.12]	[1.70]
Overall Effect	3.62***	3.58***	8.83***	13.74***	17.69***	20.30***	4.61*	8.00*
$(t-30$ to $t)$	[0.59]	[0.65]	[1.89]	[3.31]	[4.64]	[5.77]	[2.38]	[4.82]
IV F-Stat			114.00	41.86	17.41	6.94	15.10	4.03

Notes: Each column 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. See the main text for details on how the market access measures are constructed. "Overall Effect" is the sum of the contemporaneous effect and all lags shown. Each regression controls for country-year fixed effects, $\ln$ urban pop$_{t-10}$, and third order polynomials in longitude and latitude interacted with year fixed effects. Column 2 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. In column 8, the "zone of exclusion" includes both the country itself and all its neighbors. Robust standard errors, clustered by cell, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 5: Robustness: Controls, Specification, and Alternative Market Access Measures

		OLS	Excl. 5	Excl. 10	Excl. 15	Excl.Dom.
		(1)	(2)	(3)	(4)	(5)
(1)	Baseline	3.62***	8.83***	13.74***	17.69***	4.61*
	(F: _; 114.0; 41.9; 17.3; 15.1)	[0.59]	[1.89]	[3.31]	[4.64]	[2.38]
(2)	City Controls	3.27***	8.11***	13.14***	18.40***	4.33*
	(F: _; 112.0; 30.6; 14.1; 15.0)	[0.58]	[1.95]	[3.69]	[5.47]	[2.49]
(3)	Province 1960 * Year FEs	1.64**	6.53**	14.58*	15.54	6.60
	(F: _; 36.9; 8.9; 2.9; 16.3)	[0.69]	[3.28]	[7.95]	[12.38]	[4.14]
(4)	ln Urban Pop$_{t-10}$ * Year FEs	3.60***	8.92***	13.85***	17.84***	4.51*
	(F: _; 113.9; 41.9; 16.7; 15.0)	[0.59]	[1.89]	[3.31]	[4.65]	[2.38]
(5)	ln Market Access 1960	3.38***	12.45***	33.02***	88.96*	4.48
	(F: _; 41.5; 6.3; 1.0; 13.2)	[0.61]	[3.20]	[11.03]	[52.24]	[3.22]
(6)	ADL Model (2,2)	4.21***	8.16**	12.34**	18.92**	2.3
	(F: _; 55.9; 28.4; 12.1; 8.5)	[1.04]	[3.29]	[5.27]	[8.13]	[4.12]
(7)	Year FE I/O Country-Year FE	4.84***	8.11***	11.54***	12.99***	6.37***
	(F: _; 129.5; 43.6; 20.9; 14.8)	[0.60]	[1.84]	[3.16]	[4.25]	[2.08]
(8)	SE (Cluster Country)	3.62***	8.83***	13.74***	17.69***	4.61*
	(F: _; 59.6; 40.4; 13.8; 13.3)	[0.83]	[2.53]	[3.50]	[5.32]	[2.48]
(9)	MA: Fixing Pop. at t-10	3.41***	10.34***	15.72***	20.96***	6.79
	(F: _; 151.6; 47.2; 22.8; 65.8)	[0.95]	[2.23]	[3.80]	[5.74]	[4.59]
(10)	MA: Fixing Pop. at 1960	3.18***	9.80***	16.67***	25.40***	11.41**
	(F: _; 233; 70.1; 13.7; 31.6)	[1.03]	[2.05]	[3.55]	[5.62]	[5.71]
(11)	Alder (2015) Speeds	3.88***	11.20***	17.93***	22.71***	6.10**
	(F: _; 81.0; 34.2; 15.3; 15.6)	[0.64]	[2.40]	[4.28]	[5.91]	[2.95]
(12)	Shiferaw et al. (2013) Speeds	4.19***	14.51***	21.28***	26.19***	7.65**
	(F: _; 48.1; 15.2; 7.7; 10.1)	[0.69]	[3.07]	[5.57]	[7.57]	[3.67]
(13)	Sigma = 8	1.21***	3.44***	5.97***	7.17***	1.16
	(F: _; 103.5; 30.8; 7.7; 8.9)	[0.21]	[0.74]	[1.43]	[2.16]	[0.97]
(14)	Sigma = 2	13.05***	20.35***	27.37***	34.95***	23.48***
	(F: _; 158.3; 72.7; 41.7; 44.0)	[2.33]	[5.06]	[7.49]	[9.68]	[8.64]

Notes: Unless otherwise indicated, each cell reports a coefficient and standard error (clustered by cell), and a first stage F-stat (F:) corresponding to the overall effect ($t-30$ to t) from a separate regression. The five columns are variants of columns 1, 3, 4, 5, and 7 of Table 4. The dependent variable is ($\Delta_{t-10}^{t} \ln$ urban population)/100. Each regression controls for country-year FE, $\ln pop_{t-10}$, and a third order polynomial in longitude and latitude interacted with year FE. The sample consists of 4,725 cell-years. Row 2: We add the controls listed in the main text. Row 3: We add 287 province in 1960-year FE. Row 4: We interact $\ln pop_{t-10}$ with year FE. Row 5: We control for log market access in 1960. Row 6: Auto-distributed lag model, controlling for two lags of population. Row 7: Year FE instead of country-year FE. Row 8: SE clustered at the country level. Rows 9-10: $\Delta_{t-10}^{t} \ln MA_i$ calculated using $pop_{j,t-10}$ and $pop_{j,1960}$, respectively. Rows 11-12: Assuming the speeds used by Alder (2015) and Shiferaw et al. (2013), respectively. Rows 13-14: Using $\sigma = 8$ and $\sigma = 2$, respectively. *, **, *** denote significance at the ten, five, and one percent level, respectively.

Table 6: Robustness: Sampling, Largest Cities, and Population Data Quality

	OLS	Excl. 5	Excl.10	Excl.15	Excl.Dom.
	(1)	(2)	(3)	(4)	(5)
(1) Baseline	3.62***	8.83***	13.74***	17.69***	4.61*
(N:4725. F: _; 114.0; 41.9; 17.3; 15.1)	[0.59]	[1.89]	[3.31]	[4.64]	[2.38]
(2) Drop Top/Bottom 1% Δ Pop	2.80***	5.18***	8.51***	10.14**	5.76***
(N:4456. F: _; 111.5; 39.0; 19.0; 15.4)	[0.52]	[1.68]	[3.03]	[4.28]	[2.19]
(3) Drop Top/Bottom 1% Δ MA	4.51***	8.56***	13.44***	17.24***	6.58**
(N:4631. F: _; 98.1; 42.7; 13.5; 20.5)	[0.70]	[2.11]	[3.50]	[4.69]	[3.20]
(4) Drop South Africa Neighbors	3.81***	8.96***	14.07***	18.30***	4.64*
(N:4413. F: _; 114.6; 36.0; 15.5; 11.7)	[0.62]	[1.97]	[3.44]	[4.69]	[2.56]
(5) Drop 1980s	4.04***	8.98***	14.95***	21.41***	2.67
(N:3631. F: _; 48.3; 25.0; 10.3; 15.1)	[0.76]	[2.35]	[4.36]	[6.91]	[2.58]
(6) Drop 2000s	3.36***	7.35***	10.84***	10.94***	5.13*
(N:2607. F: _; 99.7; 34.3; 13.8; 12.4)	[0.65]	[1.96]	[3.22]	[4.16]	[2.63]
(7) Drop Capital + Top 1960 + 2010	3.82***	8.83***	14.44***	18.97***	7.16***
(N:4431. F: _; 116.4; 31.0; 15.0; 20.8)	[0.61]	[1.97]	[3.70]	[5.20]	[2.60]
(8) (7) + Drop Top 5 Cities 1960	3.80***	8.84***	15.08***	20.38***	7.27***
(N:4150. F: _; 135.9; 28.0; 16.6; 19.4)	[0.66]	[2.07]	[4.01]	[5.69]	[2.63]
(9) (7) + Drop Top 10 Cities 1970	3.74***	9.52***	16.93***	20.86***	8.72***
(N:3716. F: _; 120.3; 22.6; 13.9; 19.1)	[0.70]	[2.16]	[4.34]	[6.33]	[3.32]
(10) Pop Estimate Available	5.22***	7.73***	9.99***	11.24***	5.19*
(N:7369. F: _; 122.3; 44.9; 27.0; 26.2)	[0.68]	[2.01]	[2.82]	[3.52]	[2.94]
(11) Pop Est. Avail. One Prev. Year	4.17***	6.86***	10.37***	12.48***	5.31*
(N:6164. F: _; 105.0; 36.3; 20.6; 17.9)	[0.67]	[1.94]	[3.07]	[3.99]	[3.02]
(12) 2 Censuses for Start & End Year	3.31***	10.97***	15.23***	16.74***	5.92*
(N:3414. F: _; 50.0; 26.1; 11.5; 9.8)	[0.66]	[2.59]	[4.26]	[6.05]	[3.22]
(13) Excl. Start & End $\geq$5 Yrs Source	3.61***	8.95***	13.92***	17.40***	4.46*
(N:4430. F: _; 122.4; 33.8; 16.7; 14.8)	[0.60]	[1.95]	[3.39]	[4.72]	[2.54]
(14) Excl. Start or End $\geq$5 Yrs Source	1.41	6.78***	8.72**	6.41	1.82
(N:1711. F: _; 55.6; 18.2; 9.9; 10.1)	[0.86]	[2.58]	[4.23]	[5.60]	[2.98]

Notes: This table is structured as Table 5 (see the table notes for details), but we implement different robustness checks. Rows 2-3: Drop 1 percent of cells with the fastest and slowest growth in $\ln pop$ and $\ln MA$, respectively. Row 4: Drop the sample countries neighboring South Africa. Rows 5-6: Drop the 1980s and the 2000s, respectively. Row 7: Drops cells containing the capital and largest city in each country in 1960 and 2010. From the row 7 sample, rows 8 and 9 drop the cells containing the 5 and 10 largest cities in 1960 and 1970, respectively, in each country. Row 10 includes all cell-year population estimates from our source data. Row 11 does this only for the year prior to crossing the 10,000 threshold. Rows 12 restricts to country-decades with two censuses, respectively, available to interpolate start and end populations. Row 13 excludes country-periods in which the start and end dates are at least five years away from the data source. Row 14 excludes decades in which the start or end dates are at least five years away from the data source. *, **, *** denote significance at the ten, five, and one percent level, respectively.

Table 7: Robustness: Threats to Identification: Spatial Co-Investments and Reallocation

		OLS	Excl. 5	Excl. 10	Excl. 15
		(1)	(2)	(3)	(4)
(1)	Baseline	3.62***	8.83***	13.74***	17.69***
	(N: 4,725. F: _; 114.0; 41.9; 17.3)	[0.59]	[1.89]	[3.31]	[4.64]
(2)	Octant: Inner: 2-3, Outer: 5-6 (Ring)	4.41***	10.60**	8.78	7.86
	(N: 1,895. F: _; 17.3; 9.1; 2.3)	[1.17]	[4.95]	[10.28]	[14.43]
(3)	Octant: Inner: 2-3, Outer: 10-11 (Ring)	4.29***	10.82***	14.21**	18.53**
	(N: 2,197. F: _; 30.1; 10.6; 2.5)	[1.03]	[3.70]	[6.34]	[9.05]
(4)	Octant: Inner: 2-3, Outer: 15-16 (Ring)	3.78***	10.63***	15.80**	21.29**
	(N: 2,262. F: _; 45.6; 13.3; 3.9)	[0.99]	[3.57]	[7.09]	[10.61]
(5)	Quartant: Inner: 2-3, Outer: 5-6 (Ring)	4.53***	12.74**	9.80	8.45
	(N: 1,688. F: _; 15.3; 10.4; 1.7)	[1.31]	[5.12]	[10.04]	[14.78]
(6)	Quartant: Inner: 2-3, Outer: 10-11 (Ring)	4.48***	13.31***	18.04**	23.81**
	(N: 1,894. F: _; 25.1; 12.6; 3.9)	[1.20]	[4.47]	[7.90]	[10.61]
(7)	Quartant: Inner: 2-3, Outer: 15-16 (Ring)	3.34***	11.34***	16.31**	21.49*
	(N: 1,899. F: _; 36.5; 13.5; 6.3)	[1.18]	[4.08]	[7.97]	[11.44]
(8)	3x3 Mega-Cells	OLS	Excl. 4	Excl. 9	Excl. 14
	(N: 4,001; 3,948; 3,948; 3,948)	5.96***	8.54***	12.94**	12.28
	(F: _ ; 33.0; 6.6; 1.0)	[0.78]	[3.20]	[5.30]	[7.98]
(9)	5x5 Mega-Cells	OLS	Excl. 3	Excl. 8	Excl. 13
	(N: 3,394; 3,316; 3,316; 3,316)	6.65***	7.25**	8.52*	9.84
	(F: _ ; 11.0; 12.9; 4.2)	[0.96]	[3.07]	[5.00]	[6.87]
(10)	7x7 Mega-Cells	OLS	Excl. 2	Excl. 7	Excl. 12
	(N: 2,868; 2,778; 2,778; 2,778)	7.52***	12.53***	16.90**	16.61*
	(F: _ ; 34.4; 4.8; 1.1)	[1.10]	[3.39]	[6.57]	[9.35]
(11)	9x9 Mega-Cells	OLS	Excl. 1	Excl. 6	Excl. 11
	(N: 2,415; 2,320; 2,320; 2,320)	9.01***	4.09	10.3	11.97
	(F: _ ; 26.0; 10.0; 3.0)	[1.17]	[3.85]	[6.40]	[10.70]

Notes: This table is structured as Table 5 (see the table notes for details), but we implement different robustness checks. Rows 2-4: We remove the observations for which road building occurred 2-3 cells from the city and 5-6, 10-11 and 15-16 cells from the city within the same octant, respectively. Rows 5-7: We remove the observations for which road building occurred 2-3 cells from the city and 5-6, 10-11 and 15-16 cells from the city within the same quartant, respectively. Rows 9-11: 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 1x1 cell of the mega-cell. As not all mega-cells have a central cell, we lose a few observations in columns (2)-(4). *, **, *** denote significance at the ten, five, and one percent level, respectively.

Table 8: Heterogeneous Effects: Economic Geography

	Col. (1)-(3): OLS			Col. (4)-(6): Excl. 5			Excl. 10	Excl. 15
	0	1	Diff. 1 - 0	0	1	Diff. 1 - 0	Diff. 1 - 0	Diff. 1 - 0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) >Median Pop. [t-30]	4.70***	2.89***	-1.81	12.85***	5.43***	-7.42**	-15.94***	-22.36***
(F: _; 40.5; 23.0; 7.1. Share 1: 0.48)	[0.65]	[1.11]	[2.79]	[2.79]	[2.09]	[3.32]	[5.65]	[7.82]
(2) >Median Pop. in Country [t-30]	3.66***	0.56	-3.10**	8.27***	3.57	-4.70	-11.57**	-13.51**
(F: _; 27.6; 23.4; 8.2. Share 1: 0.44)	[1.05]	[1.22]	[2.07]	[2.07]	[3.92]	[4.13]	[5.84]	[6.72]
(3) >Median 1960 MA in Country	3.93***	-3.09**	-7.02***	10.92***	-1.83	-12.75	-27.45**	-41.85***
(F: _; 9.4; 8.9; 7.5. Share 1: 0.49)	[0.64]	[1.47]	[1.56]	[2.51]	[8.85]	[9.27]	[10.94]	[9.72]
(4) Railroad in 1960	3.68***	3.02***	-0.66	9.38***	5.84**	-3.54	-10.14**	-13.46***
(F: _; 59.9; 19.0; 1.1. Share 1: 0.23)	[0.66]	[1.12]	[1.25]	[2.07]	[2.87]	[3.08]	[4.37]	[5.19]
(5) Paved or Improved in 1960	4.69***	3.00***	-1.69	12.50***	6.23***	-6.27**	-11.26**	-14.91**
(F: _; 27.6; 11.6; 2.6. Share 1: 0.67)	[0.90]	[0.73]	[1.11]	[2.82]	[2.09]	[3.04]	[4.90]	[6.54]
(6) <Med. Dist. to Top 1960 Cities	3.89***	2.72***	-1.17	10.96***	1.91	-9.05***	-13.10***	-16.90***
(F: _; 44.3; 8.4; 7.7. Share 1: 0.53)	[0.77]	[0.79]	[1.06]	[2.28]	[2.01]	[2.69]	[4.40]	[5.68]
(7) >Median Distance to Border	3.94***	3.17***	-0.77	8.05***	9.55***	1.49	5.04	10.61*
(F: _; 23.9; 10.9; 5.3. Share 1: 0.50)	[0.81]	[0.77]	[1.05]	[2.11]	[2.63]	[2.89]	[4.26]	[5.42]

Notes: Each row reports results from variants of columns 3, 4, and 5 of Table 4, where the three variables of interest are interacted with the dummy variable shown at left and its negation. IV5 results show the 30-year ($t-30$ to t) effect for both groups, along with the differential between them. The IV10 and IV15 columns show the differential only. The 1st stage F-statistics and the fraction of city-years with the dummy equal to one are reported in the left column. *, **, *** denote significance at the ten, five, and one percent level, respectively.

Table 9: Heterogeneous Effects: Physical and Political Geography

	Col. (1)-(3): OLS			Col. (4)-(6): Excl. 5			Excl. 10	Excl. 15
	0	1	Diff. 1 - 0	0	1	Diff. 1 - 0	Diff. 1 - 0	Diff. 1 - 0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Land Suitability >75%	3.66***	2.06	-1.6	9.38***	-2.15	-11.53**	-11.76	-18.99*
(F: _; 56.6; 20.1; 8.3. Share 1: 0.05)	[1.99]	[2.08]	[1.94]	[1.94]	[4.90]	[5.09]	[8.70]	[10.33]
(2) Land Suitability <25%	2.89**	3.82***	0.93	6.80***	14.32***	7.51	14.59**	24.90***
(F: _; 9.2; 21.9; 6.6. Share 1: 0.16)	[1.39]	[0.63]	[1.49]	[1.78]	[5.01]	[5.21]	[7.32]	[9.42]
(3) Land Suitability >50%	3.84***	3.31***	-0.53	10.33***	5.82**	-4.51	-6.07	-8.65
(F: _; 32.3; 7.5; 2.7. Share 1: 0.47)	[0.80]	[0.77]	[1.05]	[2.33]	[2.41]	[3.00]	[5.17]	[6.96]
(4) >Median Coastal Dist.	3.33***	3.50***	0.17	7.32**	9.45***	2.13	5.32	6.84
(F: _; 50.2; 16.4; 6.1. Share 1: 0.49)	[0.90]	[0.73]	[1.13]	[3.21]	[2.11]	[3.57]	[5.57]	[7.06]
(5) Leader's Origin 150km [t-30 to t]	4.33***	2.73***	-1.60	13.16***	3.80	-9.36***	-10.68*	-14.06*
(F: _; 8.4; 6.9; 7.8. Share 1: 0.34)	[0.87]	[0.78]	[1.08]	[2.54]	[2.97]	[3.58]	[6.17]	[8.30]
(6) Leader's Origin 150km [t-10 to t]	3.96***	2.22**	-1.74	10.05***	2.72	-7.32*	-8.27	-11.56*
(F: _; 15.8; 12.1; 8.8. Share 1: 0.17)	[0.68]	[1.01]	[1.19]	[1.93]	[3.99]	[4.07]	[5.06]	[6.10]
(7) Provincial Capital in 1960	3.55***	3.62***	0.08	7.93***	10.96***	3.03	2.59	7.24
(F: _; 9.8; 20.2; 5.2. Share 1: 0.16)	[0.67]	[1.05]	[1.21]	[2.22]	[3.14]	[3.56]	[5.06]	[6.92]
(8) Provincial Capital in 2010	2.96***	4.73***	1.78	5.08**	11.91***	6.83**	9.36**	14.13**
(F: _; 22.9; 8.8; 4.0. Share 1: 0.25)	[1.04]	[1.18]	[1.98]	[1.98]	[2.93]	[3.23]	[4.70]	[6.23]

Notes: Each row reports results from variants of columns 3, 4, and 5 of Table 4, where the three variables of interest are interacted with the dummy variable shown at left and its negation. IV5 results show the 30-year ($t-30$ to t) effect for both groups, along with the differential between them. The IV10 and IV15 columns show the differential only. The 1st stage F-statistics and the fraction of city-years with the dummy equal to one are reported in the left column. *, **, *** denote significance at the ten, five, and one percent level, respectively.

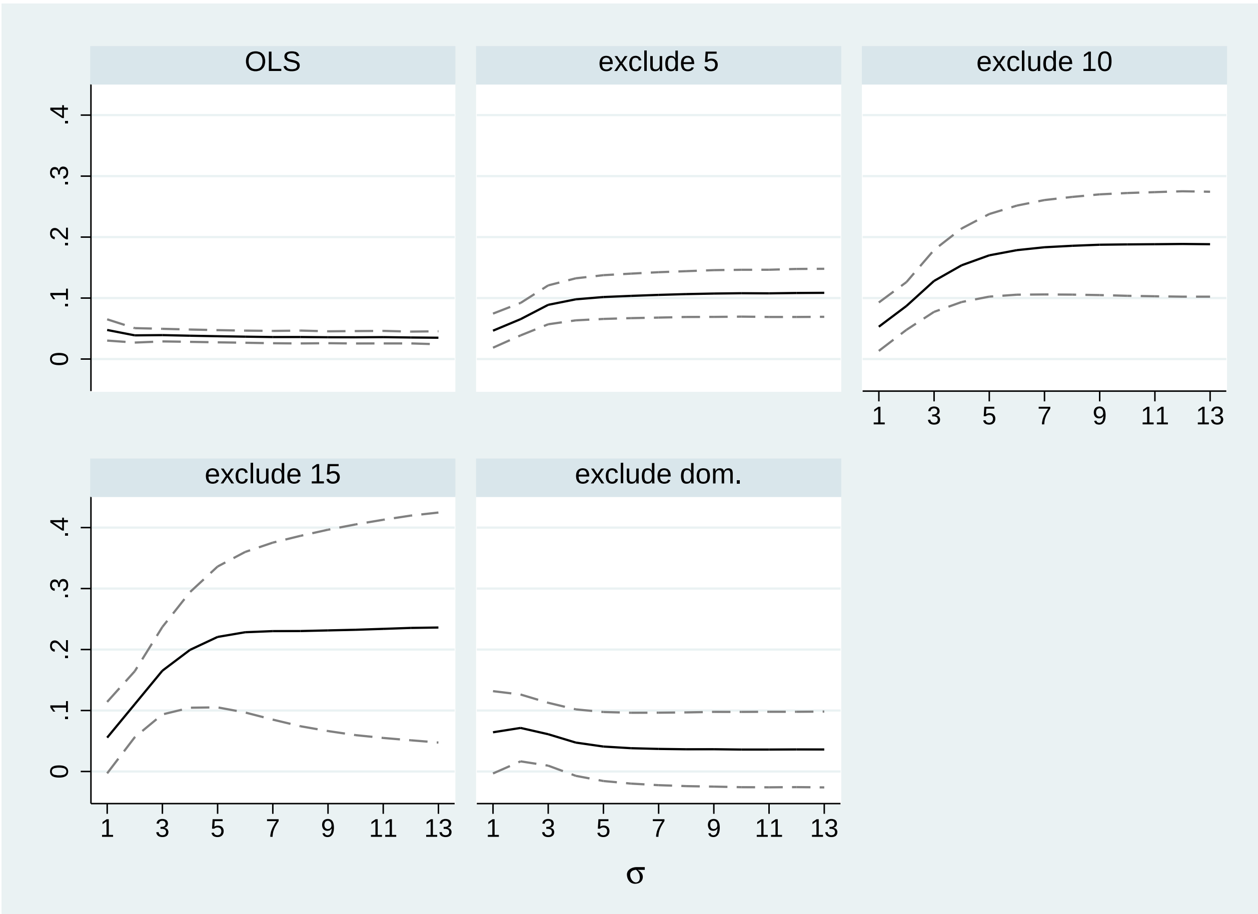
Table 10: Heterogeneous Effects: Domestic vs. Foreign Market Access

	OLS	Excl.5	Excl.10	Excl.15	Excl.Dom.
	(1)	(2)	(3)	(4)	(5)
Panel A: Overland Market Access					
(1) Baseline	3.62***	8.83***	13.74***	17.69***	4.61*
(F: _; 114.0; 41.9; 17.4; 15.1)	[0.59]	[1.89]	[3.31]	[4.64]	[2.38]
Panel B: Domestic vs. Foreign Overland Market Access					
(2) Domestic	3.45***	7.54***	8.94***	10.99***	_
(F: _; 133.9; 38.2; 18.1; 9.3)	[0.55]	[1.59]	[2.32]	[3.05]	_
(3) Domestic, Control for Foreign	3.44***	7.51***	8.94***	11.03***	_
(F: _; 142.1; 40.3; 19.3; 9.9)	[0.55]	[1.57]	[2.33]	[3.08]	_
(4) Foreign	0.37	1.13	4.45**	1.07	5.71*
(F: _; 31.6; 7.0; 20.2; 3.4)	[0.48]	[1.37]	[2.16]	[1.41]	[3.01]
(5) Foreign, Control for Domestic	0.24	0.80	4.02*	1.09	5.12*
(F: _; 31.9; 7.0; 20.1; 3.3)	[0.47]	[1.34]	[2.06]	[1.36]	[2.85]
Panel C: Overseas vs. Overland Market Access					
(6) Baseline, Control for Overseas	3.35***	10.96***	21.26***	30.90***	4.36
(F: _; 41.1; 10.3; 3.0; 12.4)	[0.64]	[2.93]	[6.74]	[11.15]	[3.00]
(7) Overseas	9.21***	10.41***	13.16***	17.94***	16.51**
(F: _; 122.9; 85.6; 43.7; 34.2)	[2.15]	[2.98]	[3.93]	[5.31]	[8.10]
(8) Overseas, Control for Overland	4.35*	6.59*	10.30**	15.35***	13.53
(F: _; 102.0; 84.9; 47.0; 37.1)	[2.36]	[3.37]	[4.40]	[5.85]	[9.48]

Notes: This table is structured as Table 5 (see the table notes for details), but we use alternative sets of market access measures based on different subsamples of the city data set. The instruments are then constructed using the same subsamples of the city data set. Rows 1 and 8: We use the main market access measures capturing overland market access (to both domestic and foreign cities). Rows 2-3: The market access measures are calculated using the population of domestic cities only. Rows 4-5: The market access measures are calculated using the population of foreign cities only. Rows 3 and 5: We control for the foreign and domestic market access measures, respectively. Rows 6-8: The overseas market access measures are calculated using the population of the 44 cities with an international port (circa 2010) only. Row 6 show the baseline results when controlling for overseas market access. Rows 7-8 show the effects of overseas market access, while simultaneously dropping the 44 cities with an international port from the regressions. Row 8: We control for the overland market access measures. *, **, *** denote significance at the ten, five, and one percent level, respectively.

A Web Appendix: Not for Publication

Figure A.1: Relationship between Sigma and the Standardized Overall Effect



Notes: This figure shows the overall effect (from t-30 to t) of a one standard deviation increase in $\Delta_{t-10}^{t} \ln MA$, $\Delta_{t-20}^{t-10} \ln MA$, and $\Delta_{t-30}^{t-20} \ln MA$ on Δ_{t-10}^{t} ln urban population)/100 by value of σ = [1; 13], with 95% confidence interval, for each of the 5 main identification strategies.