NOT FOR PUBLICATION: WEB APPENDICES

History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya

by Remi Jedwab, Edward Kerby and Alexander Moradi

— APPENDIX FIGURES —

— APPENDIX TABLES —

NOT FOR PUBLICATION

DATA DESCRIPTION, VARIABLE CONSTRUCTION AND SAMPLE SELECTION

This appendix describes in details the data we use in our analysis.

Spatial Units for Kenya:

We assemble data for 473 locations of about 16x16 km from 1901 to 2009. The 473 locations are the third level Kenyan administrative units in 1962, one year before independence. A paper map of the locations is obtained from the main report of the 1962 *Population and Housing Census*. We also use a set of maps available in Soja (1968). We then use a GIS map of sub-locations – the fourth level administrative units – in 1999 that is available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS) to reconstruct the GIS map of the locations in 1962. Sub-locations in 1999 were easily reaggregated in GIS to match the location boundaries of 1962, which we use throughout our analysis. Since the locations do not have the same size, we control for the location's area in the regressions. The locations belong to 35 administrative districts and 8 administrative provinces in 1962.

Railway Data for Kenya:

We obtain the layout of railway lines in GIS from *Digital Chart of the World*. We use Hill (1949) and Ochieng and Maxon (1992) to recreate the history of railway construction. For each line we know when it was planned, initiated and finished. Placebo lines consist of explorer routes (from the coast to Lake Victoria) and branch lines that were proposed but not built. A map of these explorer routes is obtained from the *Government Survey of Kenya* (1959). A map of the branch lines is obtained from a report of the *Colony and Protectorate of Kenya* (1926). Most of those "placebo lines" became roads later. Using the GIS road network also available from *Digital Chart of the World*, we recreate those placebo lines in GIS. We calculate for each location the Euclidean distance (km) from the location centroid to each real or placebo line. Lastly, we create a set of location dummies equal to one if the location centroid is less than X km away from the line: 0-10, 10-20, 20-30 and 30-40 km. We create a location dummy equal to one if the location contains a rail station in 1938, whose list we obtained from one of the railway reports. We do not have data on the location of railroad stations at independence. Data on aggregate railroad traffic came from various annual reports of the Colony and Protectorate of Kenya, the East African Protectorate and the Kenya Railways Corporation, as well as the World Bank (2013).

Urban Data for Kenya:

We collect urban data from lists of "urban localities" available in the reports of the *Population and Housing Censuses* 1962, 1969, 1979, 1989, 1999 and 2009. The reports consistently lists all localities above 2,000 inhabitants and their population size. Defining as a city any locality with more than 2,000 inhabitants, we obtain a geospatialised sample of 247 cities for all these years. The reports sometimes use a lower threshold than 2,000, but never for all years. We then obtain the population of the five cities of Kenya with more than 2,000 inhabitants in 1901 from *The Handbook of the Foreign Office for Kenya, Uganda and Tanzania* (1920). We compare these estimates with what we find in various historical sources available on the internet. The GeoNet data base is used to retrieve the geographical coordinates of each city. Using GIS, we recalculate total urban population for each location.

Population Data for Kenya:

Total population at the location level in 1962 is obtained by entering the 1962 *Population and Housing Census*. Total population for the year 1999 is obtained from a GIS database available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS). The GIS database reports the total population of each sub-location in 1999. Using GIS, we re-aggregate the sub-locations to match the location boundaries of the year 1962. Rather unfortunately, we do not have total population data for the year 2009. The annual total population of Kenya from 1900 to 2009 is then available on the website of *Populstats*, a database on the historical demography of all countries: http://www.populstat.info/.

Ethnic Data for Kenya:

Firstly, the 1962 *Population and Housing Census* reports the number of Europeans and Asians for each one of the 473 locations in 1962. The African population is estimated by subtracting the European and Asian populations from the total population. We use the same source to obtain the number of Kikuyus and Kalenjins for each location in 1962. Secondly, the 1962 census was the first exhaustive census in Kenya, whereas European censuses were conducted every 5 years before independence. The reports of these censuses do not display the distribution of Europeans at a fine spatial level. Instead, we obtain the total number of European voters at the location level in 1933. The *Voter Registries for the Election to the Legislative Council* of the *Official Gazette of the Colony and Protectorate of Kenya* of 1933 lists the name, sector, occupation and address of all European voters for these years. The voter registry was then geospatialised using the address of the voters. We know the number of workers for the following sectors: "agriculture", "industry", "commerce", "transport", "government", "religion", "education", "health", and "personal services". Besides, we know the number of workers for the following occupations (using the standard HISCO classification: http://hisco.antenna.nl/): "Professional, technical and related workers", "Administrative and managerial workers", "Clerical and related workers", "Sales workers", "Service workers", "Agricultural, animal husbandry and forestry workers, fishermen and hunters" and "Others". The aggregate sectoral and occupational distributions of the European settlers was then compared between the voter registry of 1933 and the European *Population Censuses* of 1931. We also obtain from the main report of 1962 *Population Census* the sectoral and occupational distributions of the European and Asian populations in 1962. Thirdly, the total number of Kenyan Europeans and Kenyan Asians for other years from 1900 to 2009 is obtained from the reports of the European and Asian *Population Censuses* in the colonial period, and the reports of *Population Censuses* post-independence. Fourthly, the 473 locations belong to 27 ethnic homelands when using the digitised ethnic map of Murdock (1959).

Commercial Agriculture Data for Kenya:

For each location in 1960, we know from the report of 1960 European *Agricultural Census* how much land is devoted to the European cultivation of four crops (in thousand acres): coffee, maize, wheat and tea (no data is available for sisal production). We use these measures in 1960 as proxies for cultivation in 1962. We have no reliable spatial data on African cultivation in 1960-62. However, Africans were not allowed to grow these crops until 1954, with the exception of maize. Data on aggregate production and the composition of exports in the colonial period is obtained from the *Annual Reports of the Colony and Protectorate of Kenya* and from reports of the various European *Agricultural*

Censuses that took place before independence. Data on aggregate production and the composition of exports in the post-independence period is obtained from FAO (2013). Lastly, we obtain a map of existing land regimes in 1938: *alienated*, *suitable for alienation*, *native reserves*, *forest*, and *rest of the colony*.[1] The map was digitized in GIS which allowed us to recreate for each location the area shares (%) of each regime.

Geographical Data for Kenya:

Data on soil aridity comes from a GIS UNEP/GRID map of agro-ecological zones available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS). The map displays the areas considered as arid in Kenya. We then use GIS to reconstruct the share of arid soils at the location level (%). Data on soil suitability for agriculture comes from Ogendo (1967). One of the maps in the study shows the areas of high agricultural potential at independence, for no crop in particular. We then use GIS to reconstruct the share of soils suitable for agriculture at the location level (%). Data on soil suitability for coffee and tea comes from the *Farm Management Handbook of Kenya* 1982-83. The handbook contains a set of maps showing the various areas of potential cultivation for coffee and tea. We then use GIS to reconstruct the shares of soils suitable for coffee and tea at the location level (%). The mean and standard deviation of altitude (m) in each location was then reconstructed using GIS topographical data from the *SRTM 90m Digital Elevation Database*. Average annual rainfall (mm) for each location for the period 1950-2000 was reconstructed using a map available on the website of the World Resources Institute (http://www.wri.org/resources). Average annual temperature for each location for the period 1950-2000 was a also reconstructed using a GIS map available from Hijmans, Cameron, Parra, Jones and Jarvis (2005): http://www.worldclim.org/. Lastly, we use *Global GIS* to obtain a GIS map of Kenya's lakes. We then create for each location the area share of lakes (%) and the Euclidean distance (km) to a lake.

Economic Geography Data for Kenya:

Firstly, we use GIS to create a "coastal location" dummy variable equal to one if the location borders the sea. Secondly, we use GIS to obtain the total area (sq km) of the location. Thirdly, the report of the 1962 *Population Census* lists the eight provincial capitals in 1962. We create a "provincial capital" dummy equal to one if the location contains a provincial capital. Fourthly, for each location, we also use GIS to get the Euclidean distances (km) to the coast and the three nodes of Mombasa, Nairobi and Kisumu. Lastly, we use a paper map of historical settlement patterns in the 19th century that is available in Soja (1968). The map shows the areas where the "major settled groups" and "pastoralist groups" lived. Using GIS, we reconstruct the area shares (%) of each group. The map also indicates the location of the "isolated groups (mainly hunters and gatherers)". Using GIS, we create a dummy variable equal to one if the location contains one of these isolated groups.

Other Transportation Networks Data for Kenya:

Roads in 1901 are described in Hill (1949) and Ochieng & Maxon (1992). The GIS maps of paved and improved roads in 1964 and 2002 are obtained from Burgess (2014) who use Michelin paper maps to recreate the 1964 and 2002 road networks in GIS (which we use as proxies for 1962 and 2009 respectively), distinguishing paved (bitumenized) and improved (laterite) roads. We also use GIS to get the Euclidean distances of the location to a paved road or an improved road in 1964 and 2002. Lastly, we use *Digital Chart of the*

[1] The source is the *Map of the Colony & Protectorate of Kenya* published in 1938 by Waterlow & Sons and Dunstable & Watford in London.

World to obtain a GIS map of the non-paved and non-improved dirt roads in 1962.

Non-Transportation Infrastructure Data for Kenya:

We have data on health, educational, institutional, communication and industrial infrastructure at the location level in 1962. In particular, we know the number of hospitals, clinics, dispensaries, secondary schools, provincial police headquarters, divisional police headquarters, police stations, police posts, post offices and postal agencies for each location in 1962. Data on health infrastructure comes from a map of *Medical Facilities* in 1960 (which we use as a proxy for 1962) that was published by the *Government Survey of Kenya* 1959. We digitised the map using GIS and estimated the number of each type of facilities (hospitals, clinics and dispensaries) for each location. Data on education infrastructure comes from a map of Secondary Schools in 1964 (which we use as a proxy for 1962) that was published in Soja (1968). We digitised the map using GIS and estimated the number of secondary schools for each location. Data on institutional infrastructure comes from a map of *Police Organisation* in 1960 (which we use as a proxy for 1962) that was published by the *Government Survey of Kenya* 1959. We digitised the map using GIS and estimated the number of each type of police stations (provincial police headquarters, divisional police headquarters, police stations and police posts) for each location. Data on industrial infrastructure comes from a map published in Ogendo (1967). The map shows the location of the main and secondary industrial centres and towns in 1962. We digitised the map using GIS and create five dummies equal to one if the location contains an important industrial town, a minor industrial town, an important industrial center, a minor industrial center, or a potential industrial center. We then have data on health and educational infrastructure at the location level in 2007-08. We know the number of hospitals, health clinics, health dispensaries, primary schools and secondary schools for each location around 2007-08. Data on health infrastructure comes from a public GIS government database of *Health Facilities* in 2008 (which we use as a proxy for 2009). We use GIS to estimate the number of each type of facilities (hospitals, health clinics and dispensaries) for each location. Data on educational infrastructure comes from a GIS government database of *Primary and Secondary Schools* in 2007 (which we use as a proxy for 2009). We use GIS to obtain the number of each type of schools for each location.

Economic Development and Human Capital for Kenya:

We use geospatialised poverty maps available at the sub location level – the unit below the location – for the year 1999 to reconstruct average poverty rates at the location level for the same year. These GIS maps are available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS). We use the poverty headcount ratio, the percentage of the population of each location living below the national poverty line (%), to estimate the share of non-poor in the location. The source of the satellite data on night lights is NOOA (2012). We follow the approach of Henderson, Storeygard and Weil (2012) and estimate average light intensity for each location for the year 2000-2001 ("2000" in our analysis). For each location i, we then estimate $MP_i = \Sigma_{j \neq i}(Y_j / D_{ij}^{\alpha})$ where Y_j is a measure of economic development of location j and D_{ij} is the network distance (in hours) via the road network in 1962 between location i and location j. Using the GIS map of paved, improved but also dirt roads in 1962, we performed in ArcGIS a least cost path analysis for each of the 473x473 pairs of locations, which allowed us to estimate the network distance in hours between each location. We assume that cars drive at speeds of 50, 35 and 20 km per hour on paved, improved and dirt roads. These parameters were obtained from Buys, Deichmann and Wheeler (2010). α is distance decay parameter. For our analysis, we create nine measures of market potential based on $Y_j = \{$total population, urban population, European crop cultivation (acres)$\}$ in 1962 and $\alpha = \{1, 2, 3\}$ (by using several α, we remain agnostic about how spatial economic interactions work). From the 10% IPUMS sample of the 2009 Population Census, we reconstruct for each constituency the shares (%) of adults aged 25 or over having completed primary, secondary or tertiary education. Constituencies were Kenya's main political units in 2009. As their boundaries do not match with the boundaries of the 473 locations, we use GIS to estimate for each location the area-weighted average of the shares using the data at the constituency level.

APPENDIX REFERENCES

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padro i Miquel. 2014. "The Value of Democracy: Evidence from Road Building in Kenya." Mimeo.

Buys, Piet, Uwe Deichmann, and David Wheeler. 2010. "Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa." *Journal of African Economies*, 19(3): 399-432.

Colony and Protectorate of Kenya. (1926). *Suggested Branch Lines. Economic and Technical Reports on the Colony and Protectorate of Kenya*. Nairobi: Government Printer.

FAO. 2013. *FAOSTAT: Trade*. Rome: Food and Agricultural Organization.?

Government Survey of Kenya. (1959). *Atlas of Kenya: a Comprehensive Series of New and Authentic Maps Prepared from the National Survey and other Governmental Sources; with Gazetteer and Notes on Pronunciation & Spelling (1st ed.)*. Nairobi: Printed by the Survey of Kenya.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2): 994-1028.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. (2005). "Very high resolution interpolated climate surfaces for global land areas." *International Journal of Climatology* 25: 1965-1978.

Hill, M. (1949). *Permanent way: The story of the Kenya and Uganda railway: Being the official history of the development of the transport system in Kenya and Uganda*. Nairobi: East African Railways and Harbours.

Kapila, N. K. (2009). *Race, Rail, and Society: Roots of Modern Kenya*. East African Educational Publishers.

MacDonald J. (1892a). *Uganda Railway Survey diaries*. RCMS 112/45, Oxford University Colonial Records Project.

MacDonald J. (1892b). *Journal of Sir James Macdonald*. RCMS 113/45, Oxford University Colonial Records Project.

Murdock, George. (1959). *Ethnographic Atlas*. Pittsburg, PA: University of Pittsburgh Press.

NOAA. (2012). *Average Visible, Stable Lights, & Cloud Free Coverage*. Version 4 DMSP-OLS Nighttime Lights Time Series. Silver Spring: National Oceanic and Atmospheric Administration

Ochieng, W.R, and R. M. Maxon. (1992). *Economic History of Kenya*. Nairobi, Kenya: East African Educational Publishers Ltd.

Ogendo, Reuben. (1967). "The Significance of Industrial Zoning to Rural Industrial Development in Kenya. A Study of the Facts and Methodology." *Cahiers d'Etudes Africaines* 7(27): 444-484.

Pringle J.W. (1892) *Uganda Railway Survey diaries*. RCMS 114/45, Oxford University Colonial Records Project.

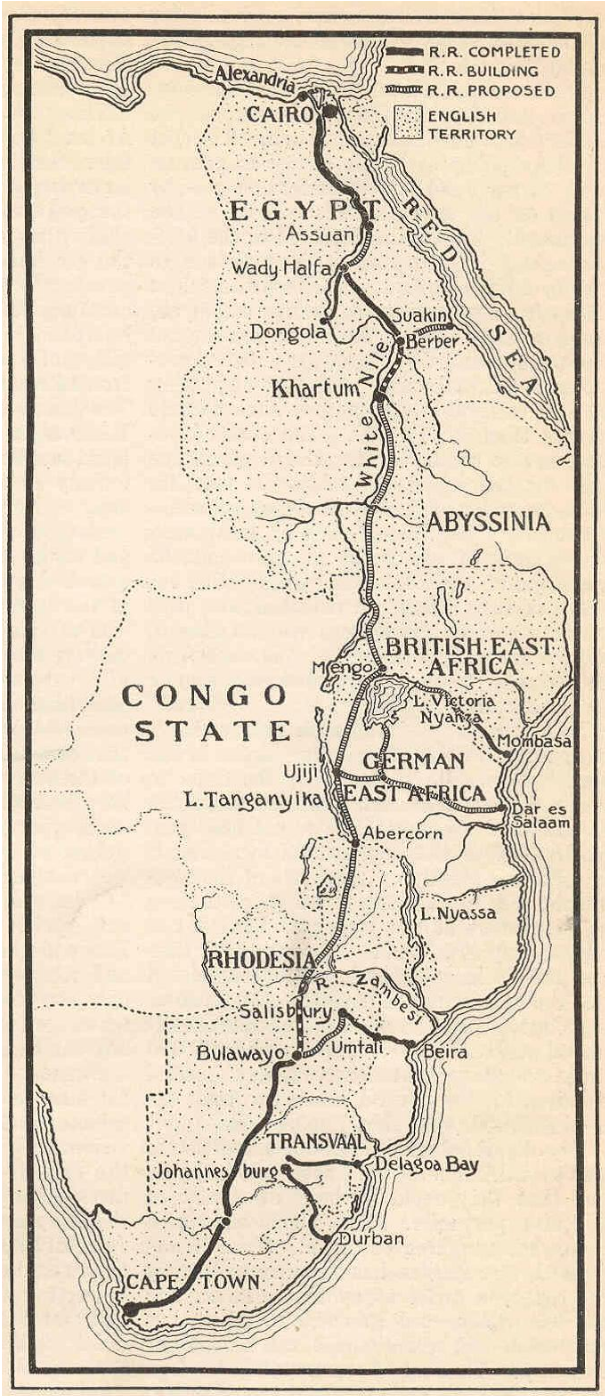
Soja, Edward W. (1968). *The Geography of Modernization in Kenya: A Spatial Analysis of Social, Economic, and Political Change*. Syracuse, NY: Syracuse University Press.

Thomson, J. B. (1896). *Joseph Thomson, African explorer: A biography*. London: S. Lowe, Marston and Co.

World Bank. 2013. *World Development Indicators*. Washington, DC: The World Bank.

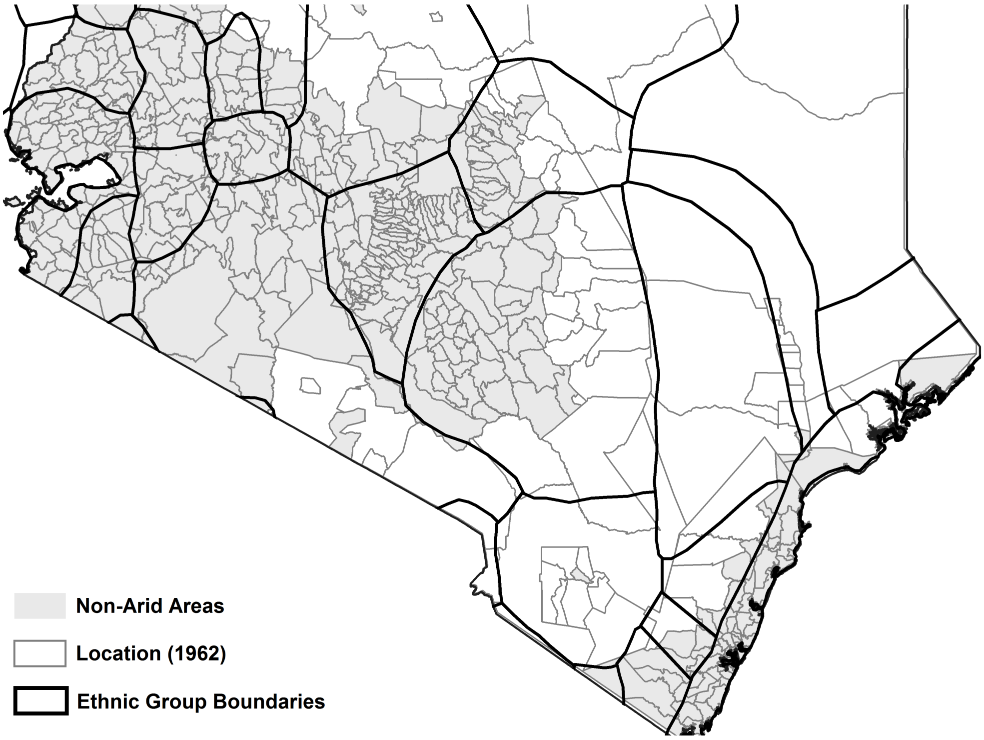
FOR ONLINE PUBLICATION: APPENDIX FIGURES AND TABLES

Web Appendix Figure 1: Map of the Cape to Cairo Railway (Early 1890s)



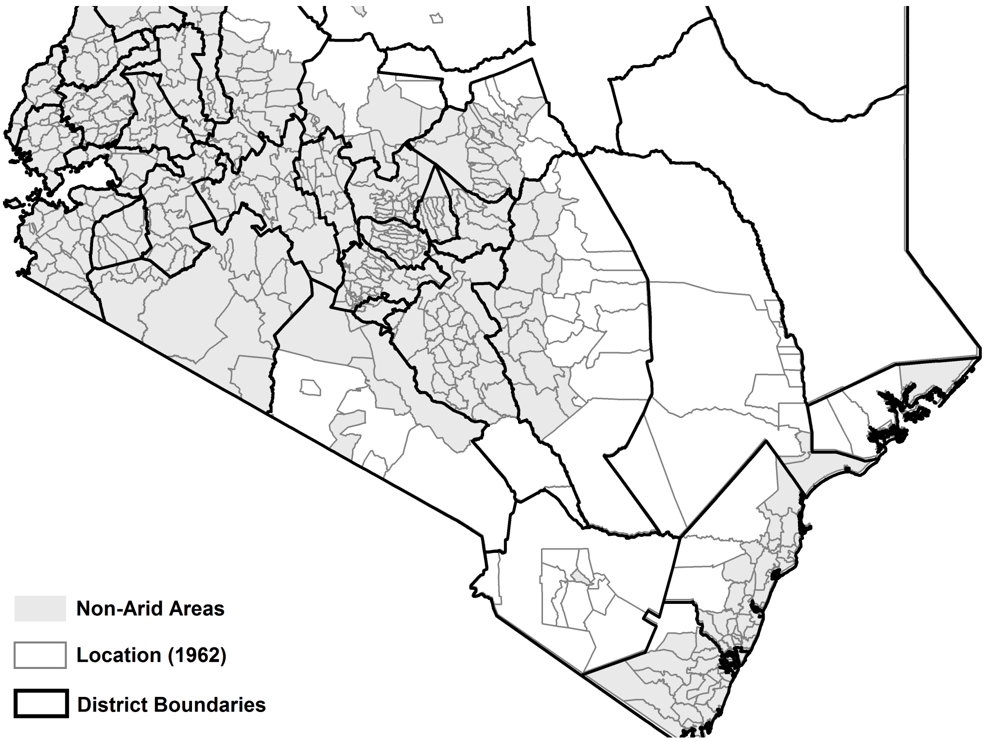
***Notes*: The map demonstrates the ambitions of the Cape to Cairo Railway, a project to unify all the British colonies of Northern, Eastern and Southern Africa by rail.** This grand scheme was the vision of Cecil John Rhodes (1853-1902), a British businessmen and mining magnate who turned his attentions to Southern African politics and imperialism. Kenya was part of "British East Africa" (the name of the East African Protectorate before 1895) in the map. The map shows that Kenya was merely a transit territory en route to the central east Africa. The source of this map is the website of the *Digital History Project*: http://www.digitalhistoryproject.com/2012/06/africa-building-cape-to-cairo-railroad.html.

Web Appendix Figure 2: Ethnic Group Boundaries



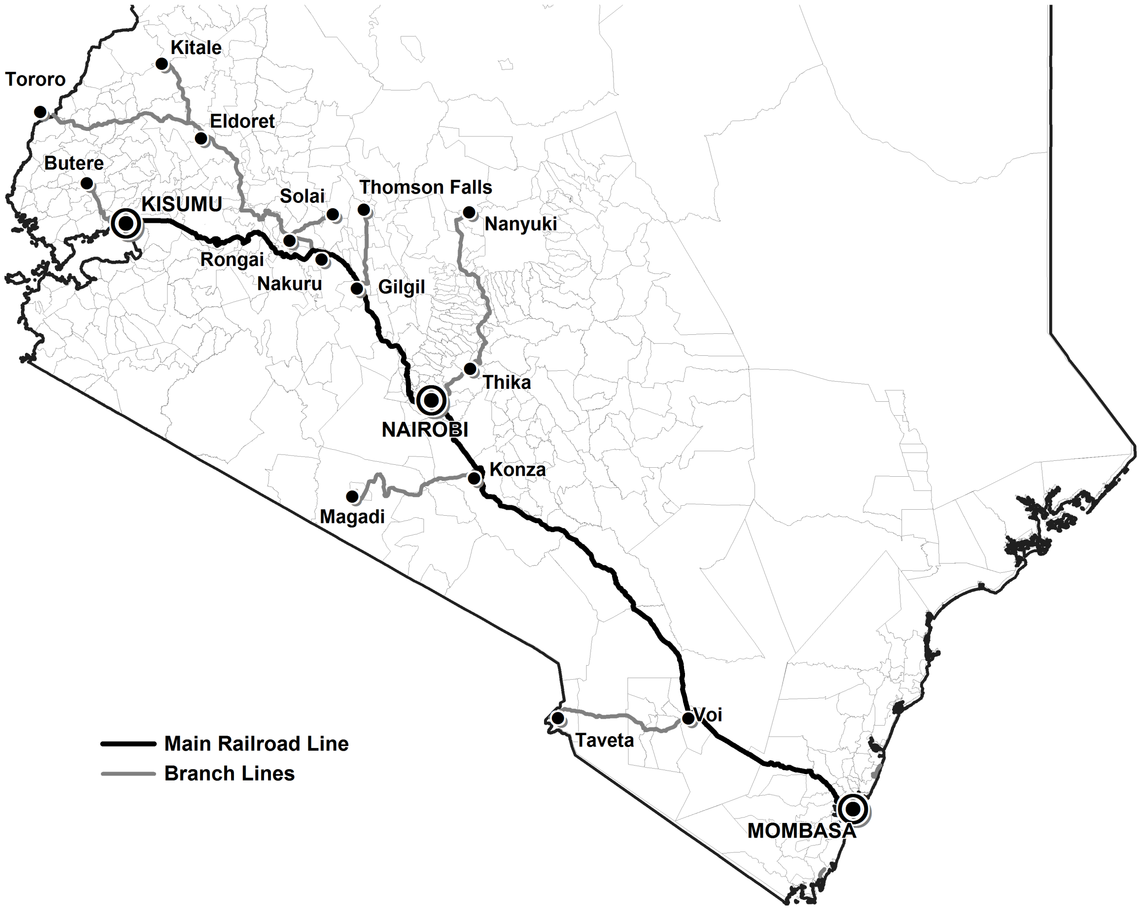
***Notes*: This map shows the boundaries for the 27 ethnic groups in Kenya.** The 403 non-arid locations belong to 21 ethnic groups. On average there are 19 locations by ethnic group. *See Web Appendix for data sources.*

Web Appendix Figure 3: District Boundaries



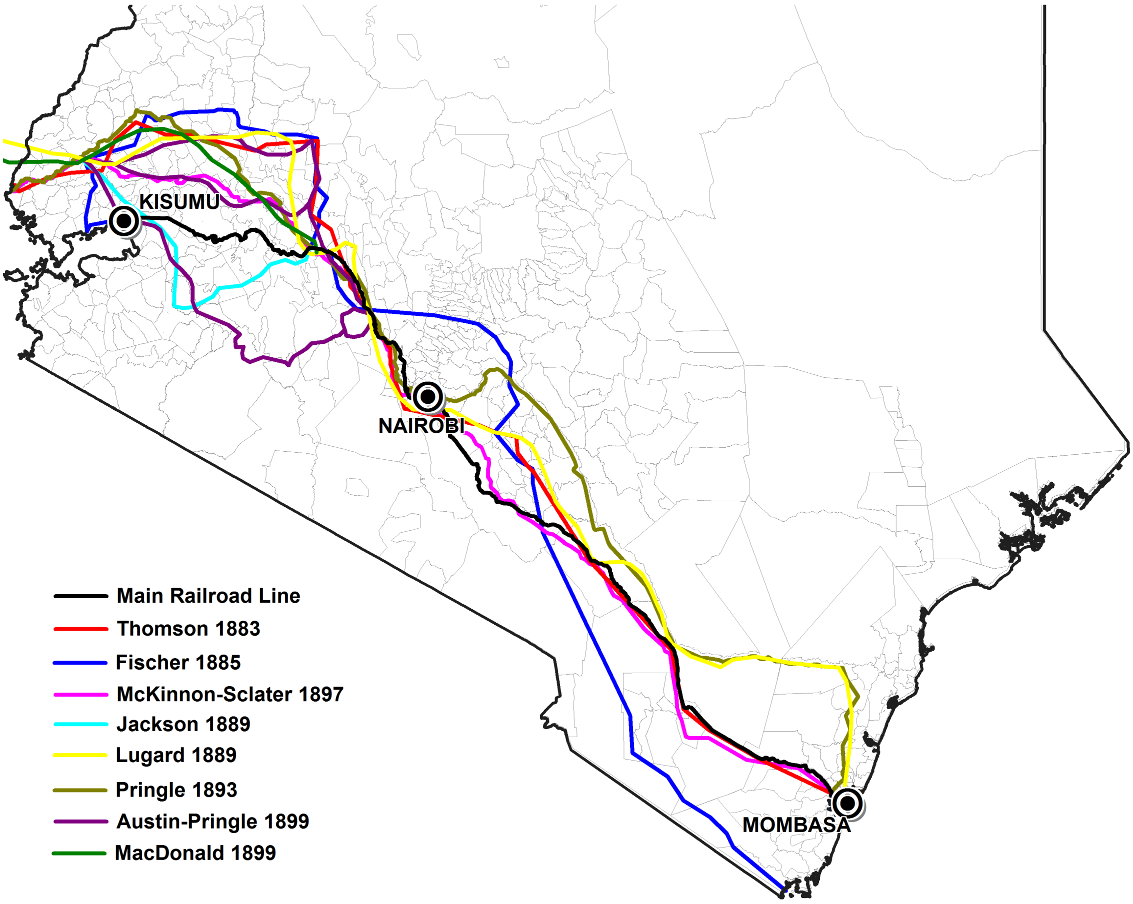
***Notes*: This map shows the boundaries for 35 districts (administrative boundaries of 1962) in Kenya.** The districts are the administrative unit below the "province" in Kenya, hence the third-level administrative unit. The 403 non-arid locations belong to 34 districts. There are thus 12 locations by district on average. *See Web Appendix for data sources.*

Web Appendix Figure 4: Main and Branch Railroad Lines



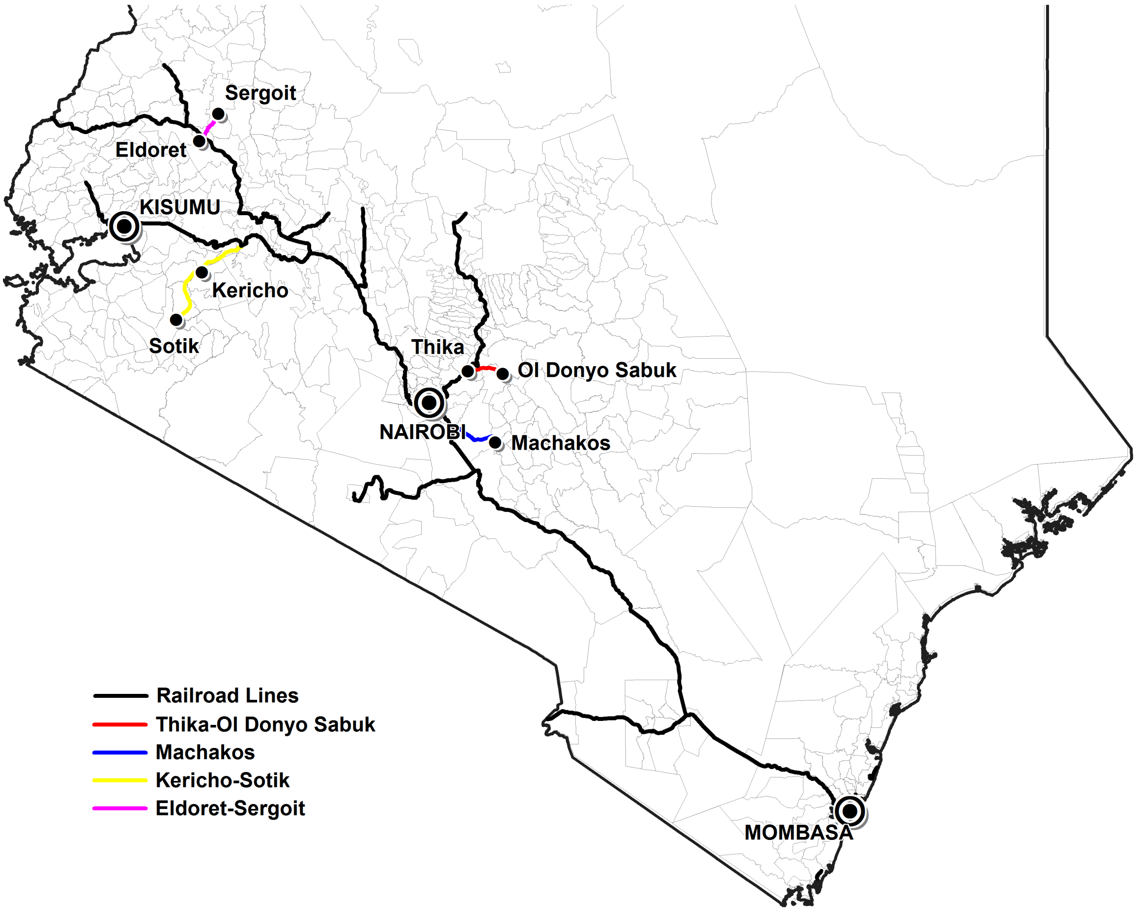
Notes: This figure shows the main line from Mombasa to Kisumu and the various branch lines that were built. **(A)** ***Main line from Mombasa to Kisumu:*** Imperial British East Africa Corporation (IBEAC) sought to build a railroad from the coast to Lake Victoria (Gunston, 2004). Suffering financial losses, the IBEAC ceded control of Uganda and Kenya to the British Government in 1896 and the Kenya-Uganda mainline was constructed between 1896-1901. The "Uganda Railway" was initially named after its original destination being Uganda and built for three principal reasons. Firstly, for strategic and geopolitical reasons. The line shielded the region against competing European powers, allowing transportation of troops. Lake Victoria is also the source of the Nile River, and the British thought that by linking Uganda to the coast they could unify all their colonies in Northern, Eastern and Southern Africa (Web Appendix Figure 1 shows the map of the "Cape to Cairo Railway", a plan to unify British Africa from south to north by rail). Secondly, Uganda was seen to hold vast wealth with further trade potential. Linking Lake Victoria to the coast would open up Uganda by reducing trade costs. Thirdly, it was deemed a "civilising" mission; bringing Christianity and the abolishment of slavery. **(B)** ***Branch lines.*** The Ugandan railroad established the general urban pattern of Kenya. The line produced its own nodes superseding the old caravan ones. Various branch lines were constructed between 1913 and 1930. No railroad was built post-1930. The following branch lines were constructed: ***(i) Nairobi-Thika (1913)***: The branch line was built to tap the fertile lands towards Mt Kenya. ***(ii) Konza-Magadi (1915)***: The branch line was built to serve the Soda Ash mines at Lake Magadi. ***(iii) Voi-Taveta (1918)***: Built linking the Northern areas of Tanzania; then included in German East Africa and captured by the Allies in 1916 during World War I. ***(iv) Nakuru-Eldoret (1926)***: Built linking European farmers who needed access to markets, especially those beyond Eldoret. ***(v) Eldoret-Kitale (1926)***: Planned as the main trunk line to Uganda, and ultimately the Congo. ***(vi) Rongai-Solai (1926)***: Built to access the Rift Valley and the agricultural lands which were dominated by coffee plantations. ***(vii) Eldoret-Tororo (1928)***: Built to connect the Kenyan Protectorate with the Ugandan Protectorate by railroad. Tororo was then a village on the Ugandan side of the border. As such, this branch line and not Eldoret-Kitale became the main trunk line to Uganda. ***(viii) Gilgil-Thomson Falls (1929)***: The branch line was built to Thomson Falls, the access point for the timber milling industry. ***(ix) Thika-Nanyuki (1930)***: The branch line was built as an extension to the Nanyuki branch line from Nairobi to Thika. ***(x) Kisumu-Butere (1930)***: The branch line was built to extend the mainline to Butere, connecting settler areas in Yala and Butere (maize and cattle) with the railroad system. *See Web Appendix for data sources.*

Web Appendix Figure 5: Main Railroad Line and Explorer Routes (from Mombasa to Lake Victoria)



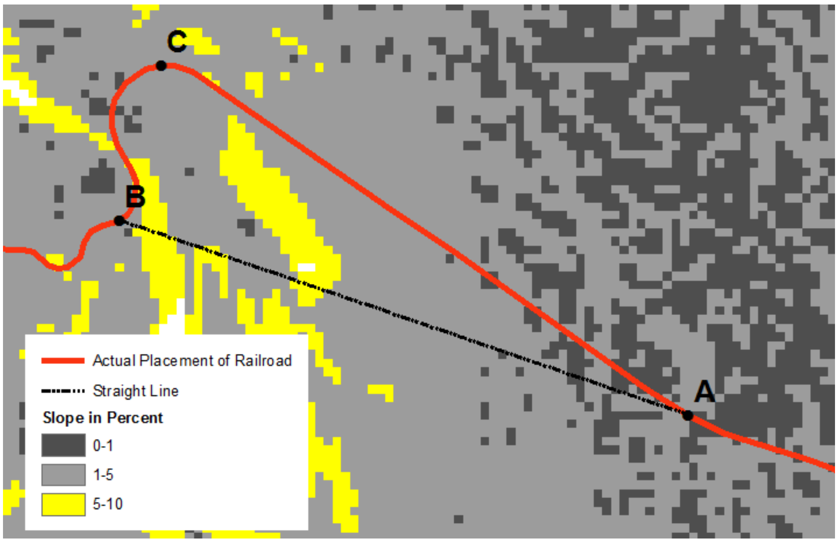
Notes*: This figure shows the various explorer routes we use in our analysis.** The explorer routes are taken from the map published in the *Government Survey of Kenya (1959)* digitising the routes in GIS. These explorer routes, from the coast to Lake Victoria provide a counterfactual for the mainline. First, various segments of these routes could have been alternatively selected to become a segment of the mainline. Second, the explorer routes traversed areas with better locational fundamentals. They often went through more densely populated areas as a result. Therefore, as we show in the main text, the economic potential of these "placebo lines" was better than for the mainline. The following explorer routes are the *"placebo lines"* we consider: ***(i) Thomson 1883: Thomson was leading a Royal Geographical Society. His travel report includes a map of his route (Thomson, 1884). Thomson joined an Arab caravan in Taveta to Njemps indicating that he was following a frequented trade route. ***(ii) Fischer 1885:*** Fischer's expedition started from Vanga, a small village on the coast and planned a route to Kampala (Seegers, 2008). We exclude his journey after reaching Kisumu, which was the railroad terminus. ***(iii) Jackson 1889:*** Jackson led an IBEAC expedition intended to open up the regions, mark out or establish trading stations, and to sign treaties with chiefs (Hill, 1950). Following the caravan route, he went via Machakos and the northern shore of Lake Naivasha. At Nakuru, the expedition deviated via Sotik. ***(iv) Lugard 1889:*** Lugard was engaged by the IBEAC to open up a trade route into the hinterland (Hill, 1950). He established Machakos as the first administrative capital of the Colony, then built a trading station at Dagoretti, close to Nairobi. He proceeded to Uganda, traveling via Naivasha to Nakuru hugging the Mau range, which he rightly thought was the shortest route to Lake Victoria and Uganda. Unable to penetrate dense forest, he instead followed the well-known caravan route via Baringo missing Guaso Ngishu which the railroad surveyors had found suitable. After Lake Baringo, he crossed the Kamasia range and the Uasin Gishu Plateau abandoning the route followed by Thomson, Fischer and Jackson, who had all circled around the north of the Nandi Hills and reaching Kavirondo by the valley of the Nzoia. ***(v) Pringle 1893:*** Led an expedition from Mombasa to Uganda. He first followed the usual caravan routes to Machakos going onto Naivasha via Nairobi. At Naivasha, Pringle hesitated between three main routes to Uganda (Pringle, 1893). He chose the one that went through more accessible territory. ***(vi) McKinnon Sclater 1897:*** Started in 1890 by the IBEAC, the McKinnon-Sclater road was an ox cart track from Mombasa to the Ugandan border, reaching Uganda in 1896, the same year rail construction began. The track was *"of the simplest kind, [...] the roughest track along which a bullock-cart would go"*(Smith, 1899). The road barely reduced trade costs, as the journey was slow and difficult. ***(vii) Austin & Pringle 1899:*** Both Royal Engineers, Captains Pringle and Austin surveyed the west of Kenya exploring three possible routes. ***(viii) MacDonald 1899:*** IBEAC promoted MacDonald to "Chief Engineer" of the railway survey leading an expedition from Nakuru to Uganda via Mumias, crossing the nile in 1899 after fighting in Uganda ceased. (MacDonald, 1899a &1899b). *See Web Appendix for data sources.*

Web Appendix Figure 6: Branch Railroad Lines and Branch Lines Planned but not Built



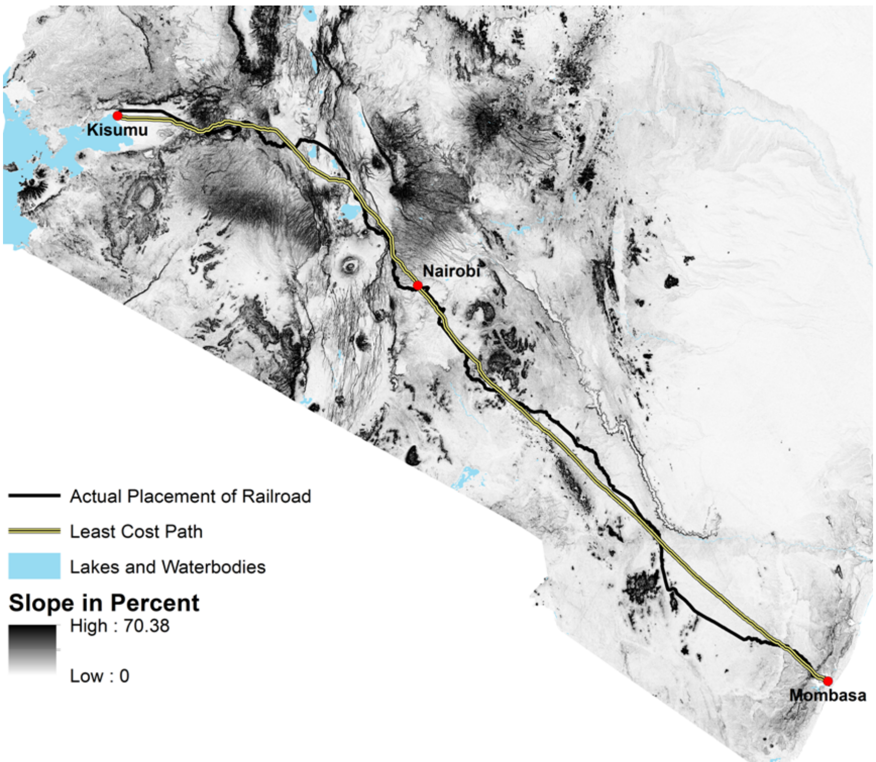
Notes*: This figure shows the various branch lines that were proposed in 1926 but not built.** We argue that the economic potential of the branch lines that were built and the branch lines that were planned but not built was very similar. Some of the proposed branch lines were dismissed because they did not have enough promise to be profitable. However all lines that were actually built were unprofitable, including in 1929 before the economic crisis *(Kenya and Uganda Railways and Harbours (1931))*. This would question the ability of the government to appraise the economic potential of various areas and profitability of specific lines at the time. These lines should thus provide a good counterfactual for the railroad lines. The following lines were the branch lines that were planned but not built *(see Colony and Protectorate of Kenya (1926) for a description of each line)*: ***(i) Eldoret-Sergoit 1926: The branch line was proposed to access the agricultural lands of the Nzoia River with crops of maize, wheat and the potential for cattle ranching and coffee cultivation. The report concluded that the line *"would ultimately pay and lead to a more intensive development of an area which is undoubtedly attractive"*. It never materialised. ***(ii) Machakos 1926:*** Lying 34 km southeast of the mainline, Machakos was established as the first capital of the Colony. Later relocation to Nairobi resulted in isolation of the European settler enclave, hence the need for a branch line. However the report notes that the proposed line would serve no extensive area of economic importance. ***(iii) Kericho-Sotik 1926:*** The branch line was motivated for its agricultural output (coffee, tea and cereals) and its potential for the processing of large quantities of bamboo for paper pulp. The line would also provide access to various African areas, which would boost trade between Europeans and the natives. Suffering isolation *"the distance from the railway (about 65 miles) being too great for crops of a bulky nature to be attempted."* Although these areas showed high potential the line was never built because expected construction costs were too high necessitating *"considerable [...] heavy earthworks, and probably viaducts."* ***(iv) Thika-Ol Donyo Sabuk 1926:*** Ol Donyo Sabuk lay within close proximity to the existing railway siding at Thika. The objective of the line was to open up more land for the cultivation of coffee. However, the report notes that the branch line was unlikely to have any effect, as it was too close to the existing branch line from Nairobi to Thika. *See Web Appendix for data sources.*

Web Appendix Figure 7: Trade-Off Between Length and Gradient for the Construction of the Least Cost Path



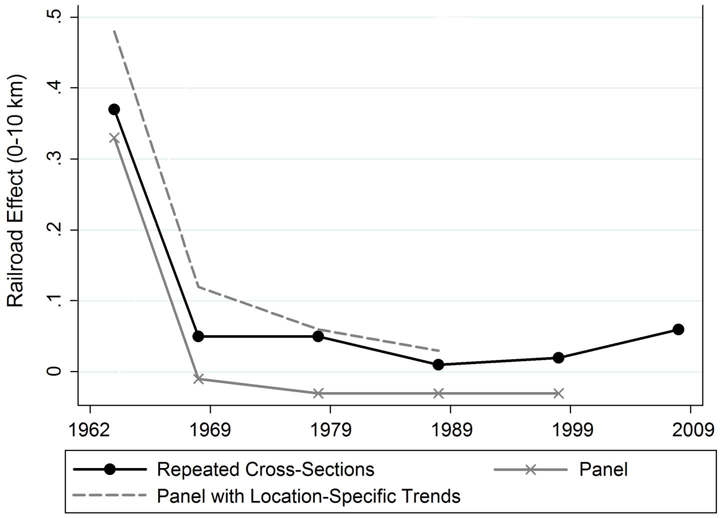
***Notes*: This map explains the derivation of rail construction costs.** We use the following sources to better understand the motivations behind the placement of the railroad and obtain various estimates of rail construction costs: *Economic and Technical Reports on the Colony and Protectorate of Kenya, Nairobi, 1921*, Hill (1949), Kapila (2009), MacDonald (1892a), MacDonald (1892b), Pringle (1892), Soja (168) and Thomson (1986). Original construction cost estimates stated that one mile of rail would cost £1,895 . This included works such as survey expenses, earth formation, clearing, permanent way and sidings, culverts and general charges. Light and heavy earthworks to overcome gradients below 5 and 10 percent would increase construction costs to £2,185 and £2,405 per mile respectively. However, these cost differences of the gradients are implausible. Firstly, they do not include construction costs of viaducts (£6,000 each) and bridges (£4,000 each) that we observe more frequently at steep gradients. Secondly, using these figures suggests that it was not cost effective to deviate from a straight line. This is at odds with the observed placement of the line. We used the following reasoning to arrive at plausible construction costs. If we observe the construction of line ABC instead of straight line AB, ABC must have been cheaper to build than AB despite of the longer distance. In the example shown above the observed line length between point A and B (via C) is 8.7 km. Line ABC fully avoided slopes of 5%-10%. In contrast, the straight line AB has only a length of 6.3 km, but would traverse 700 meters 5-10% slopes (ca. 7 yellow pixels, each 92 m). Hence, 8.7 km x cost (0-5%) <= (6.3-0.7) x cost (0-5%) + 0.7 x cost (5-10%). Setting 0-5% slope as standard 100, we conclude that traversing a 5-10% slope must be 4.5 times more expensive than traversing a 1%-5% slope. We repeated this exercise for a sample of five railroad sections. We arrived at the following cost parameters. Lakes are prohibitive and cannot be traversed; no gradient (0-1%) costs £106 per pixel; light gradients (1-5%) costs £123 per pixel and steep gradients (5-10%) costs £500 per pixel. Slopes higher than 10% were defined as barriers that cannot be crossed. This is based on the F-Class locomotive deployed at that time and which was restricted to a maximum gradient of 10%. We varied the cost of the 5-10% gradient by a factor of 10% and confirm that the shown least cost path is robust.

Web Appendix Figure 8: Least Cost Path



***Notes*: This map shows the least cost path between Mombasa, Nairobi and Kisumu, based on slopes and the presence of lakes and waterbodies.** Using the least cost path function in ArcGIS we performed a least cost path analysis, with Mombasa as origin, Kisumu as destination, with the condition of it crossing Nairobi. The parameters used to create the cost raster are detailed in Figure 1. *See Web Appendix for data sources.*

Web Appendix Figure 9: Effects of Railroads on Urban Population for Each Period, Cross-Sectional and Panel Regressions, 1901-2009



***Notes*: This graph investigates the dynamics of urban growth between 1901 and 2009. It shows the estimates of the railroad effect (for 0-10 km) when running: (i) repeated cross-sectional regressions, (ii) panel regressions with location and year fixed effects, and (iii) panel regressions with location and year fixed effects as well as location-specific linear trends.** We use panel data for 403 locations for the years = [1901, 1962, 1969, 1979, 1989, 1999, 2009]. (i) *Repeated Cross-Sections*: The graph shows the effects of repeated cross-sectional regressions where the dependent variable is urban population in year t and the variable of interest is a railroad dummy equal to one if the location is less that 10 km from a railroad line, while simultaneously controlling for urban population in year t-1. (ii) *Panel*: The graph shows the effects of panel regressions where the dependent variable is the change in the urban population between year t-1 and year t (the 1999-2009 railroad effect is omitted due to the fixed effects). We include location and year fixed effects. We use the change in the urban population as the dependent variable because we cannot add a lag of the dependent variable in panel regressions due the dynamic bias highlighted by Nickell (1981). The lag is thus not included in the regression. (iii) *Panel with Location-Specific Trends:* We run the same panel regression as model (ii) except we include location-specific linear trends (the 1989-1999 is now also omitted due to the trends). We only display the coefficients of the 10 km railroad dummy, as the other coefficients are nil. *See Web Appendix for data sources*.

WEB APPENDIX TABLE 1: COLONIAL RAILROADS AND POPULATION GROWTH, EFFECTS OF THE GEOGRAPHICAL CONTROLS, 1901-1962-2009

Dependent Variable:	Number of Inhabitants (z)			
	European 1962	**Urban 1962**	**Asian 1962**	**Urban 2009**
	Table 2 Col.(1)	*Table 2 Col.(2)*	*Table 2 Col.(4)*	*Table 5 Col.(1)*
	(1)	(2)	(3)	(4)
Rail Dummy, 0-10 km	0.78***	0.37***	0.49***	0.14
	(0.22)	(0.14)	(0.15)	(0.11)
Rail Dummy, 10-20 km	0.28***	0.09	0.06	-0.02
	(0.09)	(0.11)	(0.09)	(0.11)
Rail Dummy, 20-30 km	0.25*	0.04	0.08	-0.10
	(0.15)	(0.13)	(0.13)	(0.11)
Rail Dummy, 30-40 km	0.11	-0.14	-0.13	-0.02
	(0.10)	(0.11)	(0.13)	(0.08)
Physical Geography:				
"Coastal location" dummy	0.82**	0.14	0.11	0.09
	(0.35)	(0.17)	(0.16)	(0.19)
Distance to the coast (km)	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Area share of lakes (7%)	-3.33	-1.02	-2.45	-1.19
	(2.05)	(1.92)	(1.63)	(1.68)
Distance to a lake (km)	-0.00*	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Average annual rainfall (mm)	-0.00*	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Average annual temperature (degrees)	-0.09	-0.06	-0.01	0.02
	(0.13)	(0.13)	(0.10)	(0.12)
Altitude: mean (m)	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Altitude: standard deviation (m)	-0.00	-0.00**	-0.00**	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share of arid soils (%)	-4.44	0.43	-1.95	-3.51
	(4.57)	(2.53)	(1.90)	(2.34)
Share soils suitable for agriculture (%)	0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share soils suitable for coffee (%)	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share soils suitable for tea (%)	0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Area (sq km)	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Pre-Existing Settlement Patterns:				
Area share of "major settled groups" (%)	-0.00	-0.00*	-0.00*	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Area share of "pastoralists" (%))	0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
"Isolated groups" dummy	0.08	-0.08	-0.06	0.01
	(0.19)	(0.12)	(0.13)	(0.10)
"City in 1901" dummy	0.47	1.50**	0.48**	-0.39
	(0.98)	(0.64)	(0.22)	(0.56)
"Provincial capital" dummy	2.12	5.01*	4.28	-0.47
	(1.30)	(2.58)	(2.71)	(0.70)
Other Variables:				
European Population (z, 1962)				0.12**
				(0.06)
Asian Population (z, 1962)				-0.08
				(0.22)
Urban Population (z, 1962)				0.74***
				(0.27)
"General Business Area" Dummy				0.30
				(1.31)
Prob. > F (*Physical Geography Variables*)	0.05*	0.88	0.44	0.36
Province FE; Observations; R-squared	Y; 403; 0.20	Y; 403; 0.34	Y; 403; 0.28	Y; 403; 0.60

***Notes*: This table shows that the geographical controls have little effect on the respective growth of the European, urban and Asian populations in 1901-1962, and urban path dependence in 1962-2009.** OLS regressions using data on 403 non-arid locations for the year 1962. Robust SE's in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In column (1), the dep. var. is the z-score of the European pop. in 1962 (same regression as in column (1) of Table 2). In column (2), it is the z-score of the urban pop. in 1962 (same regression as in column (2) of Table 2). In column (3), it is the z-score of the Asian pop. in 1962 (same regression as in column (4) of Table 2). In column (4), it is the z-score of the urban pop. in 2009 (same regression as in column (1) of Table 5). All regressions include 8 province FE. *See Web Appendix for data sources.*

WEB APPENDIX TABLE 2: ALTERNATIVE IDENTIFICATION STRATEGIES, FOUR DUMMIES, 1901-1962

Strategy:	*Baseline Results* (1)	*Ethnic FE* (2)	*District FE* (3)	*Long. Lat.* (4)	*Nearest Neighbor* (5)	*Main Line* (6)	*Placebo w/o Rail* (7)	*Control: Placebo* (8)	*IV: LCP* (9)
Panel A:	*Dependent Variable*: **Number of European Inhabitants in 1962 (z)**								
Rail Dummy, 0-10 km	0.78***	0.97***	0.53**	0.74***	0.65**	0.85***	-0.01	0.71***	1.01
	(0.22)	(0.27)	(0.20)	(0.21)	(0.28)	(0.33)	(0.05)	(0.22)	(0.70)
Rail Dummy, 10-20 km	0.28***	0.47***	0.08	0.25**	0.26***	0.71**	-0.03	0.25**	0.73
	(0.09)	(0.13)	(0.10)	(0.12)	(0.07)	(0.31)	(0.09)	(0.12)	(0.75)
Rail Dummy, 20-30 km	0.25*	0.48***	0.25	0.23	0.40*	0.09	0.02	0.22	-0.06
	(0.15)	(0.18)	(0.16)	(0.17)	(0.22)	(0.09)	(0.05)	(0.14)	(0.42)
Rail Dummy, 30-40 km	0.11	0.36**	0.16	0.11	0.16*	0.15	0.07	0.20	0.33
	(0.10)	(0.14)	(0.12)	(0.11)	(0.09)	(0.12)	(0.10)	(0.12)	(0.22)
Panel B:	*Dependent Variable*: **Number of Urban Inhabitants in 1962 (z)**								
Rail Dummy, 0-10 km	0.37***	0.32**	0.41***	0.41***	0.28**	0.51*	0.04	0.41**	0.67
	(0.14)	(0.15)	(0.15)	(0.13)	(0.27)	(0.13)	(0.08)	(0.16)	(0.50)
Rail Dummy, 10-20 km	0.09	0.07	0.12	0.12	0.13	0.13	0.00	0.19*	0.03
	(0.11)	(0.11)	(0.13)	(0.13)	(0.10)	(0.11)	(0.07)	(0.12)	(0.33)
Rail Dummy, 20-30 km	0.04	0.12	0.09	0.09	0.18	-0.07	0.09	0.17	-0.32
	(0.13)	(0.13)	(0.14)	(0.14)	(0.15)	(0.09)	(0.12)	(0.14)	(0.29)
Rail Dummy, 30-40 km	-0.14	-0.07	-0.11	-0.11	0.06	-0.03	-0.04	0.04	0.25
	(0.11)	(0.10)	(0.12)	(0.12)	(0.02)	(0.20)	(0.06)	(0.13)	(0.34)
Panel C:	*Dependent Variable.*: **Number of Asian Inhabitants in 1962 (z)**								
Rail Dummy, 0-10 km	0.49***	0.63***	0.45***	0.49***	0.15**	0.57**	0.01	0.47**	0.92
	(0.15)	(0.20)	(0.16)	(0.16)	(0.06)	(0.26)	(0.05)	(0.19)	(0.56)
Rail Dummy, 10-20 km	0.06	0.19	0.04	0.05	0.06	0.25	0.02	0.11	0.00
	(0.09)	(0.13)	(0.11)	(0.11)	(0.05)	(0.18)	(0.04)	(0.12)	(0.36)
Rail Dummy, 20-30 km	0.08	0.28*	0.17	0.09	0.25	0.05	0.19	0.14	-0.32
	(0.13)	(0.16)	(0.17)	(0.15)	(0.20)	(0.09)	(0.20)	(0.13)	(0.34)
Rail Dummy, 30-40 km	-0.13	0.06	-0.05	-0.14	0.00	0.04	-0.02	0.01	0.23
	(0.13)	(0.09)	(0.12)	(0.15)	(0.04)	(0.19)	(0.05)	(0.15)	(0.35)
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

***Notes*: This table shows that results generally hold when we use the four rail dummies (0-10, 10-20, 20-30 and 30-40 km) with each identification strategy instead of one rail dummy only (0-30 km for European pop.; 0-10 km for urban and Asian pop.). The main exception is for the instrumental variables (IV) strategy (column (9)), as the railroad effects then become insignificant. The IV results of Table 2 should thus be taken with caution.** OLS regressions using data on 403 non-arid locations for the year 1962. Robust SE's in parentheses; * p<0.10, ** p<0.05, *** p<0.01. Column (1) replicates the main results of Table 2. For the sake of simplicity it employs only the 0-30 km (European pop.), 0-10 km (urban pop.) and 0-20 km (total pop.) rail dummies. In columns (2) and (3), we include 21 ethnic group fixed effects and 35 district fixed effects respectively. In column (4), we include first, second, third and fourth order polynomials of the longitude and latitude of the location's centroid. In column (5), we use nearest neighbour matching (without replacement; common support; trimming at 20%). In column (6), we show the effect for the main line, while controlling for the effect (30, 10 and 20 km respectively) of the branch lines. In column (7), we test that there are no effects when using the placebo lines as an alternative. We drop the locations less than X = 30 km (Panel A), 10 km (Panel B) and 20 km (Panel C) from a railroad line, in order to compare the placebo locations with the other control locations, while suppressing the effects from the railroad lines. In column (8), the placebo locations are the control group: First we use the locations less than 30 km (Panel A), then 10 km (Panel B) and finally 20 km (Panel C) from a placebo line. In column (9), we instrument the main line dummies (0-10; 10-20; 20-30; 30-40) by four dummies equal to one if the cell centroid is within 0-10, 10-20, 20-30 or 30-40 km from the least cost paths between Mombasa, Nairobi and Kisumu respectively, while also controlling for the percentages of area with a slope between 1 and 5%, between 5 and 10% and more than 10% respectively. The *Kleibergen-Paap rk Wald* F statistic is 2.4. We also control for the four branch line dummies (0-10; 10-20; 20-30; 30-40). *See Web Appendix for data sources.*

WEB APPENDIX TABLE 3: MATCHING ESTIMATORS, ROBUSTNESS, 1901-1962

Matching Estimator:	*Nearest Neighbor* *Baseline* (1)	*Replacement* (2)	*Radius (0.20)* (3)	*Radius (0.10)* (4)	*Radius (0.05)* (5)	*Kernel* (6)	*Local Linear* (7)
Panel A:	***Dependent Variable*: Number of European Inhabitants in 1962 (z)**						
Rail 0-30 km	0.39***	0.43***	0.41***	0.41***	0.42***	0.42***	0.42***
	(0.10)	(0.09)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)
Number of Obs.	357	357	357	357	357	357	357
Panel B:	***Dependent Variable*: Number of Urban Inhabitants in 1962 (z)**						
Rail 0-10 km	0.49*	0.42*	0.48*	0.48*	0.47*	0.47*	0.48*
	(0.26)	(0.23)	(0.26)	(0.27)	(0.26)	(0.27)	(0.27)
Number of Obs.	384	384	384	384	384	384	384
Panel C:	***Dependent Variable.*: Number of Asian Inhabitants in 1962 (z)**						
Rail 0-10 km	0.58**	0.55**	0.58**	0.58**	0.57**	0.57**	0.57**
	(0.28)	(0.25)	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)
Number of Obs.	384	384	384	384	384	384	384
Province FE	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y

***Notes*: This table shows that results are robust to using other matching estimators.** OLS regressions using data on 403 non-arid locations for the year 1962. Robust SE's in parentheses; * p<0.10, ** p<0.05, *** p<0.01. For the sake of simplicity it employs only the 0-30 km (European pop.), 0-10 km (urban pop.) and 0-10 km (Asian pop.) rail dummies. We always restrict the sample to the common support, trimming at 20% (dropping 20% of the treatment observations at which the propensity score density of the control observations is the lowest). Column (1) replicates the main results of Table 4, using nearest neighbour matching without replacement. Column (2): We use nearest neighbour matching with replacement. Columns (3)-(5): We use radius matching with the radius equal to 0.20, 0.10 and 0.05 respectively. Columns (6)-(7): We use kernel and local linear matching respectively. All regressions include 8 province FE. Table 1 lists the controls. *See Web Appendix for data sources.*

WEB APPENDIX TABLE 4: (NON-)EFFECTS FOR PLACEBO LINES, 1901-1962

Placebo Line:	Thomson 1883	Fischer 1885	Jackson 1889	Lugard 1889	Pringle 1893	McKinnon Sclater 1897	Austin Pringle 1899	Macdonald 1899
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:	***Dependent Variable*: Number of European Inhabitants in 1962 (z)**							
Placebo, 0-10 km	0.11	0.00	-0.11	0.40	-0.05	0.02	-0.02	0.61
Drop if Rail 0-10 km = 1	(0.10)	(0.08)	(0.11)	(0.26)	(0.04)	(0.08)	(0.17)	(0.61)
Panel B:	***Dependent Variable*: Number of Urban Inhabitants in 1962 (z)**							
Placebo, 0-10 km	0.26	-0.04	-0.03	0.15	0.06	0.19	-0.16	-0.15
Drop if Rail 0-10 km = 1	(0.19)	(0.09)	(0.11)	(0.11)	(0.06)	(0.18)	(0.10)	(0.11)
Panel C:	***Dependent Variable*: Number of Asian Inhabitants in 1962 (z)**							
Placebo, 0-10 km	0.11	-0.04	-0.14	0.04	-0.01	0.07	-0.06	-0.05
Drop if Rail 0-10 km = 1	(0.11)	(0.06)	(0.11)	(0.05)	(0.11)	(0.11)	(0.04)	(0.07)
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	

***Notes*: This table tests that there are no spurious effects for each of the eight explorer routes in 1901-1931. We drop the railroad cells in order to only compare the placebo cells and the other control cells.** OLS panel regressions using population data on 403 non-arid locations for the year 1962. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. This table tests that there are no spurious effects for eight placebo lines in 1962. In Panel A, the dependent variable is the z-score of the European population in 1962. In Panel B, the dependent variable is the z-score of the urban population in 1962. In Panel C, the dependent variable is the z-score of the Asian population in 1962. We drop the locations less than 10 km from a railroad line, in order to compare the placebo locations with the other control locations (N = 323), while suppressing the effects from the railroad lines. All regressions include province fixed effects (N = 8), and the same controls as in Table 2. The nodes are dropped from the analysis.*See Web Appendix for data sources.*

WEB APPENDIX TABLE 5: ROBUSTNESS AND SPECIFICATION CHECKS, 1901-1962

Panel A: **Dependent Variable = European Population (z) in 1962**							
	Baseline (0)	*Full Sample* (i)	*w/o Controls* (ii)	*Dist. to Nodes* (iii)	*Rail Stations* (iv)	*Control Roads* (v)	*Crossed by Rail* (vi)
Rail Dummy, 0-30 km (Except for (vi))	0.37*** (0.10)	0.41*** (0.11)	0.41*** (0.11)	0.29*** (0.10)	0.39*** (0.11)	0.31*** (0.11)	0.51*** (0.17)
	Dist. to Rail (vii)	*Log Instead of z* (viii)	*Conley SEs 50 km* (ix)	*City Dummy* (x)	*Urban Rate* (xi)	*5,000 Inh.* (xii)	*500 Inh. (1969)* (xiii)
Rail Dummy, 0-30 km (Except for (vii)-(viii))	-0.06*** (0.02)	1.25*** (0.22)	0.37*** (0.10)	–	–	–	–
Panel B: **Dependent Variable = Urban Population (z) in 1962**							
	Baseline (0)	*Full Sample* (i)	*w/o Controls* (ii)	*Dist. to Nodes* (iii)	*Rail Stations* (iv)	*Control Roads* (v)	*Crossed by Rail* (vi)
Rail Dummy, 0-10 km (Except for (vi))	0.34** (0.14)	0.37** (0.15)	0.52** (0.24)	0.34** (0.14)	0.55*** (0.19)	0.24** (0.12)	0.36** (0.15)
	Dist. to Rail (vii)	*Log Instead of z* (viii)	*Conley SEs 50 km* (ix)	*City Dummy* (x)	*Urban Rate* (xi)	*5,000 Inh.* (xii)	*500 Inh. (1969)* (xiii)
Rail Dummy, 0-10 km (Except for (vii)-(viii))	-0.06** (0.03)	0.67** (0.32)	0.34** (0.13)	0.28** (0.14)	3.31* (1.98)	0.35** (0.14)	0.40*** (0.13)
Panel C: **Dependent Variable = Asian Population (z) in 1962**							
	Baseline (0)	*Full Sample* (i)	*w/o Controls* (ii)	*Dist. to Nodes* (iii)	*Rail Stations* (iv)	*Control Roads* (v)	*Crossed by Rail* (vi)
Rail Dummy, 0-10 km (Except for (vi))	0.46*** (0.15)	0.48*** (0.16)	0.61** (0.01)	0.46*** (0.15)	0.60*** (0.19)	0.38*** (0.13)	0.45*** (0.16)
	Dist. to Rail (vii)	*Log Instead of z* (viii)	*Conley SEs 50 km* (ix)	*City Dummy* (x)	*Urban Rate* (xi)	*5,000 Inh.* (xii)	*500 Inh. (1969)* (xiii)
Rail Dummy, 0-10 km (Except for (vii)-(viii))	-0.07** (0.03)	1.58*** (0.32)	0.46*** (0.14)	–	–	–	–
Controls	Y	Y	N;Y	Y	Y	Y	Y

Notes*: This table shows that the results in 1901-1962 hold when performing various robustness and specification checks.** OLS regressions using data on 403 non-arid locations for the year 1962. Robust SE's in parentheses; * p<0.10, ** p<0.05, *** p<0.01. ***Column (i): We use the full sample of 473 locations. ***Column (ii):*** We drop the controls. ***Column (iii):*** We control for the Euclidean distance (km) to a node (Mombasa, Nairobi and Kisumu). ***Column (iv):*** The rail dummies are generated using the Euclidean distance to railroad stations (km) in the 1940s. ***Column (v):*** We include two dummies equal to one if the location is within 10 km from a paved road or an improved road in 1963. ***Column (vi):*** the variable of interest is a dummy equal to one if the location is crossed by the rail. ***Column (vii):*** The variables of interest are the Euclidean distance (km) of the location's centroid to the rail and its square (we do not show the effect of the square). ***Column (viii):*** The log of (European/urban/total population + 1) is the dependent variable. ***Column (ix):*** Standard errors are corrected for spatial autocorrelation using the approach of Conley (1999), with a distance cut-off of 100 km. ***Column (x):*** The dependent variable is the z-score of a dummy equal to one if the location has a city (a locality ≥ 2,000 inh.) in 1962. ***Column (xi):*** It is the urbanisation rate (%) of the location in 1962. ***Column (xii):*** It is the population that resides in localities whose population is in excess of 5,000 inh. in 1962. ***Column (xiii):*** It is the population that resides in localities whose population is in excess of 500 inh. in 1969. All regressions include 8 province fixed effects and the controls listed in Table 1. *See Web Appendix for data sources.*

WEB APPENDIX TABLE 6: RAILROADS, NON-AFRICAN SETTLEMENT AND DEVELOPMENT, 1901-2009

Dependent Variable:	**Urban Pop. 2009 (z)**		**Urban Pop. 2009 (z)**		**Urbanization Rate (%) 1999**	
	(1)	(2)	(3)	(4)	(5)	(6)
Rail Dummy, 0-10 km	0.14	-0.01	0.12	0.08	1.55	0.58
	(0.11)	(0.13)	(0.11)	(0.11)	(2.63)	(2.66)
N.Europeans (z, 1962)	0.05	0.30**			2.81	2.91
	(0.18)	(0.14)			(2.25)	(2.57)
N.Asians (z, 1962)	0.05	0.36			-2.64	0.79
	(0.22)	(0.22)			(2.07)	(3.10)
Urban Pop. (z, 1962)	0.75***	0.72***	0.70***	0.86***		
	(0.27)	(0.20)	(0.18)	(0.22)		
Urban Pop. x Rail 0-10 km		0.12		-0.02		
		(0.44)		(0.36)		
Urban Pop. x N.Europeans		-0.00				
		(0.13)				
Urban Pop. x N.Asians		-0.15				
		(0.12)				
Rail 0-10 km x N.Europeans	0.08	-0.17				
	(0.19)	(0.16)				
Rail 0-10 km x N.Asians	-0.20	-0.36				
	(0.24)	(0.47)				
Urban Pop. x Rail 0-10 km x N.Europeans		0.18				
		(0.14)				
Urban Pop. x Rail 0-10 km x N.Asians		-0.12				
		(0.24)				
Share Europeans (%, 1962)			0.08*	0.09*		
			(0.04)	(0.05)		
Share Asians (%, 1962)			0.02	0.04		
			(0.04)	(0.04)		
Urban Pop. x Share Europeans (%, 1962)				-0.03		
				(0.05)		
Urban Pop. x Share Asians (%, 1962)				-0.01		
				(0.02)		
Urbanization Rate (%, 1962)					0.99***	0.97***
					(0.07)	(0.06)
Urbanization Rate x Rail 0-10 km						0.18
						(0.18)
Urbanization Rate x N.Europeans						-0.06
						(0.05)
Urbanization Rate x N.Asians						-0.06
						(0.04)
Province FE, Controls	Y	Y	Y	Y	Y	Y
Observations	403	403	403	403	403	403
R-squared	0.60	0.65	0.60	0.63	0.53	0.54

***Notes*: This table shows that the results of Table 4 are robust to: Columns (1)-(2): Studying the interacted shocks of the demise of the railroads and the exoduses of the Europeans and Asians; Columns (3)-(4): Using the respective shares of Europeans and Asians in the location in 1962 rather than the z-scores for the same year; and Columns (5)-(6): Using the urbanization rate of each location in 1962 and 1999 rather than the z-scores of the urban population in 1962 and 2009.** OLS regressions using data on 403 non-arid locations for the year 2009. Robust SE's in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$. In columns (1)-(4), the dependent variable is the z-score of the urban population in 2009. In columns (5)-(6), the dependent variable is the urbanization rate (%) of the location in 1999. We use the year 1999 instead of the year 2009 because we do not have data on total population (only for urban population) for that year. Share Europeans (%, 1962) and Share Asians (%, 1962) are the respective shares of the European population and the Asian population in the total population of each location in 1962. All regressions include 8 province fixed effects and the controls listed in Table 1. *See Web Appendix for data sources.*

WEB APPENDIX TABLE 7: COLONIAL RAILROADS AND HISTORICAL FACTORS, 1901-1962

Dependent Variable:	Number Secondary Schools	Number Hospitals	Number Health Clinics	Number Police Stations	Number Post Offices	Paved Road Dummy	Industrial Center Dummy
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Panel A: ***Railroads and historical factors (1962)***							
Rail Dummy, 0-10 km	0.34***	0.12*	0.17*	0.38**	0.70***	0.55***	0.25***
Rail Dummy, 10-20 km	0.10	0.05	0.07	0.50***	0.32***	0.28***	0.10**
Rail Dummy, 20-30 km	0.05	0.09	0.06	0.26	0.22	0.12**	0.04
Rail Dummy, 30-40 km	0.12	0.07	-0.01	0.06	0.13	0.11*	-0.00
Panel B: ***Railroads and historical factors (1962), conditioned on historical population (1962)***							
Rail Dummy, 0-10 km	0.27***	0.06	0.14*	0.26	0.62***	0.53***	0.20***
Rail Dummy, 10-20 km	0.07	0.03	0.04	0.41***	0.28**	0.27***	0.07
Rail Dummy, 20-30 km	0.04	0.08	0.05	0.26	0.22	0.12**	0.03
Rail Dummy, 30-40 km	0.15	0.09	-0.01	0.10	0.16	0.12*	0.02
Panel C: ***Railroads and historical factors (1962), also conditioned on European settlement (1962)***							
Rail Dummy, 0-10 km	0.17**	-0.04	0.09	-0.03	0.38**	0.51***	0.15**
Rail Dummy, 10-20 km	-0.01	-0.06	0.00	0.17	0.08	0.25***	0.03
Rail Dummy, 20-30 km	-0.01	0.03	0.03	0.09	0.08	0.11*	0.00
Rail Dummy, 30-40 km	0.09	0.04	-0.03	-0.06	0.03	0.10	-0.01
Mean of the Variable	0.18	0.18	0.18	0.56	0.46	0.28	0.12
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y

Notes**: This table shows that the railroad cells are better endowed in infrastructure at independence (Panel A). This is explained by the higher population densities in the cells (Panel B), and in particular the larger European population in the cells (Panel C).** OLS regressions using data on 403 non-arid locations for the year 1962. In the interest of space, robust SE's are not reported; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the railroad dummies on seven measures of historical factors in 1962. As the measures were close to 0 in 1901, the cross-sectional regressions could be interpreted as long-differenced estimations for the period 1901-1962. In column (iv), we take into account all the police stations: provincial police headquarters, divisional police headquarters, local police stations, and local police posts. In column (v), we take into account all the post offices: local post offices, and local postal agencies. In Panel A, we regress each measure on the rail dummies. In column (vii), we take into account all the industrial centers: important industrial towns, minor industrial towns, important industrial centers, minor industrial centers, and potential industrial centers. In Panel B, we control for log(total pop. + 1) and log(urban pop. + 1) in 1962. In Panel C, we also include log(European pop. + 1) in 1962. All regressions include 8 province fixed effects and the controls listed in Table 1. *See Web Appendix for data sources.*

WEB APPENDIX TABLE 8: ADDITIONAL RESULTS ON THE CHANNELS OF PATH DEPENDENCE

Dependent Variable:	**Urban Population in 2009 (z)**			
Channel:	*Baseline*	*(9) +Infrastructure 1962*	*(10) +Crops & Indu. 1962*	*(11) +Market Pot. (6 Hours) 1962*
Sunk Investments	(9)	(10)	(11)	(i)
Urban Pop. (z, 1962)	0.74***	0.63**	0.47***	0.48***
	(0.27)	(0.25)	(0.18)	(0.18)

Notes**: Col.(i): This column shows that market potential does not reduce the effect when using three market potential measures based on the sums of total pop., urban pop. and European crop cultivations for locations within 6 hours by road from the location (excluding own location).** OLS regressions using data on 403 non-arid locations for the year 2009. Robust SE's in parentheses; * p<0.10, ** p<0.05, *** p<0.01. The specification is the same as in column (9) of Table 6. It shows how the coefficient of Urban Pop. (Z, 1962) varies as we include various controls. *Market pot. 1962 (6 hours):* Three market potential measures based on the sums of total pop., urban pop. and European crop cultivations for locations within 6 hours by road from the location (excluding own location). *See Web Appendix for data sources.*

WEB APPENDIX TABLE 9: COLONIAL RAILROADS AND CONTEMPORARY FACTORS, 2000s

Dependent Variable for the Year 2007:	**Number Primary Schools**	**Number Secondary Schools**	**Number Hospitals**	**Number Health Clinics**	**Number Health Dispensaries**
	(1)	(2)	(3)	(4)	(5)
Panel A: ***Railroads and contemporary factors (2000s)***					
Rail Dummy, 0-10 km	22.05***	7.25***	0.34**	1.22***	2.26**
Rail Dummy, 10-20 km	8.25	4.54**	-0.07	0.82***	0.65
Rail Dummy, 20-30 km	-0.13	0.31	-0.05	0.62	0.58
Rail Dummy, 30-40 km	-6.22	-1.69	-0.22	0.27	-1.37
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999-2009)***					
Rail Dummy, 0-10 km	0.67	2.09	0.15	0.56	-0.11
Rail Dummy, 10-20 km	-9.04*	0.34	-0.19*	0.25	-1.31*
Rail Dummy, 20-30 km	-8.51	-1.61	-0.10	0.28	-0.42
Rail Dummy, 30-40 km	3.70	0.80	-0.15	0.46	-0.49
Mean of the Variable	66.5	14.5	0.60	1.87	7.20
Dependent Variable: Roads: Year 2002 Education: Year 2009	**Paved Road Dummy**	**Improved Road Dummy**	**Completed Primary Education (%)**	**Completed Secondary Education (%)**	**Completed Tertiary Education (%)**
	(6)	(7)	(8)	(9)	(10)
Panel A: ***Railroads and contemporary factors (2000s)***					
Rail Dummy, 0-10 km	0.47***	-0.09	4.91***	7.00***	2.23***
Rail Dummy, 10-20 km	0.28***	-0.15**	4.06***	3.84***	1.12***
Rail Dummy, 20-30 km	0.25***	-0.07	3.82***	3.53***	1.01***
Rail Dummy, 30-40 km	0.09	-0.05	1.74**	1.77**	0.44
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999-2009)***					
Rail Dummy, 0-10 km	0.39***	-0.10	4.86***	6.72***	2.17***
Rail Dummy, 10-20 km	0.22***	-0.14*	4.06***	3.65***	1.08***
Rail Dummy, 20-30 km	0.20***	-0.06	3.82***	3.43***	0.98***
Rail Dummy, 30-40 km	0.10	-0.01	1.70**	1.84**	0.46
Panel C: ***Also conditioned on the school supply variables (1962 and 2000s)***					
Rail Dummy, 0-10 km	_	_	4.16***	5.49***	1.77***
Rail Dummy, 10-20 km	_	_	3.51***	2.91***	0.91***
Rail Dummy, 20-30 km	_	_	3.70***	3.21***	0.95***
Rail Dummy, 30-40 km			1.14*	0.90	0.13
Mean of the Variable	0.68	0.42	22.8	14.5	5.0
Province FE, Controls	Y	Y	Y	Y	Y

Notes**: Col.(1)-(7): These show that the railroad cells are better endowed in infrastructure today (Panel A), which is explained by the higher population densities in the cells (Panel B). Col.(8)-(10): The railroad cells are also better endowed in human capital today (Panel A). However, this is only partially explained by the higher population densities (Panel B) and the larger supply of schools in those cells (Panel C). This suggests that there is spatial sorting in human capital, and that the railroad cells disproportionately attract skilled workers.** OLS regressions using data on 403 non-arid locations for the 2000s. In the interest of space, robust SE's are not reported; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the rail dummies on fifteen measures of contemporary factors in the 2000s. As these measures were close to 0 for the year 1901, these cross-sectional regressions should be interpreted as long-differenced estimations for the period 1901-2000. In Panel A, we regress each measure on the rail dummies. In Panel B, we control for log total population (+1) in 1999. We also control for the log of the urban population (+1) in 2009 as we do not have data on total population in 2009. Columns (8)-(10): The primary, secondary and tertiary completion rates are estimated for adults aged 25 years or older (using data at the constituency level, so the SE's are clustered using the main constituency in each location). All regressions include province fixed effects and the same controls as in Table 1. *See Web Appendix for data sources.*