

FOR ONLINE PUBLICATION: Appendix

A.1 Roads data

According to the Michelin map legends, paved roads comprise roads that are “hard surfaced (asphalt, concrete, etc.).” Several maps distinguish paved roads with one lane only from paved roads with two lanes or more, but not consistently across years. Improved roads include both fully improved roads (“suitable for high speeds in certain sections. Regularly maintained with mechanical equipment”) and partially improved roads (“improvement is mainly confined to the difficult sections”). Many maps distinguish fully and partially improved roads, but not consistently across years. We thus use only the four categories listed, which can be consistently distinguished. In early maps, Michelin also designates some road segments as “transcontinental”. We record this information for the first map in each country. Figure A.3 shows Sierra Leone in the 1969 Michelin West Africa map and the associated GIS map.

A.2 Data on Cities, 1960-2010

We obtained population estimates of cities in 42 countries from the 1950s to the 2010s. The sources used for each country-year are listed in Table A.1.¹ For the 6 countries not in the Africapolis samples (Angola, Botswana, Malawi, Mozambique, Zambia and Zimbabwe), we use analogous sources and methods. For 5 country-years in the 1950s and 1960s, we rely on a “non-native”/European population census (WC) and assumptions regarding the population share of non-natives/Europeans in each city, in the absence of data on total population. Further assumptions were required about the population of some cities in Angola and Mozambique (JS = Our estimate). 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 sample. Results reported in Table A.8 and discussed in Section A.9 drop country-years most likely to be subject to measurement error due to these and similar concerns.

Our 2,911 cities fall within 2,789 0.1x0.1 degree grid cells, our unit of analysis, in the 39 countries of our main sample. 2,116 of these cells have estimates for at least two consecutive decades, allowing us to calculate population changes. Results reported in Table A.8 and discussed in Section A.9 include city population estimates below 10,000 where available. We refer to the capital, largest, and second largest cities in each country-year as “national cities” or “top cities”.

A.3 Construction of Market Access

Section 2.1. in the main text discusses the construction of market access. Figure A.5 shows an example for Sierra Leone between 1970 (shown in Figure A.3) and 1980. (i) We first obtain the roads in GIS for

¹For the two country-years for which an electoral census is the source of population data in the 2000s, Africapolis makes assumptions about the age structure of these cities (since voters had to be at least 18 to vote in these countries) to reconstruct the total population. Since these population estimates are subject to more measurement error, we show in row 14 of Table A.8 that our main results are robust to not using the population estimates from the 2000s.

1969 (top left), 1971, 1976 and 1983 (top middle). (ii) We then assign the time cost of traversing the cell, using the speed of the fastest road surface passing through the cell, as the cell's value (bottom left), interpolating between 1969 and 1971, and 1976 and 1983, to obtain values for 1970 and 1980, respectively. (iii) Analogously, we assign each city's populations to the cell in which it falls (top right). (iv) We then use Dijkstra's algorithm to obtain the least cost path (LCP) from each city/cell to each other city/cell. For example, we show here the LCPs from Bo to all other cities (bottom middle). (v) Using the populations, this trade cost proxy between each pair of cities, and θ , we calculate market access in the first year and the second year, and the change in market access (bottom right). Figure A.6 shows the total change in market access between 1960 and 2010 for each of the 187,900 cells among the 39 countries of the main sample.

A.4 Demographic rates

To estimate birth and death rates, the two components of population change besides migration, we use data from the Demographic and Health Surveys (DHS). The DHS is the largest demographic survey program in the world. Since 1984 it has carried out hundreds of surveys in nearly a hundred countries. In all of our sample countries where such surveys have taken place, they include a full birth history for a nationally representative sample of women aged 15 to 49. Many surveys, especially since the mid-1990s, include geographic coordinates of sampled clusters, allowing us to assign them to individual cities.

We select 16 countries based on the following three criteria for the availability of georeferenced surveys, in order to calculate birth and death rates for the 1990s and 2000s, and in 8 countries, the 1980s. First, a country must have a survey with fieldwork beginning December 2009 or later, in order to capture 2009 births. Second, it must have a survey completing fieldwork in December 2005 or earlier, in order to capture all births by 33-year-olds in 1990. We limit our sample to women aged 15 to 33 as a compromise between capturing births at all ages and including countries that did not have an appropriate survey before 2003. Third, a country must have a survey no more than 16 years before its first post-2009 survey, in order to capture births at that time by women too old to be surveyed in the post-2009 survey. The survey fulfilling this third criterion can be the same as the one fulfilling the second criterion, depending on the timing of the one fulfilling the first criterion. Sixteen countries meet these criteria: Benin, Burkina Faso, Cameroon, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Kenya, Malawi, Mali, Namibia, Nigeria, Senegal, Tanzania, Uganda, and Zimbabwe. They have a total of 51 relevant surveys between 1990 and 2014 (excluding one particularly small survey in Burkina Faso and one in Mali). Each country has two to five relevant surveys; all but Tanzania have at least three.

We limit the DHS sample to urban survey clusters, and assign them to the nearest one of our city-cells in the same country if it is within 15 km of the city-cell's centroid. We use this 15 km buffer to account for a) the distance of up to 8 km between a cell centroid and its corner, b) up to 2 km of

random error added to urban survey clusters by DHS to promote anonymity, and c) finite width of cities. Survey clusters not within 15 km are dropped.

Each DHS survey is nationally representative, and most are representative for subnational regions, but none is explicitly representative at the level of individual cities. A country typically surveys entirely different clusters in consecutive surveys, and therefore many of our sample cities, especially the small ones, are sampled in no surveys or only one. Even those that are surveyed may have a small sample of women.

Having thus defined our samples, we calculate a crude fertility rate (CFR) for each city-decade, where the denominator is the woman-years of exposure in the age range 15-33 in a given decade, and the numerator is the number of births to these women in these years. Both are weighted using DHS sampling weights, to account for the DHS' two-stage clustered sampling design. The infant mortality rate (IMR) is calculated using the same sample of births, excluding those in the last year of a decade and those in the year prior to the survey, to avoid censoring. The infant mortality rate is the share of those live births in which the infant died before age one.

Using data from the World Development Indicators (WDI), we multiply each city-decade's CFR by its country's share of women 15-33 in the overall population (calculated as the share 15-29 plus 0.8 times the share 30-34) to get estimates of each city-decade's crude birth rate (CBR). Again using WDI data for the sample countries and period in five-year increments, we regress the crude death rate (CDR) on IMR and year and country fixed effects. We apply the slope and fixed effects from that regression to our city-decade IMR values to generate predicted CDR for each city-decade. The city-decade-specific rate of natural increase (RNI) is then simply the CBR minus the CDR. Given the small values of RNI involved (so that the approximation $x \approx \ln(1 + x)$ works well), we interpret these percentage changes as good approximations of the natural increase component of $\Delta \ln(\text{population})$.

A.5 Other Data

Rainfall. Historical precipitation data are from Willmott and Matsuura (2009).² We calculate average annual precipitation (in mm) over the period 1900 to 1960 for each cell.

Land Suitability for All/Food/Cash Crops. Land suitability for food crops and cash crops are from IIASA and FAO (2012), based on geographical characteristics circa 2010.³ We use the "Water supply: rain-fed" and "Input level: Low input level" variants, as the use of irrigation, fertilizer and other inputs is low compared to the rest of the world. We calculate the share of land in each cell, alone and with its neighbors, that is suitable for at least one crop, separately for food crops, cash crops, and all crops. We also define as "cash crop cells" cells for which land suitability for cash crops is above 90%.

²Available at http://climate.geog.udel.edu/~climate/html_pages/archive.html

³Available at <http://www.fao.org/nr/gaez/en/>. Food crops: Buckwheat, Barley, Foxtail millet, Maize, Oat, Pearl millet, Indica dryland rice, Wetland rice, Rye, Sorghum, Wheat, Chickpea, Cowpea, Green gram, Dry pea, Phaseolus bean, Pigeon pea, Cabbage, Carrot, Cassava, Onion, Sweet potato, Tomato, White Potato, Yams. Cash crops: Cacao, Coffee, Coconut, Cotton, Flax, Groundnut, Jatropha, Oil palm, Olive, Rape, Sunflower, Soybean, Tea, Tobacco, Citrus, Sugarbeet, Sugarcane.

Mines. Eros and Candelario-Quintana (2006) shows the locations of 303 mines in the 39 countries. For 288 mines, we use various sources to find the year of opening. We exclude the remaining 15 mines as we did not find the year of opening, but verified that they are small in terms of quantity produced.

Provincial/Regional Capitals and Boundaries. The *Statoids* database of Law (n.d.) and Wikipedia (n.d.) provide information on first-level administrative unit (“province”/“region”) capitals in 1960 and in 2010. The sample includes 343 provincial capitals in 1960 and 481 in 2010. Current boundaries are from <https://www.diva-gis.org/gdata>; 2010 and 1960 are reaggreated from those, reallocating underlying districts as necessary.

Railroads. The location of all railroad lines in sub-Saharan Africa, and the year each entered service, are from Jedwab and Moradi (2016).

Ports. The traffic of the 44 main international ports in sub-Saharan Africa circa 2005, are from Ocean Shipping Consultants, Ltd. (2009). Passenger traffic (000s), cargo entering (i.e. imports, kilotons) and cargo leaving (i.e. exports, kilotons) for 65 main international ports circa 1960 are from Ady (1965).

Airports. The locations of 466 main airports (civilian/public, joint military/civilian, military and other) in sub-Saharan Africa are from National Imagery and Mapping Agency (n.d.).

Customs Posts/Border Crossings. The locations of 837 customs posts circa 2010 are from the most recent Michelin map in each region. We treat these as proxies for economically relevant border crossings.

Natural Parks. Natural park boundaries c. 2015 are from World Database on Protected Areas (2015). We define as “natural park cells” the 26,252 cells in which more than 50% of the area belongs to a natural park.

Leaders and historical ethnic group boundaries. We collected the locality of birth and ethnicity of the 189 heads of state of the 39 sample countries between 1960 and 2010. Our main sources are the English and French versions of Wikipedia (n.d.), verified when possible 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 focus on selected countries and/or periods. To our knowledge, we are the first to collect these data for virtually all of sub-Saharan Africa from independence to date. We also record the main historic ethnic group in each cell according to two group classifications: the Murdock (1959) map and the GREG map compiled by Weidmann et al. (2010). We check whether each head of state’s locality of birth is in his or her ethnic homeland. When they differ, typically for those born abroad or in the capital city, we define locality of origin as the historical residence of the parents (as indicated by Wikipedia (n.d.)) if it is within the boundaries of the leader’s ethnic group in Murdock (1959) and/or GREG. Otherwise, we use the centroid of the ethnic homeland in either Murdock (1959) or GREG.

Night Lights. Night lights data for 1992 (which we use as a proxy for 1990), 2000, and 2010 are produced and distributed by the US National Oceanic and Atmospheric Administration, based on data from the

Defense Meteorological Satellite Program.⁴ These datasets provide annual average estimates of light emitted into space from each 30-second pixel (approximately 1 square km at the equator). We average these values across all pixels on land within each of our 0.1 degree grid squares.

Conflict. Conflict date and location data are from the Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset (GED) version 19.1 (Sundberg and Melander, 2013). Our conflict variable indicates whether a cell-year contained area within 25 km of a conflict, with a best estimate of deaths greater than zero, and a spatial precision that is either a point, a 25-km radius, or a second subnational level administrative unit, that occurred between year $t-9$ and year t (inclusive). This means that events with no deaths and events with a coarser spatial precision (first-level administrative unit, country, or a residual category including along a river or near a border) are excluded. We chose this specification carefully. Summing deaths would be unwise given that we cannot distinguish the area over which the deaths occurred. Similarly, counting conflict events would not be appropriate, given extreme differences in intensity. Including events with no deaths would be more subject to bias in terms of the level of press coverage allowing the event to be found by the authors of the database, and limiting to conflicts with deaths is standard in the literature using this dataset. Including the roughly 10% of events coded at the first subnational administrative level is possible, but all the conflicts are coded as point locations, so it's not clear what spatial extent to use for such cases, especially when province borders have changed. The problem of how to assign a conflict to a location is even more salient for the residual category of spatial precision (e.g. along a river or near a border), and there are even fewer of these (4% of conflict events). Country-level spatial precision is easy to map, but it doesn't mean the conflict actually took place throughout the country, just that the location within the country is unknown, and they represent only about 1% of conflicts.

Trans-African Highway (TAH) Network. We assigned segments in our main road database to the planned TAH using detailed maps from ADB and UNECA (2003).

Total Population. Cell-level estimates of total population for 1960–1990 are from Nelson and Deichmann (2004) and those for 2000–2010 are from CIESIN (2016).

Other City-Level Controls. We use several other physical geographic characteristics following Jedwab and Moradi (2016). Cell-level mean and standard deviation of elevation, in meters is calculated from the Shuttle Radar Topography Mission version 3 (SRTM3 DTED1) 90-meter data (Farr et al., 2007).⁵ Distance from each cell to a river is calculated using navigable rivers from VMAP/GlobalGIS.⁶

Urbanization Rates. United Nations (2015a) provides the total population of each country in each sample year. We use this total and the population of cities over 10,000 in our sample to calculate consistently defined national urbanization rates.

⁴Available at <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>.

⁵Available at <http://www2.jpl.nasa.gov/srtm/>

⁶Available at http://www.agiweb.org/pubs/globalgis/metadata_qr/perennial_rivers_qk_ref.html

Per capita GDP. Per capita GDP (1990 International Geary-Khamis \$) 1950–2010 comes from Bolt and van Zanden (2014), based on Maddison (2008), and in some cases from International Monetary Fund (2019). Values for a few countries for 2010 are constructed using 2008 values and per capita PPP GDP growth rates from World Bank (2016).

Democracy. Polity IV (2015) reports a “Combined Polity Score” for each country-year on a -10 to +10 scale, with scores strictly below 5 classified as non-democratic.

Wars and Refugees. Center for Systemic Peace (2015b) reports international and civil wars, and Center for Systemic Peace (2015a) reports the number of foreign refugees hosted, for each country-year.

Droughts. CRED (2013) reports droughts by country-year.

A.6 Main Sample, Descriptive Statistics and Baseline Specification

Our main sample consists of 4,725 city-decades in which the city had a population of at least 10,000 inhabitants in both the initial and final years. The covered decades are the 1980s, 1990s, and 2000s (2000-2010). The two prior decades are dropped due to lags in our main specification. The descriptive statistics of the main variables are shown in Table A.3.

Our vector formulation of market access for all cities in year t , MA_t , can be written as the product of the matrix T_t of transformed transport costs between all N cities in year t and the column vector P_t of all N city populations in year t ($MA_t = T_t P_t$) more explicitly as follows:

$$\begin{bmatrix} MA_{1t} \\ MA_{2t} \\ \vdots \\ MA_{N-1,t} \\ MA_{Nt} \end{bmatrix} = \begin{bmatrix} \sum_{d \neq 1} P_{dt} \tau_{1dt}^{-\theta} \\ \sum_{d \neq 2} P_{dt} \tau_{2dt}^{-\theta} \\ \vdots \\ \sum_{d \neq N-1} P_{dt} \tau_{N-1,dt}^{-\theta} \\ \sum_{d \neq N} P_{dt} \tau_{Ndt}^{-\theta} \end{bmatrix} = \begin{bmatrix} 0 & \tau_{12t}^{-\theta} & \cdots & \cdots & \tau_{1Nt}^{-\theta} \\ \tau_{21t}^{-\theta} & 0 & \cdots & \cdots & \tau_{2Nt}^{-\theta} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & \tau_{N-1,N,t}^{-\theta} \\ \tau_{N1t}^{-\theta} & \tau_{N2t}^{-\theta} & \cdots & \tau_{N,N-1,t}^{-\theta} & 0 \end{bmatrix} \begin{bmatrix} P_{1t} \\ P_{2t} \\ \vdots \\ P_{N-1,t} \\ P_{Nt} \end{bmatrix} \quad (\text{A.1})$$

where the first equality applies the definition of city o 's market access at the beginning of section 2.1. to each element of the vector.

IV Specification with lead and third lag. Table A.5 Panel A shows the effect of the lead of the change in market access ($t;t+10$) when using the three IV strategies. It is small and never significant. Panel B includes a third lag. It is also small and insignificant, and the overall effect differs little from the baseline, though its inclusion does result in a weaker instrument and smaller and less precise contemporaneous terms.

A.7 Instrumental Variables Strategy and Measurement Error

Given the construction of the instruments, their measurement error is likely correlated with measurement error in the endogenous variables in our context. In this section we consider the econometric implications of this, starting from the simplest possible model:

$$y = \beta x + \epsilon \quad (\text{A.2})$$

where we only have data on

$$\tilde{x} = x + u. \quad (\text{A.3})$$

We make the classical measurement error assumptions:

$$E(u) = 0 \quad (\text{A.4})$$

$$\text{cov}(x, u) = 0 \quad (\text{A.5})$$

$$\text{cov}(\epsilon, u) = 0. \quad (\text{A.6})$$

We further assume that x is endogenous:

$$\text{cov}(\epsilon, x) \neq 0 \quad (\text{A.7})$$

and that our instrument, z is exogenous, and strong:

$$\text{cov}(x, z) > 0 \quad (\text{A.8})$$

$$\text{cov}(\epsilon, z) = 0. \quad (\text{A.9})$$

Departing here from standard treatments of measurement error with instruments, we assume that our IV is mismeasured:

$$\tilde{z} = z + v, \quad (\text{A.10})$$

that the classical measurement error assumptions extend to z :

$$E(v) = 0 \quad (\text{A.11})$$

$$\text{cov}(z, v) = 0 \quad (\text{A.12})$$

$$\text{cov}(\epsilon, v) = 0, \quad (\text{A.13})$$

and that measurement error in the instrument z is *positively* correlated with the measurement error in x :

$$\text{cov}(u, v) > 0. \quad (\text{A.14})$$

Under the assumptions above, it can be shown that there is still attenuation bias in the IV estimator:

$$\hat{\beta}_{IV} = \frac{\text{cov}(y, \tilde{z})}{\text{cov}(\tilde{x}, \tilde{z})} = \frac{\text{cov}(\beta x + \epsilon, z + v)}{\text{cov}(x + u, z + v)} = \frac{\beta \sigma_{xz} + \beta \sigma_{xv} + \sigma_{\epsilon z} + \sigma_{\epsilon v}}{\sigma_{xz} + \sigma_{xv} + \sigma_{uz} + \sigma_{uv}} = \frac{\beta \sigma_{xz}}{\sigma_{xz} + \sigma_{uv}} \quad (\text{A.15})$$

Define IV attenuation bias:

$$\lambda_{IV} = \hat{\beta}_{IV} / \beta = \frac{\sigma_{xz}}{\sigma_{xz} + \sigma_{uv}} \quad (\text{A.16})$$

It is well known that OLS attenuation bias is given by:

$$\lambda_{OLS} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}. \quad (\text{A.17})$$

Therefore, instrumenting reduces the affect of attenuation bias iff:

$$\frac{\sigma_x^2}{\sigma_u^2} < \frac{\sigma_{xz}}{\sigma_{uv}} \quad (\text{A.18})$$

In words, instrumenting reduces attenuation bias if and only if instrument strength (σ_{xz}) relative to correlation of the errors (σ_{uv}) is more than the signal-to-noise ratio in x ($\frac{\sigma_x^2}{\sigma_u^2}$). We expect that measurement error in IV15 will be less correlated with measurement error in overall MA than measurement error in IV5 is, but IV5 is also stronger than IV15. It is thus ambiguous how $\frac{\sigma_{xz}}{\sigma_{uv}}$ varies between IV5 and IV15.

A.8 Robustness Checks for the Main Identification Strategies

Table A.6 reports several robustness checks related to the main identification strategies.

Co-Investment and radial extension. Rows 1–15 vary the definitions of the inner and outer rings in excluding cities due to potential co-investment or radial extension outward or inward. In rows 1–6, the inner ring is still 2–3 cells but the outer ring is 10–11 or 15–16. By construction, IV10 and IV15 (IV15) already exclude co-investment in an outer ring of 5–6 (10–11) cells, so these combinations are not reported. The sample is reduced by more than 50%, and the instrument excluding up to a radius of 15 cells is weak, but results are generally consistent with the baseline. Radial extension inward requires dropping the 1980s because we don't know whether second lag changes in the 1960s follow inner changes in the 1950s. Excluding observations with potential extension inward decreases the sample size by a further 40–48%. This is less than the analogous 56–66% of observations with radial extension outward, consistent with the idea that on average road improvement proceeded outward away from cities over time, in line with the patterns of road decentralization described in Section 3.2.. In rows 7–15, results are similar with a 1–2 cell inner ring (i.e. cells closer to the city).

Road Construction by Distance to Cities by Decade. For the 2,126 cells in our sample, we show that road changes are more likely to take place closer to the city in early decades relative to later decades. More specifically, for 2,126 cells x 5 decades x 30 rings (0–1 cell away, 1-2, ... 29–30) = 318,900 observations, we estimate the share of total changes in road costs in that decade that comes from each ring. Note that a higher road cost means a lower-quality road. Appendix Figure A.7 shows the shares of the 1–29 rings by decade, dropping the central cell. Roads were built closer to cities in the 1960s and 1970s, and also in the 2000s when very few roads were built. This is again broadly in line with the patterns of road decentralization described in Section 3.2..⁷

In row 16, results are also similar if the convex hull is constructed for the nearest city of at least 50 thousand. The resulting exclusion zones are weakly smaller since the nearest city of 50 thousand or more is at least as close as the nearest city of 100 thousand or more. Rows 17–19 use older

⁷Regressing share on ring by decades, the 1970s have the steepest average slope (-0.00074***), followed by the 1960s (-0.00045***), the 2000s (-0.00017**), the 1980s (-0.00012***), and the 1990s (j0.00001).

populations in constructing market access. Rows 17 and 18 hold population fixed at its 1960 level in the instruments alone, and in the instruments and instrumented MA, respectively. In both cases, the 1980s are dropped because they are the only decade in which an included lag of ΔMA is constructed using the population of the other cities in 1960. In row 19, instrumented MA is constructed using beginning-of-period population (the instrument are always constructed in this manner). In all cases, results are similar to baseline. Row 20 adds a control for growth in a 16–30 cell ring to the specification controlling for local growth in Table 2; results are similar to that specification.

Table A.7 reports specifications that drop cities especially likely to see endogenous road-building to them, and add related controls.

Drop Isolated Cities. Effects are similar but slightly smaller if we: (i) drop “isolated” city-years, i.e. urban cells that are in the top 25% of the Euclidean distance to the nearest cell with a city at any point between 1960 and 2010 (row 1); or (ii) drop country-years where cities are relatively far from each other on average, i.e. country-years where the average Euclidean distance between cells with a city at any point between 1960 and 2010 is in the top 25% in the sample (row 2). The fact that the effects are smaller when excluding less connected cities or countries is consistent with results on heterogeneous effects.⁸

Drop Potential Growth Hubs. In rows 3–11, we drop selected cities with observable characteristics that may cause them to grow and to have roads to be built towards them, even from far away. In row 3, we drop from the sample each country’s 5 largest cities and national and regional capitals from both 1960 and 2010 (instead of 1960 only as in the main text). Results are similar to though a bit smaller than the 1960-only version in the main text. However, if roads promote city growth, and larger cities are more likely to become regional capitals, we may under-estimate the effects when dropping the new ones.

In rows 4–11, we drop any city that is within 100 km of: (4) a national city (the capital, largest, and second largest cities of each country in either 1960 or 2010); (5) a mine open at any time between 1960 and 2010; (6) a cell whose land suitability for cash crops is above 90%; (7) the hometown of any of the country’s heads of state between 1960 and 2010; (8) a port in either 1960 or 2005; (9) an airport in 2007; (10) a border crossing in 2010; and (11) a natural park cell in 2015. Results are generally consistent with the baseline.

Outliers. Rows 12–13 drop the cell-years with the largest and smallest one percent of increases in population and market access, respectively, to ensure that outliers are not driving results. Indeed, outliers might be more endogenous, as the specific unobservable factors that explain their very slow/fast population/market access growth could be correlated with our road changes far away, thus invalidating the exclusion restriction. Results are generally consistent with the baseline.

⁸We use cells that are “urban” at any point in 1960–2010 because we do not want our measures of city or country isolation to be mismeasured simply because we do not want have good population data below 10,000.

Regional mean reversion. In rows 14–15, we account explicitly for regional mean reversion. In row 14, we interact the lag of log population with country-year fixed effects, as the importance of local increasing returns/mean reversion may have been changing differentially across countries over time. In row 15, we control for log market access in 1960, as governments may have been targeting places that were relatively less connected initially. Results are generally consistent with the baseline.

More city-level controls. Row 16 adds many controls proxying for physical, economic and political geography.⁹ Row 17 controls for the best surface in a city's own cell. Row 18 includes an indicator for a conflict in a city-decade. Row 19 includes a control for national GDP growth in the nearest neighboring country. In all four cases, coefficients are very similar to baseline.

Rows 1–8 of Table A.8 drop country-years that might be different in important ways that affect the exclusion restriction.

Colonies. Countries that are still colonies may see more road and non-road investments towards the capital city or other places of interest to the colonial government. The exclusion restriction may not be satisfied, despite the fact that we include country-year fixed effects. However, rows 1–2 of Table A.8 shows results are similar if we drop countries that became independent after 1970 or 1960.

Wars. Rows 3–5 show that results are similar if we drop country-decades associated with: (3) interstate or independence wars; (4) civil wars; or (5) any wars. In the last two cases, the 15-cell instrument is substantially weaker and the coefficient shrinks more.

Refugees. Countries can also be indirectly affected by a war through refugee inflows. If refugees live in camps that drive city growth, and roads are built towards refugee camps, the exclusion restriction may not be satisfied, despite the fact that we include country-year fixed effects. Since we do not have historical data on the location and population size of refugee camps, we instead drop country-decades most likely to be affected by refugees. Rows 6–7 of Table A.8 show that results are similar if we drop the country-decades in which the (6) mean and (7) maximum annual number of refugees during the decade is larger than their respective sample means (89,000 and 253,000).

Multi-year Droughts. Row 8 shows that results are similar if we drop the country-decades in which a multi-year drought (i.e., a drought that lasted at least two years) took place.

A.9 Robustness Checks: Measurement Error

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

⁹The controls include dummies if the cell contains the capital / largest / second largest city or a regional capital in 1960 or 2010, and the log of the Euclidean distances to these cities, dummies if the cell is within 100 km from a top city in 1960 or 2010, a mine, a cash crop region, a president's hometown, a port in 1960 or 2005, an airport in 2007, a border crossing in 2010, or a natural park in 2015, and the log of the Euclidean distances to these locations, dummies if the cell is on the coast or crossed by a river, and the log of the Euclidean distances to the coast/a river, the mean and standard deviation of altitude (to control for ruggedness), and average rainfall in 1900–1960.

been different from other cities in ways that are correlated with road building. Second, it requires interpolation and extrapolation, sometimes several years away from censuses or other estimates. This affects both the dependent variable and the variables of interest. As long as this interpolation and extrapolation does not systematically overestimate or underestimate either of these, measurement error will be classical, biasing estimates downward.

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

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

Rows 15–18 restrict the sample to country-decades with the population estimates most likely to be reliable. Row 15 restricts the sample to country-decades whose beginning and end populations are each based on at least two *census* populations, as opposed to other sources. Row 16 excludes country-decades for which the initial and final populations are both at least 5 years from a population data source. Results are similar to baseline. Row 17 excludes country-decades for which the initial *or* final populations are more than 5 years, respectively, from a population data source, reducing the sample by more than 50 percent. The point estimates are reduced as well, and while the 15-cell instrument is weakened somewhat, the 5-cell and 10-cell instruments remain strong and suggest effects that are significant, if reduced by up to a third from baseline. Finally, row 18 restricts the sample to the 33 countries with Africapolis data, with little effect.

Road data quality. Rows 13 and 14 showed that results are similar when excluding data from the 1960s and the 2000s, two decades when the road data are most likely to be incomplete. In addition, we verify that results are similar if we drop country-decades or countries that we believe poor road data. More precisely, for each of the 39 countries \times 5 decades = 195 country-decades in the sample, we compute the growth rate in total paved road kilometers (recalling that paved roads were the main

form of investment during our period), and compare it to an analogous estimate from official national sources compiled by Canning (1998), or World Bank (2015) when no estimate is available from Canning (1998). The mean of the absolute value of this difference is about 0.2, meaning that we typically over- or under-estimate paved road building by 20%. Results are broadly similar if we drop: (i) country-decades for which the absolute value difference is more than the mean in our main sample (row 19); (ii) whole countries for which the mean absolute value difference is more than the mean in the sample (row 20). In rows 21 and 22, we also drop country-decades or countries for which official data is missing.

A.10 Robustness Checks: Functional form

Table A.9 explores the robustness of results to other changes to our main specifications.

Our baseline specification controls for initial log population, since it is standard in the literature. However, its potential endogeneity is a concern. It also controls for country-year fixed effects and year-specific polynomials in longitude and latitude. Row 1 of Table A.9 shows that the baseline OLS estimate is larger, and the IV10 and IV15 estimates somewhat smaller, with no controls. Rows 2 and 3 progressively add the other controls, and row 4 replaces country-decade fixed effects with decade fixed effects, with little change. Row 5 clusters standard errors at the country level, to account for the fact that much of the data is collected by country. The estimated standard errors are only slightly larger.

Our baseline specification with three distinct periods of market access change allows effects to vary across three decades during and after road construction. A challenge of this specification is that three instruments are needed to identify the three endogenous variables. In row 6, we investigate this tradeoff by collapsing these three changes into one change from $t - 30$ to t . The instrument set, with only one instrument for one endogenous variable, is substantially stronger, but point estimates fall by about two-thirds. This likely reflects two aspects of the change. First, pooling doubles the variance of $\Delta \ln(\text{Market Access})$ across observations, while the variance of city growth remains the same. In standardized terms, 30-year market access has half as large an effect as the pooled 10-year changes (a one standard deviation in market access growth is associated with a 0.25-0.44 standard deviation increase in population growth vs. 0.46-0.88 when using 10-year market access). Second, and more importantly, there is meaningful decadal variation in market access during the $(t-30,t)$ period that is being averaged away, imposing the restriction that changes in $(t-30,t-20)$, $(t-20,t-10)$ and $(t-10,t)$ all have the same effects on population growth in $(t-10,t)$. However, Table 1 has shown that the contemporaneous effect differs from the lagged effects. Table 3 will show that effects differ even more strongly across lags when studying night light growth. Allowing effects to vary over time is thus essential.

Rows 7 and 8 report specifications reported by Donaldson and Hornbeck (2016). Row 7 reports the reduced form effects of the instruments on city population growth. We prefer the IV specification

because it is more comparable with the rest of the literature. Row 8 shows that results are similar if we weight observations by their initial population size in $t - 10$.

Market Access Measures. Rows 9 and 10 replace the speeds assumed at baseline with those used by Alder (2017) and Shiferaw et al. (2013) (see Table A.2). Magnitudes increase modestly. In row 11, we allow railroad travel (at the speed of paved roads) in constructing our market access. Next, our baseline analysis assumes $\theta = 3.8$ in defining market access. Assigning larger (smaller) values of θ mechanically increases (decreases) the coefficients on market access, as it shrinks (widens) the variation in market access, without altering the variation in city growth. This can be seen in rows 12 and 13 where we use $\theta = 8$ and $\theta = 2$ respectively. However, as in Donaldson and Hornbeck (2016), the effect of a one standard deviation increase in $\Delta_{t-10}^t \ln MA$, $\Delta_{t-20}^{t-10} \ln MA$, and $\Delta_{t-30}^{t-20} \ln MA$ is quite stable for values of $\theta \geq 2$ (Figure A.8). Our results thus do not depend on the choice of a specific trade elasticity. We also verify that higher trade elasticities mechanically weaken our instruments (available upon request), rendering far away road changes less relevant. Lastly, we find similar effects if we assign border crossing costs of 4 or 24 hours (rows 14–15). Unfortunately, no data exist on specific border crossing costs for the 39 sample countries from 1960 to 2010.

Iceberg Costs. Row 16 replaces our measures of costs with one more closely related to the iceberg cost of Donaldson and Hornbeck (2016). Teravaninthorn and Raballand (2008) report costs of \$2–\$10 per km for trucks across several routes in sub-Saharan Africa. This corresponds well with the factor of five difference in unit costs we assume between paved and dirt costs. We thus assign costs of \$1.5, \$2, \$3, and \$10 per km to highways, paved, improved and dirt roads, respectively. The iceberg cost measure is 1 plus this dollar value divided by the value of goods transported, for which we assume \$10,000. We do not pursue this as a main specification because we expect substantial variation in the value of goods transported. The iceberg specification allows us to include a city's own population in its measure of market access (row 17). Coefficients in both rows are larger. In part this reflects lower variation in market access. However, the effect of a one standard deviation increase in market access is now a 0.24–0.31 standard deviation increase in population (vs. 0.43–0.85 in the baseline regressions).

Edge Countries. 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 true in the four sample countries that border South Africa: Botswana, Mozambique, Namibia, and Zimbabwe. In row 18, removing these countries has little effect. Likewise, magnitudes are broadly similar but slightly smaller if we drop countries nearest to North Africa (Chad, Mali, Mauritania, Niger and Sudan; row 19) or the Arabian Peninsula (Djibouti, Eritrea, Somalia and Sudan; row 20).

Surface-Specific Effects. In rows 21–22, defining market access changes due to paving alone and improving alone suggests that if anything market access changes due to improving have slightly larger effects. This difference is not statistically significant, and of course it is sensitive to the speed assigned to each surface.

A.11 Robustness Checks: Access to Other Characteristics, Night Lights, and Reallocation

Market access to other characteristics. Our measures of market access use the urban population of the other cells in each initial year as weights for the road changes in each period. Table A.10 shows that the overall IV effects of market access to cities are lower but remain strong if we simultaneously control for the effects of market access to total population (columns 1-4; using the total population of the other cells in each initial year as weights) and the effects of market access to mines (columns 5-8; using dummies for whether the other cells had a mine opened in each initial year as weights). In each case, instruments for the new variables are constructed analogously, and in the case of total population, this makes the overall instrument set very weak. There is no evidence of an independent effect of either total population or mines. The total population data are derived primarily from province-level data and are therefore likely more mismeasured than the urban population data. Ideally we would consider rural population separately, but since total population is mapped independently of cities, its measurement error would be even larger, and nonclassical. Similarly, since the mines measure is binary, independent of production level, it is likely to be severely mismeasured.

Night Lights. Table A.11 shows results for 3,591 cell-years for which we have data on both population and night lights (the 1990s, with lights starting in 1992, and 2000s), excluding cities in gas flaring regions. In Panel A, the outcome is log change in population. Effects are similar to those for the full sample (Table 1). In Panel B, the outcome is the log change in per capita night lights, with zero values replaced by the smallest nonzero sample value. Results are similar to those for total lights (Table 3). In panel C, the effect on population controlling for contemporaneous night lights change are also substantially positive but limited to the two lags.

Reallocation. Table A.12 repeats the tests of rows 5–8 of Table 4 for reallocation within 3x3, 5x5, 7x7 and 9x9 *mega-cells*, but restricts the sample to mega-cells that do not contain the capital or any of the 5 largest cities or regional capitals of each country in 1960. These mega-cells are the least likely to be destinations of long distance migration, especially if there are ethnic differences across mega-cells. Results are again noisy but similar to Table 4.

A.12 Robustness Checks: Heterogeneous Effects

Tables A.13–A.16 report robustness checks on the heterogeneous specifications.

Remoteness. Row 1 of Table 5 showed larger effects for more remote cities defined in terms of distance to the country's top (capital, largest and second largest) cities in 1960. Rows 1–5 of Table A.13 shows results are generally similar when using alternative definitions of remoteness, though differences are not always significant in the IV5 specification. Row 1 compares cities above and below 1960 median market access in the country, and row 2 those initially (in $t - 30$) bigger and smaller than their country's

median city. Rows 3–5 vary the Table 5 row 1 specification in other ways: (i) expanding attention to top cities (capital, largest, 2nd largest) in both 1960 and 2010 (row 3); (ii) using the continental instead of country-specific median distance (row 4); and (iii) dropping the top cities themselves (row 5). Alternatively, rows 6–10 show mostly positive effects of being less connected in terms of transportation infrastructure: (i) If the cell has no paved/improved road in 1960 (row 6) or no railroad in 1960 (row 7); (ii) If the cell is farther away than the median Euclidean distance in the country to a 1960 port (row 8), a 2005 port (row 9), or a 2007 airport (row 10). The results of row 1 on market access are broadly similar if we define remote cities based on 1960 market access smaller than the national 25th percentile (row 11) or 75th percentile (row 12) or the continental median (row 13). Results are in the same direction, but instruments, especially in row 12, are substantially weaker. The results of row 2 on population size are broadly similar if we define smaller cities based on a $t - 30$ population smaller than: the national 25th percentile (row 14) or 75th percentile (row 15); the continental median (row 16); or the national median in a sample dropping the top cities in 1960 and 2010 (row 17).

Land Suitability. Rows 1–8 of Table A.14 shows variants of Table 5 row 2. Cities in areas where land suitability is over 75% grow relatively slower when market access increases (row 1). The significance of the differences (for IV5 and IV15) are striking given that the high suitability group represents only 5% of the sample and its coefficients are imprecisely estimated as a result. Limiting attention to food (rows 2–3) or cash (rows 4–5) crops only gives similar results when splitting suitability at 25%, but somewhat smaller effects (and in the case of cash crops, substantially weaker instruments) when splitting at 75%. Redefining suitability more narrowly in terms of average (1960–2010) rainfall follows a similar pattern when splitting at the 25th percentile (row 7); when splitting at the 75th percentile IV differences are small and imprecisely estimated. In row 8, we do not find any significant differential effect based on distance to a mine using our IV strategy, though OLS estimates imply smaller effects near mines. Land-labor ratios differ across sectors, so we should not expect roads to have the same population effects for all economic activities.

Leader Favoritism. Rows 9–17 of Table A.14 report perturbations of the leader's origin results of Table 5, row 3. In Table 5, we interact each market access change variable (between $t - 30$ and $t - 20$, between $t - 20$ and $t - 10$, and between $t - 10$ and t) with a dummy equal to one if the leader was in power for more years than the mean duration (2 years) in the specific decade among the sample of 4,725 observations (thus, also between $t - 30$ and $t - 20$, between $t - 20$ and $t - 10$, and between $t - 10$ and t). Row 9 of Table A.14 splits the sample at 9 years of rule in the decade (90th percentile) instead of 2. Rows 10 and 11 limit the dummy to non-democratic leaders (Polity IV score under 5), using the mean and 90th percentile non-democratic durations (1 and 7 years). Rows 12 and 13 assign leaders to cities based on a radius of 250 instead of 150 km, using the mean and 90th percentile tenure durations (3 and 10 years). Rows 14 and 15 use ethnic boundaries from Murdock (1959) instead of circles around place of birth to define ethnic homelands (mean 2 and 90th percentile 7 years). Overall, leaders' ethnic

territories consistently see lower point estimates, though many are now insignificant. Point estimates of differentials are generally larger for long-serving and non-democratic leaders.

Conversely, and unlike large cities in general, 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. This differential is only significantly different from zero when considering 2010 regional capitals (row 17), whose status could have been jointly determined along with road locations, not 1960 regional capitals (rows 16), but the differential sign in these two variants is consistent across all IV specifications. Overall, this suggests that roads built for different kinds of “political” reasons may have different effects.

First stage heterogeneity and local average treatment effects (LATEs). Table A.15 reports first stage regression results where the market access measures are interacted with dummies for farther than the median distance from the country’s top cities in 1960 (panel A), lower than national median market access in 1960 (panel B), and population below the national median in $t - 30$ (panel C). See main text for discussion.

Domestic/Foreign/Overland/Overseas Market Access. Table A.16 reports variants of Table 6. Row 1 of weights city populations in the foreign/domestic specification by their country’s GDP per capita, to account for the possibility that connections to wealthier cities/countries have larger effects than those to poorer ones, with minimal impact. Rows 2–5 show little change in overland/overseas results when: (i) excluding cities within 50 km of a 2005 port (row 2); (ii) including overseas cities in the overland market access (row 3); (iii) fix port population to its 1960 level when calculating overseas market access (row 4); and (iv) defining overseas cities based on 1960 instead of 2005 ports (row 5).

A.13 Aggregate Urban Effects of Road Upgrading 1960-2010

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

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

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

where $\Delta_R \ln MA_{ot}$, as defined in the main text, includes only changes in roads, not population, between $t - 10$ and t , and the $\hat{\beta}$ terms are lag-specific estimated effects of changes in market access from our baseline specifications. $\Delta_R \ln MA_{ot-10}$ and $\Delta_R \ln MA_{ot-20}$ are assumed to be zero for $t = 1970$ and $t = \{1970, 1980\}$, respectively, in the absence of data about road-building in the 1940s and 1950s.¹⁰

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

Then counterfactual log population in year t (i.e. in the absence of changes between $t - 10$ and t caused by roads built between $t - 30$ and t) is defined as:

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

Thus, absolute population change due to those roads is: $\tilde{\Delta}P_{ot} = P_{ot} - \tilde{P}_{ot}$.

We restrict attention to 5,903 cell-decades with intensive margin growth (i.e. urban population above 10,000 in both $t-10$ and t) and non-missing market access. This comprises our main estimation sample plus the 1960s and 1970s. Each of these cell-decades is affected by up to 3 decades of change in market access, but for those in the 1960s and 1970s, we only observe those that took place after 1960.

Total urban population of the 39 sample countries increased by 203.5 million between 1960 and 2010, of which 171 million represent intensive margin growth. The first three columns of Table A.17 (column (1)) report the share of urban population growth coming from intensive margin growth that could be explained by road upgrades in 1960-2010, based on the 3 instruments. In row 1, this is based on our main average effects specification. Summing across all cities and decades, our estimated contribution of road building to city growth 1960–2010:

$$\sum_t \sum_o \tilde{\Delta}P_{ot} = 5.1 \text{ to } 10.5 \text{ million}$$

depending on whether we use the average IV5 or IV15 estimates. Between 1960 and 2010, the total urban population of the 39 countries increased by 203.5 million, of which 171 million reflected the intensive margin growth we study. The 5.1–10.5 million new urban residents thus represent 3.0–6.2% of intensive margin growth.

The urbanization rate of the 39 countries increased from 9.2% in 1960 to 27.9% in 2010, but excluding extensive margin growth (new cities) it would have been 23.8% in 2010. This implies a 14.6 percentage point increase between 1960 and 2010. If the “extra” urban residents caused by road upgrades had stayed in rural areas, the 39 countries’ overall urbanization rate would be 0.6–1.3 percentage points lower than its actual rate of 23.8% in 2010 (see column (2) of Table A.17). Column (3) reports these estimates as a fraction of the 14.6 percentage point increase actually experienced. The average effects suggest that 4.1–8.9% of the intensive margin growth in urban share between 1960 and 2010 was due to the road upgrades.

Depending on the heterogeneity variable considered, we find contributions that are between 35% lower and 25% higher. More precisely, allowing for heterogeneity in our coarse way widens the range of effects to 3.1–13.1 million new urban residents, accounting for 1.8–7.7% of intensive margin growth in urban population or 0.4–1.6 percentage points of the urbanization rate and 2.7–11.0% of intensive

margin increase in the urban share.¹¹ We regard this modest widening of the range of potential effects as mildly surprisingly. However, it is likely limited by the binary form of heterogeneity we consider.

These estimates are conservative in the narrow sense that they apply changes to individual decades, rather than compounding them, and because they do not include the contribution of roads to extensive margin urban growth. In addition, they assume no reallocation, which would reduce the estimated aggregate effects, or other general equilibrium effects, which could increase or decrease them. While we found little evidence of reallocation in Section 3.5., we certainly cannot rule it out. Thus, these aggregate results should be taken with caution.

Trans-African Highways 2010–2040. Another way to interpret these results is in the context of proposed roads. The idea of a Trans-African Highway (TAH) network has been discussed since at least the early 1970s and was operationalized in a proposal 30 years later by the United Nations Economic Commission for Africa and the African Development Bank (ADB and UNECA, 2003). Using a map of the TAH network we constructed from this document (Appendix Figure A.9), we find that its complete implementation would require construction of 44,000 km of highways in sub-Saharan Africa, 42,000 km of which are in the 39 sample countries. In our data, there are only 1,490 km of highways in the 39 countries in 2010. By comparison, India had 24,000 km (Government of India, 2016) and China 111,900 km (Government of China, 2016). The TAH network would thus represent a 2,740% increase in highway length.¹²

Assuming travel speeds of 80 (or alternatively 100) kph along the TAH roads, we estimate by how much the market access of each city in 2010 would have increased (due to roads only) had the TAH roads been built by 2010.¹³ We then use the same methodology as described above to estimate the potential aggregate effects of the TAH network between 2010 and 2040.¹⁴

More precisely, we focus on the 2,768 cells for which urban population was above 10,000 in 2010 applying average 30-year effect estimates from row 6 of Table A.9 and analogous heterogeneous effects estimates (available upon request) to the market access changes we assume are associated with TAH construction.¹⁵

Our average effects reported in Table row 6 of Table A.9 imply that the urban population of the 39 sample countries would increase by 2.1–7.6 million, a range that overlaps with the 1960–2010

¹¹We now define initially small cities as those below the country median population in 1960 rather than t-30, because this exercise includes the 1960s and 1970s. Otherwise, the groups for the heterogeneous effect exercise are defined using the same thresholds and dummies as in our main analysis on the sample of 4,725 observations.

¹²Using construction cost data from Collier et al. (2015), we estimate that the cost of building the TAH network is 12–15% of 2010 regional GDP (vs. 17% for 1960–2010 road upgrades).

¹³80 kph is the baseline highway speed in the rest of our analysis. 100 kph reflects the possibility that the TAH will be built to a higher standard than existing highways. Note that we include TAH segments in South Africa, Lesotho and Swaziland in calculating market access.

¹⁴We use the 30-year change in market access specification for this exercise because city populations in the intermediate years 2020 and 2030 are unknown.

¹⁵We define the favoritism dummy using only leaders from the 2000s, since this is when the TAH plan was designed. Otherwise, the groups for the heterogeneous effect exercise are defined using the same thresholds and dummies as before.

effect estimates on the low end. Since the urban population of the 39 countries was 222.3 million in 2010, this is a 0.9–3.4% increase, a small fraction of the 206% overall increase (intensive and extensive margin) predicted by United Nations (2015b) based on country-specific urban definitions.¹⁶ Allowing for heterogeneity expands these ranges to 1.1–9.8 million new urban residents, representing a 0.5–4.4% increase in the urban population. Thus, depending on the heterogeneity variable considered, effects are between 50% lower and 30% higher. United Nations (2015b) reports an urbanization rate of 33.6% for the sample regions in 2010, and a projected increase to 49.0% by 2040. Adding our estimated TAH effect to the United Nations (2015b) projection implies a 2040 urbanization rate 0.3–1.0 percentage points higher when using the average effects (details available upon request). Heterogeneous effect estimates imply values between half lower and one third higher. Again, these results do not account for urban reallocation, which could be more important given a higher initial urbanization rate, other general equilibrium effects, and extensive margin growth. They should thus be taken with caution.

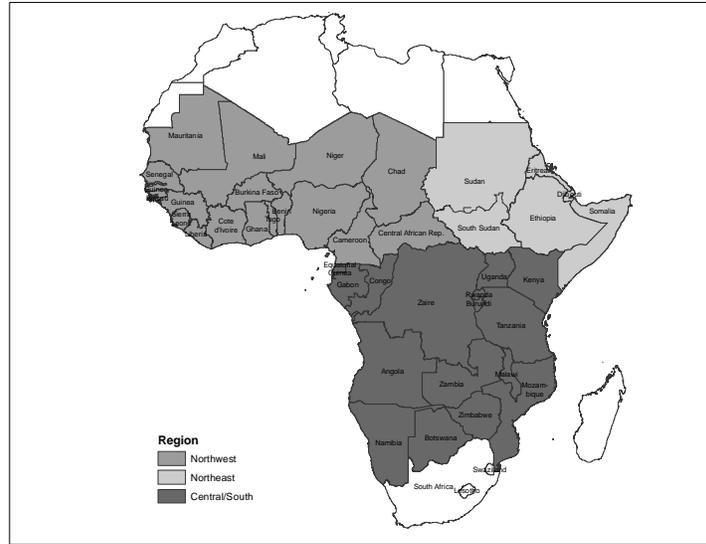
REFERENCES

- ADB and UNECA, “Review of the Implementation Status of the Trans African Highways and the Missing Links,” Technical Report, African Development Bank and United Nations Economic Commission For Africa 2003.
- Ady, Peter H, *Oxford Regional Economic Atlas of Africa*, Oxford, UK: Oxford University Press, 1965.
- Alder, Simon, “Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development,” mimeo, University of North Carolina 2017.
- Bolt, Jutta and Jan Luiten van Zanden, “The Maddison Project: collaborative research on historical national accounts,” *Economic History Review*, 08 2014, 67 (3), 627–651.
- Burgess, Robin, Rémi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel, “The Value of Democracy: Evidence from Road Building in Kenya,” *American Economic Review*, June 2015, 105 (6), 1817–51.
- Canning, David, “A Database of World Stocks of Infrastructure: 1950-1995,” *World Bank Economic Review*, 1998, 12, 529–548.
- Center for Systemic Peace, *Forcibly Displaced Populations, 1964-2008*, Vienna, VA: Center for Systemic Peace, 2015.
- , *Major Episodes of Political Violence, 1946-2016*, Vienna, VA: Center for Systemic Peace, 2015.
- CIESIN, “Gridded Population of the World, Version 4 (GPWv4): Population Count,” Technical Report, Center for International Earth Science Information Network - CIESIN - Columbia University 2016.
- Collier, Paul, Martina Kirchberger, and Måns Söderbom, “The Cost of Road Infrastructure in Low and Middle Income Countries,” *The World Bank Economic Review*, 2015, 30 (3), 522–548.
- CRED, *EM-DAT: International Disaster Database*, Louvain, Belgium: Centre for Research on the Epidemiology of Disasters (CRED), Université Catholique de Louvain, 2013.
- Donaldson, Dave and Richard Hornbeck, “Railroads and American Economic Growth: A ‘Market Access’ Approach,” *Quarterly Journal of Economics*, 2016, 131 (2), 799–858.
- Eros, J.M. and Luisette Candelario-Quintana, *Mineral Operations of Africa and the Middle East: Derived from USGS Open-File report 2006-1135*, Reston, VA: U.S. Geological Survey, 2006.
- 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).
- Fearon, James D., Kimuli Kasara, and David D. Laitin, “Ethnic Minority Rule and Civil War Onset,” *American Political Science Review*, February 2007, 101, 187–193.
- Francois, Patrick, Ilia Rainer, and Francesco Trebbi, “How Is Power Shared in Africa?,” *Econometrica*, 2015, 83 (2), 465–503.
- Government of China, *Transportation Overview*, Beijing, China: Ministry of Transport of the People’s Republic, 2016.
- Government of India, *Basic Road Statistics of India 2013-14 and 2014-15*, Delhi, India: Ministry of Road Transport & Highways, 2016.

¹⁶These definitions are on average less restrictive than ours.

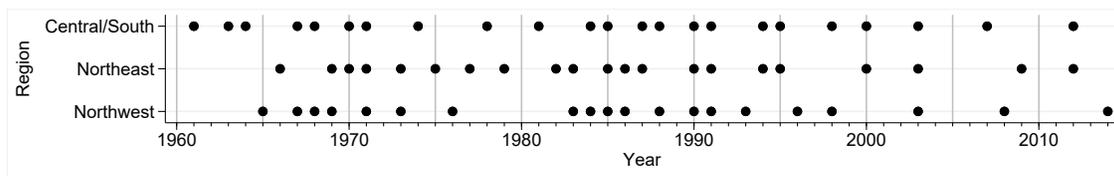
- Hodler, Roland and Paul A. Raschky, "Regional Favoritism," *The Quarterly Journal of Economics*, 2014, 129 (2), 995–1033.
- IIASA and FAO, "Global Agro-ecological Zones (GAEZ v3.0)," Technical Report, IIASA and FAO 2012.
- International Monetary Fund, "World Economic Outlook: Global Manufacturing Downturn, Rising Trade Barriers," Technical Report October 2019.
- Jedwab, Rémi and Alexander Moradi, "The Permanent Economic Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics & Statistics*, 2016, 98 (2), 268–284.
- Law, Gwillim, "Administrative Divisions of Countries ("Statoids")," <http://www.statoids.com/>. Accessed: 2015–2016.
- Maddison, Angus, *Maddison Project Database* 2008.
- Murdock, George, *Africa: Its Peoples and their Culture.*, New York: McGraw-Hill, 1959.
- National Imagery and Mapping Agency, "VMAP_1V10 - Vector Map Level 0 (Digital Chart of the World)."
- Nelson, Andrew and Uwe Deichmann, "African Population database documentation," Technical Report, United Nations Environment Programme and CIESIN 2004.
- Ocean Shipping Consultants, Ltd., "Beyond the Bottlenecks: Ports in Africa," African Infrastructure Country Diagnostic Background Paper 8, Washington, DC 2009.
- Polity IV, *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013*, Vienna, VA: Center for Systemic Peace, 2015.
- Shiferaw, Admasu, Mans Soderbom, Eversusalem Siba, and Getnet Alemu, "Road Infrastructure and Enterprise Dynamics in Ethiopia," Unpublished 2013.
- Sundberg, Ralph and Erik Melander, "Introducing the UCDP Georeferenced Event Dataset," *Journal of Peace Research*, 2013, 50 (4), 523–532.
- Teravaninthorn, Supee and Gael Raballand, *Transport Prices and Costs in Africa: a Review of the International Corridors* 2008.
- United Nations, *World Population Prospects: The 2015 Revision*, New York, NY: United Nations, 2015.
- , *World Urbanization Prospects: The 2014 Revision*, New York: Department of Economic and Social Affairs, Population Division, 2015.
- Weidmann, Nils B., Jan Ketil Rød, and Lars-Erik Cederman, "Representing ethnic groups in space: A new dataset," *Journal of Peace Research*, 2010, 47 (4), 491–499.
- Wikipedia, "Administrative Divisions of [Country] X; List of Heads of State of [Country] X; Profile of Head of State Y," <https://www.wikipedia.org/>. Accessed: 2016.
- Willmott, Cort J. and Kenji Matsuura, "Terrestrial Precipitation: 1900-2008 Gridded Monthly Time Series (1900 - 2008) (V 2.01)," 2009.
- World Bank, *World Development Indicators*, Washington: World Bank, 2015.
- , *World Development Indicators*, Washington: World Bank, 2016.
- World Database on Protected Areas, *Protected Areas*, Cambridge, UK: United Nations Environment World Conservation Monitoring Centre, 2015.

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



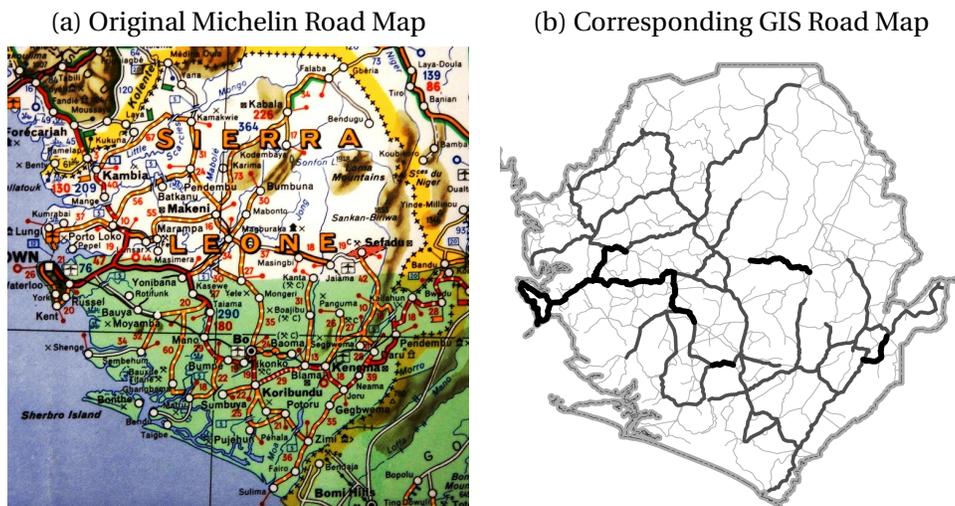
Notes: This figure shows the 39 countries of our main analysis, as well as South Africa, Lesotho and Swaziland, which contribute roads and their largest 20, 1, and 1 cities, respectively, to the calculation of market access. All analysis includes South Sudan in Sudan to reflect the situation during the sample period.

Figure A.2: Map Years for Each Map Region



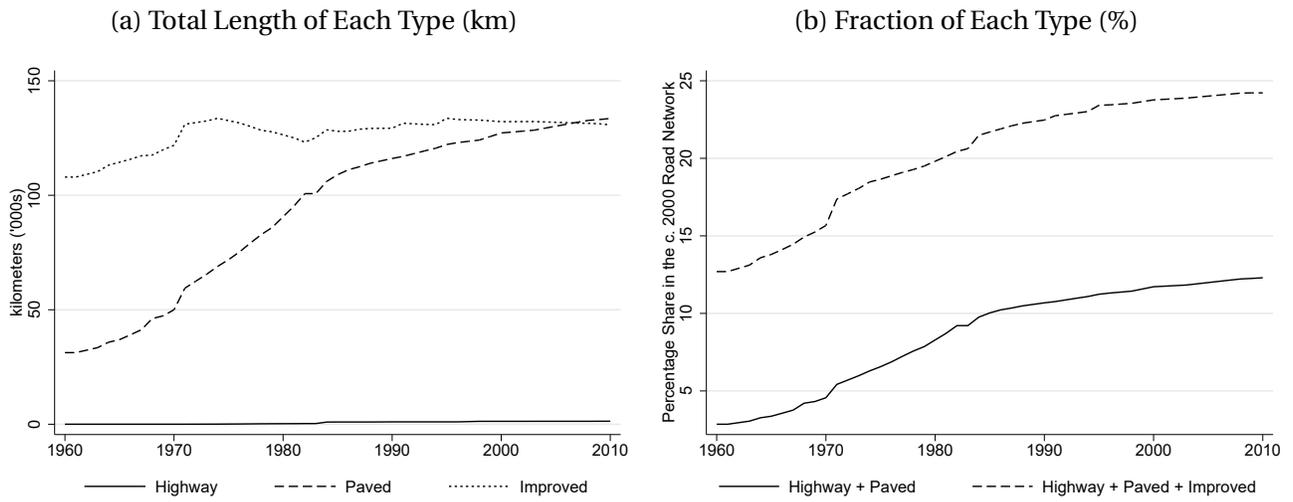
Notes: This figure shows the years for which we have a map for each of the three sections of the road maps for sub-Saharan Africa. There are 64 maps in total: 20 for the Northwest region, 21 for the Northeast, and 23 for the Central/South. The average gap between maps across regions is under 2.5 years, and the longest is 7 years.

Figure A.3: Michelin Road Map and Corresponding GIS Road Map of Sierra Leone in 1969



Notes: The left panel shows the section of the original Michelin map for Sierra Leone in 1969. Different colors and patterns correspond to different types of roads. The right panel shows our GIS map, with three aggregated road categories only: paved (thick black), improved (thick grey) and tracks (thin grey). As the tracks shown in the GIS map are mostly from Nelson and Deichmann (2004), they differ from the Michelin map in some places.

Figure A.4: Road network evolution by type in the 39-country sample, 1960–2010



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

Figure A.5: Steps to Obtain Market Access for Sierra Leone between 1970 and 1980

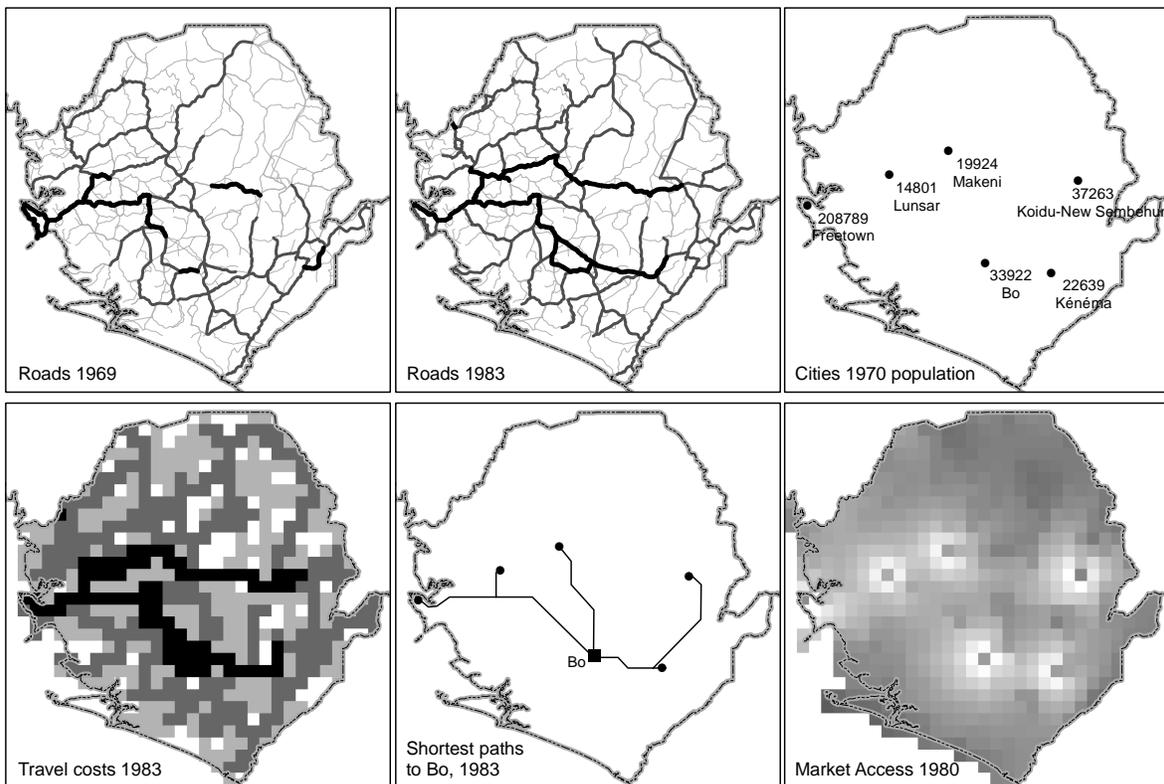
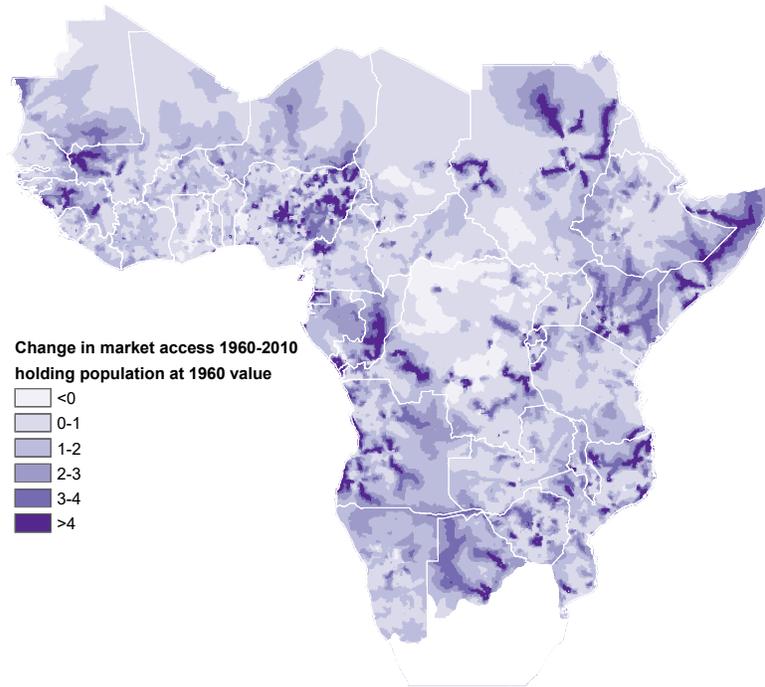
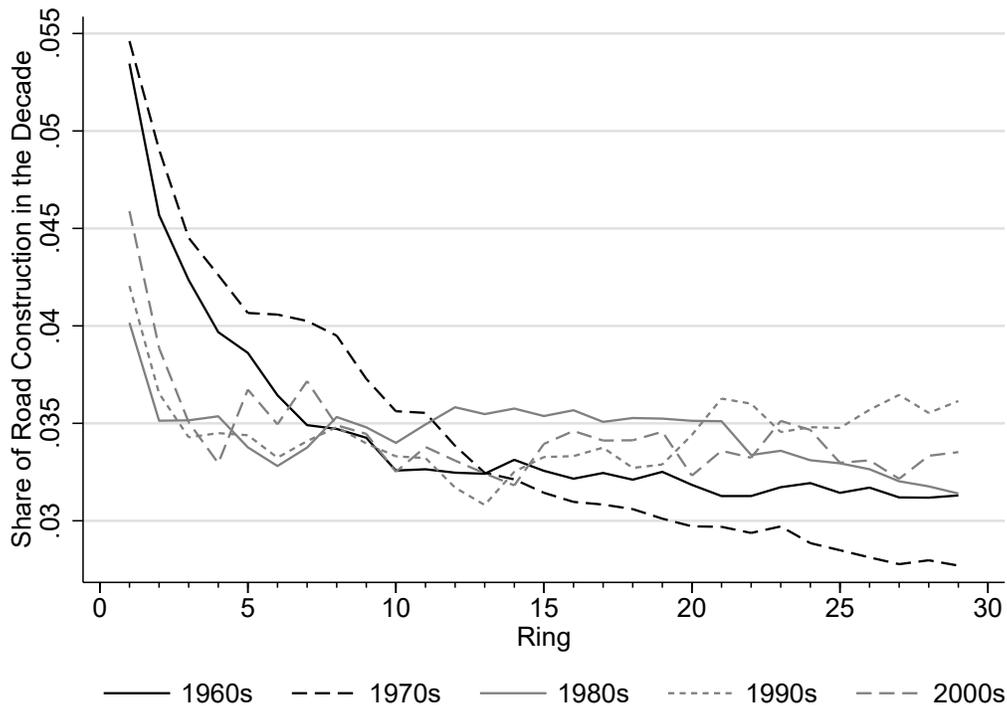


Figure A.6: Change in Market Access 1960–2010 for the 39 Sample Countries, 1960 population



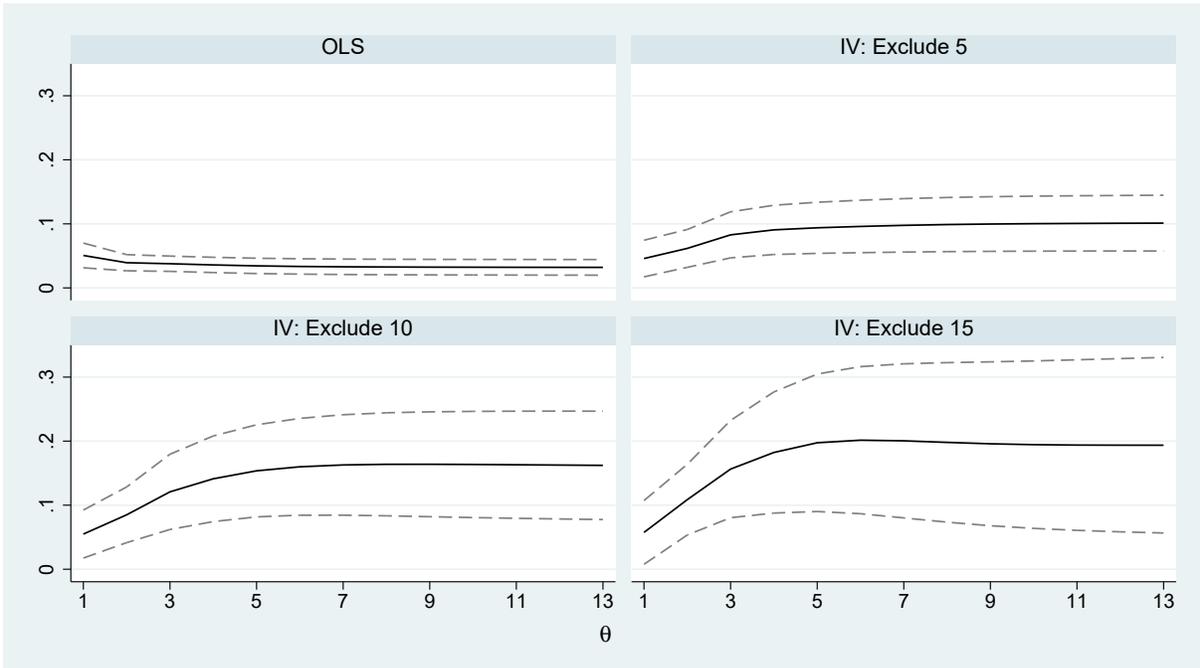
Notes: This figure shows the change in market access between 1960 and 2010, holding city populations fixed at their 1960 levels.

Figure A.7: Share of Road Construction in Inner vs. Outer Rings, 1960-2010



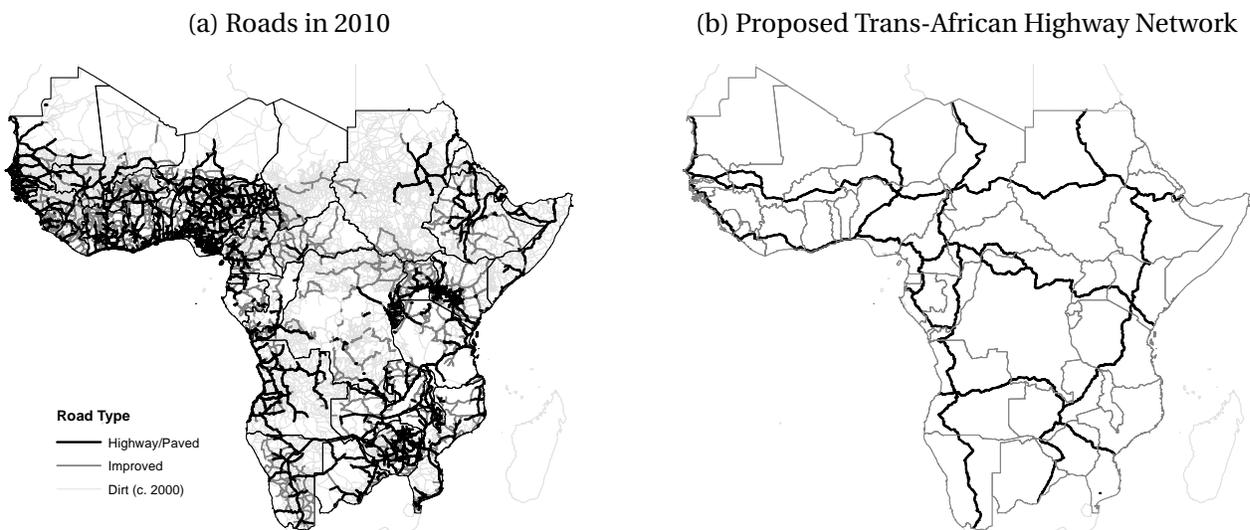
Notes: For each decade, averaging across all cities, this figure shows the share of total road construction taking place 0–1, 1–2, ..., 29–30 cells away from the city.

Figure A.8: Relationship between θ 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 $\Delta_{t-10}^t \ln$ urban population)/100 for values of $\theta = [1; 13]$, with 95% confidence interval, for each of the 5 main identification strategies.

Figure A.9: Roads in 2010 and Proposed Trans-African Highway Network



Notes: Left panel shows the highways and paved/improved/dirt roads in the 39 sub-Saharan African countries of our sample in 2010. Right panel shows the proposed Trans-African Highway network, from ADB and UNECA (2003).

Table A.2: Speeds (Km / Hour) Assumed in City-to-City Distance Calculations

Category	This Paper	India (Alder, 2017)		Ethiopia (Shiferaw et al., 2013)	
Highway	80	Golden Quadrilateral	75	Post-Rehab Asphalt	70
Paved Road	60	Conventional Highways	35	Pre-Rehab Asphalt	50
Improved Road	40	Roads of Lower Quality	25	Pre-Rehab Federal Gravel	35
Dirt Road	12	Unpaved/No Roads	10	Pre-Rehab Regional Gravel	25
No Road	6	Unpaved/No Roads	10	Pre-Rehab Earth	20

Table A.3: Descriptive Statistics for the Main Sample of City-Period Observations

Main Variable:	Mean	Std. Dev.	Min	Max
Δ_{t-10}^t ln Urban Population	0.318	0.209	-1.533	2.343
Δ_{t-10}^t ln MA (Market Access)	0.655	0.892	-8.236	10.618
Δ_{t-20}^{t-10} ln MA (Market Access)	0.901	1.099	-8.236	11.537
Δ_{t-30}^{t-20} ln MA (Market Access)	1.161	1.288	-8.236	13.291
ln Urban Population _{t-10}	10.247	0.990	9.210	15.902

Table A.4: First Stage Regressions

Dep.Var. Δ ln MA	(1) IV: Exclude 5			(2) IV: Exclude 10			(3) IV: Exclude 15		
	Δ_{t-10}^t	Δ_{t-20}^{t-10}	Δ_{t-30}^{t-20}	Δ_{t-10}^t	Δ_{t-20}^{t-10}	Δ_{t-30}^{t-20}	Δ_{t-10}^t	Δ_{t-20}^{t-10}	Δ_{t-30}^{t-20}
Δ_{t-10}^t ln MA Roads	1.78*** [0.10]	-0.36*** [0.09]	-0.18* [0.10]	2.18*** [0.16]	-0.84*** [0.22]	-0.48*** [0.17]	2.20*** [0.38]	-1.35*** [0.34]	-0.57** [0.28]
Δ_{t-20}^{t-10} ln MA Roads	0.07 [0.07]	1.64*** [0.10]	-0.37*** [0.05]	-0.00 [0.11]	1.87*** [0.25]	-0.73*** [0.16]	0.17 [0.22]	1.85*** [0.38]	-1.44*** [0.29]
Δ_{t-30}^{t-20} ln MA Roads	0.16*** [0.06]	0.09 [0.07]	1.63*** [0.09]	0.21** [0.09]	0.20** [0.10]	1.92*** [0.20]	0.34** [0.14]	0.54*** [0.17]	2.31*** [0.32]

Notes: Each set of three columns reports the 1st stage regressions corresponding to the three endogeneous variables in the corresponding column of Table 1, Panel B. Robust SEs, clustered by 1960 province, are in brackets. N=4,725

Table A.5: Robustness checks: IVs with a lead and with a third lag

	OLS	IV: Exclude 5	IV: Exclude 10	IV: Exclude 15
	(1)	(2)	(3)	(4)
<i>Panel A: Lead</i>				
$\Delta_{t-10}^t \ln$ Market Access	1.52*** [0.41]	3.32*** [1.01]	4.74*** [1.51]	4.90* [2.89]
$\Delta_{t-20}^{t-10} \ln$ Market Access	1.10*** [0.34]	2.91*** [0.93]	4.76*** [1.49]	4.27** [2.12]
$\Delta_{t-30}^{t-20} \ln$ Market Access	0.76** [0.30]	0.41 [0.74]	0.24 [1.03]	-0.07 [1.30]
$\Delta_{t+10}^t \ln$ Market Access	0.72 [0.56]	1.93 [1.42]	2.62 [2.81]	3.28 [4.60]
Overall Effect ($t - 30$ to t)	3.39*** [0.71]	6.65*** [2.02]	9.73*** [3.04]	9.11** [3.99]
First stage Kleibergen-Paap F		37.40	28.19	6.919
<i>Panel B: Three lags</i>				
$\Delta_{t-10}^t \ln$ Market Access	1.57*** [0.46]	0.56 [1.33]	1.40 [2.50]	2.26 [4.01]
$\Delta_{t-20}^{t-10} \ln$ Market Access	1.49*** [0.36]	4.51*** [1.03]	7.81*** [1.62]	8.81*** [2.92]
$\Delta_{t-30}^{t-20} \ln$ Market Access	0.80** [0.32]	3.15** [1.26]	4.79** [2.13]	7.32** [3.00]
$\Delta_{t-40}^{t-30} \ln$ Market Access	0.31 [0.25]	0.04 [0.76]	0.31 [1.15]	-1.35 [1.65]
Overall Effect ($t - 30$ to t)	3.87*** [0.79]	8.22*** [2.38]	14.00*** [3.79]	18.39*** [5.42]
Overall Effect ($t - 40$ to t)	4.18*** [0.83]	8.26*** [2.62]	14.32*** [3.98]	17.04*** [5.25]
First stage Kleibergen-Paap F		40.91	22.10	7.634

Notes: Each column in each panel is a separate regression of $100\Delta_{t-10}^t \ln$ urban population on the change in market access measures shown. Each regression includes the same controls as Table 1. In columns 2–4 measures of $\Delta \ln$ Market Access that exclude road changes within the radius shown instrument for the market access change measures. “Overall Effect” is the sum of the contemporaneous effect and lags specified. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance. In Panel A, t indexes years 1990 and 2000, for 2,607 cell-years. In Panel B, t indexes years 2000 and 2010, for 3,630 cell-years.

Table A.6: Robustness Checks: IV Strategies and Exclusion Restriction

	OLS	IV5	IV10	IV15
(1) Co-Investment: Inner: 2-3, Outer: 5-6 (N=1,890; F: -; 19.1; 8.7; 2.0)	4.02*** [1.23]	10.63** [4.55]		
(2) Co-Investment: Inner: 2-3, Outer: 10-11 (N=2,197; F: -; 38.4; 16.7; 3.0)	4.28*** [0.99]	10.35*** [3.62]	13.75** [5.69]	
(3) Radial Ext. Out.: Inner: 2-3, Outer: 5-6 (N=2,098; F: -; 79.2; 9.5; 6.6)	4.11*** [0.95]	9.10*** [2.50]		
(4) Radial Ext. Out.: Inner: 2-3, Outer: 10-11 (N=1,804; F: -; 71.1; 24.2; 5.6)	3.51*** [0.94]	8.54*** [2.19]	12.47*** [3.93]	
(5) Radial Ext. In.: Inner: 2-3, Outer: 5-6 (N=2,502; F: -; 93.4; 41.6; 9.9)	3.38*** [0.88]	8.41*** [2.36]		
(6) Radial Ext. In.: Inner: 2-3, Outer: 10-11 (N=2,672; F: -; 91.9; 49.4; 11.6)	3.64*** [0.90]	8.30*** [2.36]	11.60*** [3.44]	
(7) Co-Investment: Octant: Inner: 1-2, Outer: 5-6 (N=2,533; F: -; 47.7; 21.7; 2.2)	3.90*** [0.97]	12.01*** [3.90]		
(8) Co-Investment: Octant: Inner: 1-2, Outer: 10-11 (N=2,654; F: -; 60.1; 26.8; 3.2)	4.89*** [0.90]	12.21*** [3.28]	18.75*** [5.80]	
(9) Co-Investment: Octant: Inner: 1-2, Outer: 15-16 (N=2,686; F: -; 78.4; 29.7; 7.3)	4.18*** [0.86]	10.22*** [2.70]	13.34*** [4.37]	16.81** [6.92]
(10) Rad. Ext. Out.: Octant: Inner: 1-2, Outer: 5-6 (N=1,971; F: -; 71.0; 8.9; 6.3)	4.28*** [0.97]	8.57*** [2.59]		
(11) Rad. Ext. Out.: Octant: Inner: 1-2, Outer: 10-11 (N=1,546; F: -; 88.0; 7.4; 6.7)	3.84*** [0.97]	8.99*** [2.55]	15.15*** [4.30]	
(12) Rad. Ext. Out.: Octant: Inner: 1-2, Outer: 15-16 (N=1,315; F: -; 43.5; 5.2; 2.9)	4.88*** [1.04]	11.70*** [2.84]	18.77*** [5.76]	28.17** [11.83]
(13) Rad. Ext. In.: Octant: Inner: 1-2, Outer: 5-6 (N=2,961; F: -; 100.4; 20.2; 9.1)	3.76*** [0.83]	9.57*** [2.49]		
(14) Rad. Ext. In.: Octant: Inner: 1-2, Outer: 10-11 (N=3,014; F: -; 79.6; 46.3; 11.8)	3.90*** [0.82]	8.80*** [2.38]	12.41*** [3.56]	
(15) Rad. Ext. In.: Octant: Inner: 1-2, Outer: 15-16 (N=3,190; F: -; 121.0; 34.6; 11.4)	3.14*** [0.81]	8.95*** [2.45]	12.33*** [3.69]	15.16*** [5.33]
(16) Excl. Changes Convex Hull 50k (N=4,725; F: -; 63.6; 43.3; 13.9)	3.47*** [0.63]	7.93*** [2.17]	11.35*** [3.34]	16.08*** [4.59]
(17) Fixing Pop. to 1960 in IVs & Drop 1980s (N=3,629; F: -; 15.3; 7.0; 5.1)	3.84*** [0.79]	6.47*** [2.16]	10.84*** [3.25]	16.28*** [5.58]
(18) Fixing Pop. to 1960 in MAs & Drop 1980s (N=3,629; F: -; 156.7; 36.3; 21.3)	2.80* [1.53]	8.55*** [2.79]	15.91*** [4.65]	26.35*** [7.58]
(19) Fix. Pop. at t-10 in Market Access (MA) (N=4,725; F: -; 84.6; 20.8; 9.3)	3.36*** [1.06]	9.95*** [2.68]	14.91*** [3.84]	19.60*** [5.85]
(20) Ctrl Reg. Growth (Prov., Ethnic, Circle 0-15&15-30) (N=3,479; F: -; 31.2; 8.8; 4.1)	2.41*** [0.68]	6.19*** [2.25]	8.41** [3.33]	9.47** [4.75]

Notes: See text for details. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table A.7: Other City Exclusions and Controls

	OLS	IV5	IV10	IV15
(1) Excl. > 75th Pctile Dist. to Nearest City (N=3,543; F: -; 24.5; 5.5; 2.1)	2.84*** [0.68]	7.87*** [2.82]	11.05*** [2.75]	9.87*** [2.68]
(2) Excl. > 75th Avg. Dist. between Cities (N=3,551; F: -; 20.5; 7.9; 4.8)	3.06*** [0.70]	5.65** [2.27]	9.67*** [3.52]	12.62*** [4.29]
(3) Excl. National, Reg'l (Incl. 2010) & Top 5 Cities (N=3,460; F: -; 36.0; 6.0; 4.7)	3.41*** [0.74]	5.94** [2.41]	8.22* [4.81]	7.86 [6.38]
(4) More than 100 km from National Cities (N=3,391; F: -; 101.4; 12.9; 9.4)	3.65*** [0.73]	9.78*** [2.58]	16.94*** [4.29]	22.89*** [6.39]
(5) More than 100 km from Any Mine (N=3,202; F: -; 100.7; 10.4; 6.2)	3.68*** [0.74]	9.38*** [2.38]	15.18*** [4.13]	21.03*** [6.57]
(6) More than 100 km from Cash Crop Cells (N=4,606; F: -; 110.3; 46.8; 11.1)	3.65*** [0.65]	8.65*** [2.14]	12.93*** [3.21]	16.49*** [4.54]
(7) More than 100 Km from President's Origin (N=3,060; F: -; 103.7; 14.8; 6.2)	3.86*** [0.81]	9.86*** [2.69]	16.13*** [4.61]	22.11*** [7.65]
(8) More than 100 km from Port (N=4,317; F: -; 93.6; 33.5; 12.3)	3.42*** [0.65]	8.75*** [2.22]	13.81*** [3.32]	17.43*** [4.77]
(9) More than 100 km from Airport (N=3,383; F: -; 104.6; 55.6; 19.4)	3.24*** [0.77]	10.16*** [2.37]	14.27*** [3.57]	17.94*** [5.18]
(10) More than 100 km from Border Crossing (N=2,355; F: -; 52.6; 12.1; 7.6)	3.46*** [0.97]	7.92*** [2.55]	11.52*** [3.75]	15.60*** [5.88]
(11) More than 100 km from Natural Park Cell (N=2,888; F: -; 107.4; 19.1; 15.1)	3.91*** [0.88]	9.41*** [2.50]	13.76*** [3.92]	16.43*** [5.66]
(12) Drop Top/Bottom 1% Δ Population (N=4,631; F: -; 98.9; 22.4; 12.3)	2.73*** [0.57]	4.96** [1.92]	8.08*** [3.06]	9.61** [4.20]
(13) Drop Top/Bottom 1% Δ MA in Any Decade (N=4,456; F: -; 85.3; 36.1; 13.2)	4.21*** [0.74]	7.89*** [2.48]	12.14*** [3.47]	15.78*** [4.48]
(14) In Urban Pop t-10 x Country-Year FEs (N=4,725; F: -; 95.8; 29.3; 13.7)	3.47*** [0.62]	7.97*** [2.12]	12.81*** [3.22]	16.43*** [4.71]
(15) Control for In Market Access 1960 (N=4,723; F: -; 40.2; 4.9; 1.6)	3.11*** [0.68]	10.69*** [3.27]	26.18*** [9.34]	62.55** [31.31]
(16) City-Level Controls (N=4,725; F: -; 100.7; 19.3; 10.2)	2.90*** [0.54]	7.76*** [2.01]	12.56*** [3.46]	17.53*** [5.48]
(17) Control for Own Cell Surface (N=4,725; F: -; 86.7; 38.0; 9.6)	3.36*** [0.66]	8.90*** [2.40]	14.13*** [3.58]	18.25*** [5.10]
(18) Incl. Local Conflict as Control (N=3,631; F: -; 53.5; 29.6; 9.9)	3.75*** [0.76]	8.26*** [2.25]	13.54*** [3.76]	19.55*** [6.12]
(19) Ctrl for Neighboring Country's Growth (N=4,725; F: -; 100.1; 41.8; 12.0)	3.47*** [0.63]	8.39*** [2.17]	12.95*** [3.23]	16.53*** [4.60]

Notes: See text for details. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table A.8: Robustness Checks: Country-Level Shocks and Data Quality

	OLS	IV5	IV10	IV15
(1) Excl. 6 Countries with Independence Yr > 1970 (N=4,367; F: -, 96.0; 12.8; 8.9)	3.64*** [0.67]	8.68*** [2.27]	13.71*** [3.40]	17.78*** [4.81]
(2) Excl. 15 Countries with Independence Yr > 1960 (N=3,892; F: -, 103.9; 11.9; 7.8)	4.12*** [0.70]	8.37*** [2.39]	13.10*** [3.79]	17.54*** [5.67]
(3) Excl. Country-Decades with International War (N=4,458; F: -, 95.2; 57.7; 10.2)	3.56*** [0.67]	8.47*** [2.28]	13.01*** [3.32]	16.51*** [4.77]
(4) Excl. Country-Decades with Civil War (N=2,771; F: -, 16.8; 6.6; 2.8)	3.18*** [0.68]	7.01*** [2.37]	10.42*** [3.56]	10.25** [4.33]
(5) Excl. Country-Decades with Any War (N=2,741; F: -, 17.3; 6.3; 2.8)	3.21*** [0.71]	6.96*** [2.36]	10.65*** [3.40]	10.49** [4.08]
(6) Mean # Refugees in Decade < Mean in SSA (N=3,489; F: -, 39.7; 36.3; 7.6)	3.52*** [0.65]	7.55*** [2.07]	10.86*** [2.67]	11.89*** [3.60]
(7) Max # Refugees in Decade < Mean in SSA (N=3,344; F: -, 39.7; 36.3; 6.9)	3.61*** [0.65]	7.98*** [2.11]	11.17*** [2.79]	12.02*** [3.96]
(8) Excl. Country-Decades with Multi-year Drought (N=4,669; F: -, 84.3; 34.6; 13.2)	3.42*** [0.62]	7.44*** [2.03]	11.81*** [3.26]	15.54*** [4.69]
(9) Balanced Sample (Always $\geq 10,000$) (N=3,264; F: -, 89.6; 43.6; 9.3)	3.37*** [0.68]	7.48*** [1.98]	10.81*** [2.97]	12.74*** [4.46]
(10) Population Threshold = 20,000 (N=2,471; F: -, 85.6; 37.4; 8.7)	2.69*** [0.80]	7.68*** [2.87]	11.96*** [4.41]	14.79** [6.90]
(11) Pop. Estimate < 10,000 Available (N=7,369; F: -, 62.5; 22.7; 11.7)	5.10*** [1.01]	7.45*** [2.34]	9.28*** [2.79]	10.58*** [3.45]
(12) Pop. Est. < 10,000 Avail. One Prev. Year (N=6,164; F: -, 69.7; 18.4; 11.4)	4.01*** [0.80]	6.33*** [2.21]	9.86*** [3.00]	11.65*** [3.90]
(13) Drop 1980s (N=3,631; F: -, 53.2; 29.5; 9.9)	3.84*** [0.79]	8.18*** [2.34]	13.70*** [3.86]	19.82*** [6.21]
(14) Drop 2000s (N=2,607; F: -, 42.7; 35.8; 9.0)	3.35*** [0.72]	7.03*** [2.11]	10.41*** [3.17]	10.51*** [3.99]
(15) 2 Censuses for Both Start & End Year (N=3,414; F: -, 30.7; 14.1; 5.7)	3.14*** [0.65]	10.26*** [2.64]	14.33*** [3.48]	15.77*** [5.29]
(16) Excl. if Both Start & End ≥ 5 Yrs from Source (N=4,430; F: -, 98.5; 24.0; 10.6)	3.46*** [0.63]	8.36*** [2.13]	13.05*** [3.21]	16.29*** [4.59]
(17) Excl. if Start or End ≥ 5 Yrs from Source (N=1,711; F: -, 43.7; 8.4; 7.8)	1.32 [0.91]	6.22** [2.65]	8.01** [3.98]	6.39 [5.20]
(18) Drop non-Africapolis Countries (N=4,241; F=102.4; 12.9; 8.0)	3.67*** [0.67]	8.70*** [2.33]	13.54*** [3.57]	17.56*** [5.14]
(19) Excl. if > Mean Diff. Official Road Data (N=2,246; F: -, 147.8; 10.0; 4.3)	3.89*** [0.88]	10.81*** [2.77]	17.30*** [4.84]	22.63*** [7.33]
(20) Excl. if Country > Mean Diff. Official (N=3,659; F: -, 82.5; 27.3; 9.2)	3.95*** [0.77]	9.48*** [2.66]	15.58*** [4.13]	20.28*** [6.07]
(21) Excl. if > Mean Diff. Official or Missing (N=1,032; F: -, 23.9; 1.4; 1.3)	1.43 [1.09]	6.84 [4.73]	20.65 [12.91]	23.19 [16.14]
(22) Excl. if Country > Mean Diff. or Missing (N=3,258; F: -, 34.5; 28.0; 7.6)	4.33*** [0.77]	10.84*** [2.81]	16.81*** [3.75]	22.02*** [5.50]

Notes: See text for details. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table A.9: Robustness Checks: Functional Form

	OLS	IV5	IV10	IV15
(1) No controls (N=4,725; F: .; 92.3; 75.7; 21.5)	5.36*** [0.66]	7.43*** [2.08]	9.69*** [2.85]	10.27*** [3.46]
(2) Controls: Country-Year FEs Only (N=4,725; F: .; 86.3; 45.8; 15.9)	3.42*** [0.60]	7.26*** [2.09]	10.69*** [2.92]	12.84*** [3.92]
(3) Controls: Country-Year FEs & Lagged Pop. (F: .; 86.55; 46.59; 15.94)	3.37*** [0.61]	7.29*** [2.06]	10.71*** [2.90]	12.86*** [3.84]
(4) Year FE Instead of Country-Year FE (N=4,725; F: .; 92.3; 27.2; 13.0)	4.67*** [0.70]	7.85*** [2.21]	10.81*** [3.33]	11.87*** [4.42]
(5) SE (Cluster Country) (N=4,725; F: .; 69.8; 54.4; 13.3)	3.47*** [0.80]	8.35*** [2.42]	12.89*** [3.10]	16.49*** [4.60]
(6) $\Delta_{t-30}^t \ln(\text{Market Access})$ (N=4,725; F: .; 202.1; 73.1; 45.6)	1.05*** [0.20]	2.39*** [0.58]	3.69*** [0.83]	4.45*** [1.14]
(7) Reduced Form Effect of the IVs (N=4,725; F: .; .; .; .)		12.75*** [3.25]	19.52*** [5.13]	23.27*** [8.17]
(8) Obs. Weighted by Initial Pop. t-10 (N=4,725; F: .; 62.3; 23.9; 9.2)	2.67*** [0.93]	7.68*** [2.66]	9.21** [4.04]	11.90** [6.03]
(9) Alder (2017) Speeds (N=4,725; F: .; 72.5; 35.6; 10.7)	3.66*** [0.68]	10.64*** [2.75]	17.11*** [4.19]	21.76*** [5.64]
(10) Shiferaw et al. (2013) Speeds (N=4,725; F: .; 48.1; 13.1; 6.2)	4.05*** [0.76]	14.22*** [3.31]	21.29*** [5.10]	26.67*** [6.82]
(11) Road + Railroad MA (N=4,725; F: .; 53.9; 27.9; 10.1)	3.40*** [0.63]	8.61*** [2.42]	14.26*** [3.72]	17.04*** [4.69]
(12) Sigma = 8 (N=4,725; F: .; 105.9; 19.5; 6.3)	1.12*** [0.22]	3.25*** [0.87]	5.34*** [1.33]	6.21*** [2.03]
(13) Sigma = 2 (N=4,725; F: .; 75.2; 45.9; 28.8)	13.80*** [2.62]	19.87*** [5.90]	27.50*** [8.54]	35.26*** [11.03]
(14) Border Cost: 4 Hours (N=4,725; F: .; 104.8; 48.7; 13.8)	3.32*** [0.61]	8.77*** [2.19]	12.40*** [3.19]	15.83*** [4.51]
(15) Border Cost: 24 Hours (N=4,725; F: .; 119.9; 46.8; 13.2)	3.07*** [0.59]	8.06*** [1.99]	11.16*** [2.86]	13.94*** [3.94]
(16) Iceberg Costs (Container Value = 10,000 USD) (N=4,725; F: .; 847.3; 213.7; 70.0)	49.90** [19.75]	52.75** [20.70]	46.14** [21.55]	40.79 [25.20]
(17) Iceberg Costs & Including Own City Size (N=4,725; F: .; 6.7; 4.0; 2.6)	72.08*** [20.67]	58.76** [23.60]	54.48** [26.91]	52.77 [36.28]
(18) Drop South Africa Neighbors (N=4,413; F: .; 96.3; 14.5; 8.9)	3.72*** [0.67]	8.44*** [2.25]	13.14*** [3.35]	17.02*** [4.62]
(19) Drop North Africa Neighbors (N=4,238; F: .; 83.9; 23.2; 12.5)	3.27*** [0.63]	5.24*** [1.88]	7.66** [3.17]	8.88** [4.06]
(20) Drop Arabian Peninsula Neighbors (N=4,378; F: .; 76.2; 18.8; 6.5)	2.53*** [0.58]	5.21*** [1.84]	9.06** [3.54]	10.42** [4.73]
(21) Paving only (N=4,725; F: .; 49.8; 44.9; 10.4)	3.23*** [0.65]	8.82*** [2.49]	14.71*** [4.32]	23.13*** [7.11]
(22) Improving only (N=4,725; F: .; 36.2; 11.6; 5.2)	3.82*** [0.79]	11.18*** [3.48]	20.32*** [5.85]	24.48*** [8.32]

Notes: See text for details. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Table A.10: Effects on Market Access to Cities vs. Market Access to Other Characteristics

Other Characteristic:	Col.(1)-(4): Total Population				Col.(5)-(8): Open Mine			
	(1) OLS	(2) IV5	(3) IV10	(4) IV15	(5) OLS	(6) IV5	(7) IV10	(8) IV15
Overall Effect: Cities	2.31*** [0.67]	11.22*** [3.08]	9.73*** [3.54]	10.46** [4.63]	3.35*** [0.66]	11.16*** [2.92]	8.31*** [2.60]	6.52*** [2.47]
Overall Effect: Other	2.82*** [0.91]	-3.79 [3.26]	-0.72 [4.56]	-0.99 [6.64]	0.75 [0.73]	-6.77* [3.69]	0.11 [3.47]	4.40 [5.33]
IV F-Stat		2.1	1.5	0.8		8.6	6.0	3.9

Notes: Main sample of 4,725 cell-years. Robust SEs, clustered by 1960 province. *, **, *** = 10, 5, 1% significance.

Table A.11: Effects on Population and Night Lights for the Sample with Night Lights Data

	(1) OLS	(2) IV: Excl. 5	(3) IV: Excl. 10	(4) IV: Excl. 15
<i>Panel A:</i> $100\Delta_{t-10}^t \ln \text{Urban Population}$				
$\Delta_{t-10}^t \ln \text{Market Access}$	1.58*** [0.46]	0.55 [1.31]	1.29 [2.50]	2.76 [4.22]
$\Delta_{t-20}^{t-10} \ln \text{Market Access}$	1.49*** [0.36]	4.50*** [1.02]	7.68*** [1.59]	9.30*** [3.05]
$\Delta_{t-30}^{t-20} \ln \text{Market Access}$	0.79** [0.32]	3.13** [1.26]	4.69** [2.16]	7.68** [3.18]
Overall Effect	3.86*** [0.79]	8.18*** [2.34]	13.67*** [3.85]	19.74*** [6.20]
<i>Panel B:</i> $100\Delta_{t-10}^t \ln(\text{Light Intensity Per Capita})$				
$\Delta_{t-10}^t \ln \text{Market Access}$	-0.46 [3.18]	21.75** [9.98]	42.20*** [12.60]	68.70*** [20.32]
$\Delta_{t-20}^{t-10} \ln \text{Market Access}$	0.56 [2.94]	9.77 [8.05]	1.33 [12.54]	-5.15 [17.56]
$\Delta_{t-30}^{t-20} \ln \text{Market Access}$	0.20 [2.15]	0.77 [4.66]	-4.44 [7.62]	-11.12 [11.35]
Overall Effect	0.30 [5.31]	32.29*** [11.87]	39.08** [19.00]	52.43* [28.43]
1st stage Kleibergen-Paap F		53.21	29.79	9.83
<i>Panel C:</i> $100\Delta_{t-10}^t \ln \text{Urban Population}$				
$\Delta_{t-10}^t \ln \text{Market Access}$	1.56*** [0.46]	0.39 [1.33]	1.03 [2.55]	2.40 [4.30]
$\Delta_{t-20}^{t-10} \ln \text{Market Access}$	1.47*** [0.35]	4.41*** [1.02]	7.64*** [1.60]	9.25*** [3.05]
$\Delta_{t-30}^{t-20} \ln \text{Market Access}$	0.76** [0.32]	3.09** [1.27]	4.66** [2.16]	7.67** [3.19]
$100\Delta_{t-10}^t \ln \text{light}$	0.76*** [0.24]	0.67*** [0.24]	0.59** [0.25]	0.48* [0.28]
Overall Effect of MA	3.79*** [0.79]	7.89*** [2.36]	13.33*** [3.90]	19.33*** [6.29]
1st stage Kleibergen-Paap F		53.05	29.27	9.840

Notes: See Table 3. N=3,591. Robust SEs, clustered by 1960 province. *, **, *** = 10, 5, 1% significance.

Table A.12: Robustness Checks: Reallocation

(1) 3x3 Excl. National, Regional & Top 5 (N=3,066; F: \downarrow ; 31.7; 12.7; 3.1)	7.06*** [1.03]	8.44*** [3.08]	13.25** [5.49]	15.26** [6.65]
(2) 5x5 Excl. National, Regional & Top 5 (N=2,468; F: \downarrow ; 7.7; 7.5; 6.9)	7.78*** [1.27]	6.66** [2.96]	8.65* [4.93]	8.72 [6.28]
(3) 7x7 Excl. National, Regional & Top 5 (N=1,974; F: \downarrow ; 27.6; 4.0; 1.5)	8.63*** [1.49]	12.55*** [3.26]	14.75*** [5.41]	13.22 [8.43]
(4) 9x9 Excl. National, Regional & Top 5 (N=1,561; F: \downarrow ; 20.6; 10.5; 4.0)	9.90*** [1.66]	4.38 [3.36]	10.08* [6.04]	9.36 [8.90]

Notes: See Table 4. Outcome variable is $100\Delta_{t-10}^t \ln(\text{Light Intensity})$. N=3,591 cell-decades.

Table A.13: Heterogeneous Effects: Population and Market Access: Alternative Specifications

	OLS Diff. (1)	Col. (2)–(4): IV5			IV10 Diff. (5)	IV15 Diff. (6)
		0 (2)	1 (3)	Diff. (4)		
(1) < Median 1960 MA (F: \downarrow ; 16.9; 7.3; 2.1. Sh.: 0.56)	7.33*** [1.70]	6.11 [4.87]	9.03*** [2.72]	2.92 [5.26]	6.46 [7.06]	18.90** [7.54]
(2) < Median Pop. t-30 (F: \downarrow ; 28.8; 17.8; 7.2. Sh: 0.49)	3.39*** [1.25]	3.96 [3.74]	7.75*** [2.09]	3.79 [3.60]	10.58* [5.65]	12.71* [6.61]
(3) >Med. Dist. Top 1960 & 2010 Cities (F: \downarrow ; 40.1; 19.7; 2.1. Sh: 0.49.)	7.40*** [1.26]	5.49** [2.54]	8.87*** [2.42]	3.38 [2.45]	7.63** [3.16]	13.17*** [4.92]
(4) >Med. Dist. Top 1960 Relative to SSA (F: \downarrow ; 5.7; 2.4; 2.3. Sh: 0.50.)	7.48*** [1.58]	-1.73 [2.07]	9.51*** [2.54]	11.25*** [3.04]	17.19*** [3.65]	20.59*** [5.26]
(5) >Med. Dist. Top 1960 Cities (Drop Top) (F: \downarrow ; 59.7; 13.4; 2.6. Sh: 0.50.)	7.49*** [1.39]	2.36 [2.72]	9.21*** [2.24]	6.85** [2.66]	13.08*** [3.93]	21.77*** [6.25]
(6) No Paved/Improved Road 1960 (F: \downarrow ; 27.4; 7.4; 2.4. Sh.: 0.33)	1.40 [1.17]	6.37*** [1.99]	11.29*** [2.98]	4.91* [2.68]	8.65* [4.44]	12.07* [6.33]
(7) No Railroad 1960 (F: \downarrow ; 48.1; 14.8; 0.7. Sh: 0.77.)	0.99 [1.43]	5.51* [3.27]	8.81*** [2.39]	3.30 [3.36]	10.01** [4.59]	12.82** [5.59]
(8) >Med. Dist. 1960 Port (F: \downarrow ; 6.5; 2.1; 1.6. Sh.: 0.50)	-0.94 [1.17]	3.30* [1.99]	10.53*** [2.69]	7.22** [3.21]	11.52** [4.62]	13.07** [6.23]
(9) >Med. Dist. 2005 Port (F: \downarrow ; 6.7; 1.6; 1.0. Sh.: 0.50)	-0.97 [1.16]	3.30* [1.96]	10.63*** [2.69]	7.33** [3.17]	11.95*** [4.60]	14.46** [6.20]
(10) >Med. Dist. 2007 Airport (F: \downarrow ; 15.4; 3.5; 2.0. Sh.: 0.50)	0.43 [1.08]	3.61** [1.66]	11.62*** [2.81]	8.01*** [2.87]	10.25** [4.53]	11.37* [6.47]
(11) <25th %ile 1960 MA (F: \downarrow ; 34.7; 8.1; 2.0. Sh.: 0.24)	5.82*** [1.25]	4.03 [3.57]	8.87*** [3.02]	4.84 [4.19]	19.21*** [7.26]	39.16*** [12.33]
(12) <75th %ile 1960 MA (F: \downarrow ; 2.8; 0.2; 0.1. Sh.: 0.74)	8.92** [3.60]	1.92 [6.84]	9.84*** [2.49]	7.91 [7.07]	15.13 [17.10]	26.33 [53.34]
(13) <Med. 1960 MA Relative to SSA (F: \downarrow ; 3.6; 6.8; 4.9. Sh.: 0.50)	5.36*** [1.72]	-2.89 [4.83]	9.99*** [2.59]	12.87** [5.14]	19.31** [8.12]	15.71* [8.49]
(14) <25th %ile Pop. t-30 (F: \downarrow ; 30.6; 15.1; 6.9. Sh: 0.52)	3.47*** [1.26]	3.90 [3.86]	7.84*** [2.11]	3.93 [3.65]	11.39** [5.67]	14.09** [6.69]
(15) <75th %ile Pop. t-30 (F: \downarrow ; 15.6; 9.7; 4.1. Sh: 0.74)	2.26* [1.28]	6.37*** [2.32]	10.79*** [2.53]	4.42 [2.79]	11.23** [5.13]	18.88** [8.11]
(16) <Med. Pop. t-30 Relative to SSA (F: \downarrow ; 34.0; 21.9; 5.9. Sh: 0.25.)	1.76 [1.19]	5.15** [2.20]	12.01*** [2.45]	6.87*** [2.52]	15.25*** [5.07]	22.79*** [7.82]
(17) <Med. Pop. t-30 (Drop Top Cities) (F: \downarrow ; 33.6; 22.2; 6.7. Sh: 0.52.)	2.92** [1.37]	3.88 [3.93]	7.72*** [2.09]	3.84 [3.73]	10.56* [5.78]	12.54* [6.75]

Notes: Each row reports results from variants of Table 5. The 1st stage F-statistics (“F”) and the share of city-years with the dummy equal to one (“Sh”) are reported in the left column.

Table A.14: Heterogeneous Effects: Land Suitability and Political Economy: Alternative Specifications

	OLS Diff. (1)	Col. (2)–(4): IV5 0 1 Diff. (2) (3) (4)			IV10 Diff. (5)	IV15 Diff. (6)
(1) Land Suitability >75% (F: .; 54.5; 17.7; 6.7. Sh: 0.05)	-1.27 [2.20]	8.89*** [2.18]	-2.24 [4.50]	-11.13** [4.79]	-10.08 [8.01]	-15.61* [9.01]
(2) Food Crop Suitability > 75% (F: .; 54.9; 19.8; 6.7. Sh: 0.04.)	-1.14 [2.34]	8.79*** [2.18]	-0.30 [4.43]	-9.09* [4.63]	-7.28 [6.59]	-12.31 [7.65]
(3) Food Crop Suitability < 25% (F: .; 18.2; 8.2; 3.5. Sh: 0.12.)	0.04 [1.51]	6.64*** [2.12]	13.43*** [4.75]	6.79 [4.94]	15.49** [6.77]	26.58*** [8.92]
(4) Cash Crop Suitability > 75% (F: .; 2.1; 0.4; 0.1. Sh: 0.03.)	-2.22 [3.39]	8.47*** [2.18]	2.20 [7.07]	-6.27 [7.23]	-5.79 [11.20]	-5.67 [14.17]
(5) Cash Crop Suitability < 25% (F: .; 21.5; 5.3; 3.5. Sh: 0.14.)	-1.27 [1.50]	6.71*** [2.11]	12.53*** [4.34]	5.82 [4.42]	12.84** [6.07]	22.94*** [8.03]
(6) >75 %tile Average Rainfall (F: .; 38.1; 14.8; 5.8. Sh: 0.74.)	-3.83*** [1.34]	7.94*** [2.47]	8.70*** [2.93]	0.76 [3.25]	3.93 [5.88]	9.14 [9.06]
(7) <25 %tile Average Rainfall (F: .; 14.2; 8.5; 5.3. Sh: 0.24.)	-1.37 [1.74]	8.29*** [2.28]	16.59*** [6.07]	8.30 [6.14]	16.33** [7.54]	23.28*** [8.83]
(8) > Med. Dist. Mine t-30 (F: .; 51.1; 14.6; 5.9. Sh.: 0.50)	4.23*** [1.33]	8.04*** [2.53]	8.24*** [2.65]	0.20 [2.93]	0.57 [4.62]	2.22 [7.18]
(9) Leader's Origin 150km (90th %ile) (F: .; 9.5; 7.6; 8.3. Sh: 0.11)	-3.51** [1.56]	8.73*** [1.76]	3.45 [6.38]	-5.29 [6.40]	-5.94 [8.72]	-7.95 [10.45]
(10) Non-Democratic Leader's Orig. 150km Mean (F: .; 21.3; 11.1; 8.2. Sh: 0.21)	-0.79 [1.26]	9.28*** [1.81]	2.14 [4.03]	-7.13* [4.04]	-5.82 [5.06]	-8.07 [6.09]
(11) Non-Demo. Leader's Orig. 150km (90th %ile) (F: .; 68.7; 24.3; 8.1. Sh: 0.11)	-2.66* [1.56]	9.26*** [1.88]	-3.16 [3.23]	-12.42*** [3.54]	-17.78*** [5.91]	-21.98*** [8.13]
(12) Leader's Origin 250km (Mean) (F: .; 27.3; 4.8; 7.8. Sh: 0.39)	-1.50 [1.07]	9.62*** [1.92]	3.29 [3.05]	-6.33* [3.26]	-4.42 [4.68]	-6.67 [6.07]
(13) Leader's Origin 250km (90th %ile) (F: .; 11.4; 2.5 ; 8.3 . Sh: 0.20)	-2.56* [1.44]	8.72*** [1.89]	2.37 [3.52]	-6.35 [3.89]	-11.18* [6.24]	-15.24* [9.05]
(14) Leader's Ethnicity (Murdock) (Mean) (F: .; 14.8; 12.6; 4.7. Sh: 0.21)	-1.34 [1.32]	9.15*** [1.87]	3.83 [4.94]	-5.31 [5.11]	-2.78 [6.87]	-5.40 [8.30]
(15) Leader's Ethnicity (Murdock) (90th %ile) (F: .; 10/0; 12.8; 4.3. Sh: 0.13)	-2.07 [1.51]	8.90*** [1.94]	2.83 [3.34]	-6.06 [3.86]	-8.67 [5.59]	-11.44 [7.75]
(16) Provincial Capital in 1960 (F: .; 16.2; 16.7; 4.8. Sh.: 0.16)	-0.08 [1.18]	7.42*** [2.43]	10.89*** [3.07]	3.48 [3.43]	2.64 [4.68]	7.47 [6.37]
(17) Provincial Capital in 2010 (F: .; 18.6; 5.1; 3.7. Sh.: 0.23)	1.21 [1.12]	4.91** [2.28]	11.34*** [2.93]	6.43** [3.08]	8.04* [4.74]	12.92** [6.19]

Notes: Each row reports results from variants of Table 5. The 1st stage F-statistics ("F") and the share of city-years with the dummy equal to one ("Sh") are reported in the left column.

Table A.15: First Stage Regressions: Heterogeneity with Respect to City Size and Remoteness

Dependent Variable: Δ In Market Access	IV: Exclude 5			IV: Exclude 10			IV: Exclude 15		
	t-10;t	t-20;t-10	t-30;t-20	t-10;t	t-20;t-10	t-30;t-20	t-10;t	t-20;t-10	t-30;t-20
<i>Panel A: Distance to Top Cities 1960</i>									
Δ_{t-10}^t In MA due to Roads Only	1.97*** [0.17]	-0.30*** [0.10]	-0.16 [0.10]	2.54*** [0.33]	-0.64*** [0.23]	-0.53* [0.29]	2.75*** [0.36]	-0.85* [0.44]	-0.63 [0.48]
Δ_{t-20}^{t-10} In MA due to Roads Only	0.17 [0.11]	2.27*** [0.15]	-0.44*** [0.09]	1.37*** [0.35]	3.59*** [0.46]	-0.90*** [0.38]	2.20*** [0.62]	3.26*** [0.59]	-0.54 [0.68]
Δ_{t-30}^{t-20} In MA due to Roads Only	-0.21 [0.22]	0.20 [0.18]	2.20*** [0.20]	0.33 [0.59]	2.78*** [0.78]	4.47*** [0.69]	2.00** [0.90]	4.34*** [1.26]	4.06*** [1.23]
> MedDist. Top $\times \Delta_{t-10}^t$ In MA Roads Only	-0.30 [0.19]	-0.13 [0.14]	-0.08 [0.18]	-0.49 [0.39]	-0.38 [0.31]	-0.09 [0.36]	-0.74 [0.60]	-0.84 [0.55]	-0.03 [0.57]
> MedDist. Top $\times \Delta_{t-20}^{t-10}$ In MA Roads Only	-0.13 [0.11]	-0.73*** [0.19]	0.04 [0.11]	-1.47*** [0.33]	-1.83*** [0.42]	0.13 [0.41]	-2.16*** [0.57]	-1.48** [0.68]	-1.05 [0.80]
> MedDist. Top $\times \Delta_{t-30}^{t-20}$ In MA Roads Only	0.38 [0.23]	-0.14 [0.16]	-0.64*** [0.23]	-0.17 [0.60]	-2.68*** [0.80]	-2.67*** [0.75]	-1.75* [0.93]	-3.96*** [1.29]	-1.94 [1.19]
<i>Panel B: Market Access 1960.</i>									
Δ_{t-10}^t In MA due to Roads Only	2.08*** [0.26]	-0.05 [0.14]	0.12 [0.25]	2.98*** [0.46]	0.31 [0.39]	-0.90 [1.00]	5.07*** [1.39]	-0.20 [2.38]	-5.53** [2.36]
Δ_{t-20}^{t-10} In MA due to Roads Only	-0.06 [0.25]	2.04*** [0.25]	-0.04 [0.12]	0.87** [0.37]	2.93*** [0.45]	0.30 [0.34]	2.70*** [1.02]	4.73*** [1.80]	-0.32 [1.57]
Δ_{t-30}^{t-20} In MA due to Roads Only	0.02 [0.11]	-0.26 [0.47]	1.80*** [0.29]	0.18 [0.33]	0.62** [0.29]	2.73*** [0.29]	4.10** [2.08]	1.74 [2.40]	3.94* [2.10]
< Med MA 1960 $\times \Delta_{t-10}^t$ In MA Roads Only	-0.40 [0.26]	-0.44*** [0.16]	-0.48* [0.26]	-0.93* [0.51]	-1.35*** [0.45]	0.25 [1.01]	-3.01** [1.44]	-1.31 [2.37]	4.81** [2.42]
< Med MA 1960 $\times \Delta_{t-20}^{t-10}$ In MA Roads Only	0.08 [0.25]	-0.51* [0.26]	-0.50*** [0.15]	-1.01** [0.39]	-1.24*** [0.44]	-1.35*** [0.37]	-2.70** [1.05]	-3.15* [1.84]	-1.58 [1.60]
< Med MA 1960 $\times \Delta_{t-30}^{t-20}$ In MA Roads Only	0.08 [0.13]	0.27 [0.46]	-0.34 [0.30]	-0.11 [0.35]	-0.63* [0.33]	-1.17*** [0.32]	-3.99* [2.09]	-1.54 [2.42]	-2.22 [2.08]
<i>Panel C: City Size t-30.</i>									
Δ_{t-10}^t In MA due to Roads Only	1.69*** [0.12]	-0.30*** [0.10]	-0.15 [0.10]	2.12*** [0.20]	-0.52** [0.23]	-0.43** [0.20]	2.69*** [0.45]	-0.87** [0.38]	-0.63 [0.40]
Δ_{t-20}^{t-10} In MA due to Roads Only	0.15 [0.10]	1.85*** [0.15]	-0.43*** [0.09]	0.22 [0.34]	2.33*** [0.33]	-0.76*** [0.24]	0.71 [0.63]	3.16*** [0.62]	-1.55*** [0.39]
Δ_{t-30}^{t-20} In MA due to Roads Only	-0.02 [0.10]	0.21 [0.14]	1.79*** [0.19]	-0.24* [0.13]	0.45** [0.21]	2.32*** [0.44]	-0.30 [0.29]	0.74 [0.64]	3.48*** [0.67]
< MedPop. t-30 $\times \Delta_{t-10}^t$ In MA Roads Only	0.15 [0.15]	-0.12 [0.14]	-0.07 [0.15]	0.14 [0.42]	-0.62* [0.33]	-0.10 [0.33]	-0.79 [0.59]	-0.72 [0.52]	0.14 [0.54]
< MedPop. t-30 $\times \Delta_{t-20}^{t-10}$ In MA Roads Only	-0.09 [0.12]	-0.25 [0.20]	0.07 [0.10]	-0.27 [0.36]	-0.51 [0.46]	0.04 [0.29]	-0.55 [0.66]	-1.42** [0.67]	0.10 [0.45]
< MedPop. t-30 $\times \Delta_{t-30}^{t-20}$ In MA Roads Only	0.19* [0.11]	-0.13 [0.12]	-0.17 [0.21]	0.47*** [0.15]	-0.26 [0.23]	-0.43 [0.50]	0.67** [0.31]	-0.22 [0.65]	-1.21* [0.68]

Notes: 4,725 obs. Coefficients of interest highlighted in bold. Robust SEs, clustered by 1960 province, in brackets. *, **, *** = 10, 5, 1% significance.

Table A.16: Foreign vs. Domestic and Overland vs. Overseas Market Access: Alternative Specifications

	(1) OLS	(2) IV5	(3) IV10	(4) IV15
(1) Domestic GDP-weighted	3.12*** [0.56]	6.82*** [1.82]	8.73*** [2.65]	10.89*** [3.59]
Foreign GDP-weighted	2.42** [1.14]	0.85 [3.20]	0.07 [3.38]	-0.90 [3.68]
First stage F-statistic		27.39	9.06	4.15
(2) Overland drop w/in 50km of ports	2.82*** [0.68]	6.52*** [2.55]	4.95* [2.60]	3.45 [2.78]
Overseas drop w/in 50km of ports	4.21 [2.84]	3.65 [3.73]	10.04** [5.02]	16.35** [6.92]
First stage F-statistic		46.51	29.14	9.18
(3) Overland include overseas	2.68*** [0.68]	7.85*** [2.71]	6.04** [2.76]	4.17 [3.03]
Overseas	4.03 [2.85]	0.63 [4.08]	7.16 [5.26]	14.27* [7.40]
First stage F-statistic		47.13	26.98	7.74
(4) Overland pop 1960	1.53 [1.23]	7.42** [2.91]	5.95* [3.08]	3.96 [3.33]
Overseas pop 1960	6.93** [2.85]	3.44 [3.47]	8.31* [4.92]	15.11** [6.95]
First stage F-statistic		45.93	43.51	10.30
(5) Overland ports 1960	3.01*** [0.60]	7.25*** [2.10]	5.89*** [2.03]	3.99* [2.14]
Overseas ports 1960	0.71 [2.33]	0.63 [3.44]	5.39 [4.28]	12.06** [5.41]
First stage F-statistic		47.76	44.35	31.18

Table A.17: Estimated Aggregate Urban Effects of Road Upgrading 1960-2010

	(1) Share of Urban Pop. Growth 1960-2010 (%)			(2) Diff. in Urban Share 2010 (%)			(3) Share of Increase in Urban Share 1960-2010 (%)		
	IV5	IV10	IV15	IV5	IV10	IV15	IV5	IV10	IV15
(1) Average Effects	3.0	4.7	6.2	-0.6	-1.0	-1.3	4.1	6.8	8.9
(2) Het.: Distance to Top Cities 1960	1.8	2.5	3.1	-0.4	-0.5	-0.7	2.7	3.4	4.8
(3) Het.: Market Access in 1960	2.7	5.3	7.7	-0.6	-1.1	-1.6	4.1	7.5	11.0
(4) Het.: Population in 1960	1.8	2.1	3.4	-0.4	-0.4	-0.7	2.7	2.7	4.8
(5) Het.: Land Suitability (25%)	2.8	4.0	4.9	-0.6	-0.9	-1.0	4.1	6.2	6.8
(6) Het.: Land Suitability (75%)	3.1	4.7	6.3	-0.7	-1.0	-1.3	4.8	6.8	8.9
(7) Het.: Leader Favoritism	2.0	3.7	4.9	-0.4	-0.8	-1.1	2.7	5.5	7.5
(8) Het.: Regional Capitals 1960	3.3	4.8	6.8	-0.7	-1.0	-1.5	4.8	6.8	10.3
(9) Het.: Regional Capitals 2010	3.3	4.6	6.6	-0.7	-1.0	-1.4	4.8	6.8	9.6

Table A.18: Estimated Aggregate Urban Population Effects of the TAH 2010-2040

Speed for TAH Roads (kph):	Predicted Percentage Increase in Urban Population (%)					
	80-IV5	80-IV10	80-IV15	100-IV5	100-IV10	100-IV15
(1) Average Effects	0.9	1.5	1.8	1.8	2.8	3.4
(2) Het.: Distance to Top Cities 1960	0.5	0.9	1.0	0.9	1.7	1.8
(3) Het.: Market Access in 1960	0.9	1.9	2.3	1.7	3.7	4.4
(4) Het.: Population in 1960	0.9	1.3	1.5	1.7	2.5	2.9
(5) Het.: Land Suitability (25%)	0.9	1.3	1.5	1.7	2.5	2.8
(6) Het.: Land Suitability (75%)	0.9	1.4	1.7	1.8	2.8	3.4
(7) Het.: Leader Favoritism	0.9	1.4	1.7	1.7	2.7	3.3
(8) Het.: Regional Capitals 1960	1.1	1.5	2.0	2.0	3.0	3.9
(9) Het.: Regional Capitals 2010	1.1	1.5	1.9	2.1	2.9	3.7