Out of Africa: Colonial Railroads, White Settlement and Path Dependence in Kenya*

Remi Jedwab[†] and Edward Kerby[‡] and Alexander Moradi[§]

October 2013

Abstract: Little is known about the extent and forces of path dependence in developing countries. We examine whether locational fundamentals (i.e., geography) or increasing returns (i.e., historical shocks) are the main determinants for the distribution of economic activity across space. Since new investments in transportation infrastructure constitute "shocks to economic geography", the construction of the colonial railroad provides a natural experiment. Using fine spatial data for Kenya over one century, we find that colonial railroads had a strong causal impact on White settlement, commercial agriculture and urban growth before independence. Railroads fell into disrepair after independence, yet these patterns of economic geography persisted. While colonial sunk investments (i.e., schools, hospitals and roads) partly contributed to path dependence, railroad cities mainly persisted because their early emergence served as a mechanism to coordinate contemporary investments in the subsequent period. Our findings are important as they inform regional economic policy in shaping future economic progress.

Keywords: Path Dependence; Transportation Technology; Colonialism
JEL classification: R4; R1; O1; O3; N97

---

*We are grateful to Richard Hornbeck, Stephen Smith, Anthony Yezer and seminar audiences at the EHA annual meeting, George Washington and Sussex for helpful comments.

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

[‡]Edward Kerby, Department of Economic History, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK (email: e.m.kerby@lse.ac.uk).

[§]Alexander Moradi, Department of Economics, University of Sussex, Jubilee Building, Falmer, BN1 9SL, UK (email: A.Moradi@sussex.ac.uk). Also affiliated with the Center for the Study of African Economies, University of Oxford, Department of Economics, Manor Road, Oxford, OX1 3UQ.

From Mombasa is the starting-point of one of the most wonderful railways in the world [...] a sure, swift road along which the white man and all that he brings with him, for good or ill, may penetrate into the heart of Africa as easily and safely as he may travel from London to Vienna. [...] The British art of "muddling through" is here seen in one of its finest expositions. Through everything - through the forests, through the ravines, through troops of marauding lions, through famine, through war, through five years of excoriating Parliamentary debate, muddled and marched the railway.

Winston Churchill, *My African Journey*, 1908

1. INTRODUCTION

The literature has argued over whether locational fundamentals (i.e., the geographical endowments of various locations) or increasing returns (i.e., localized historical shocks) are the main determinants for the distribution of economic activity across space. If locational fundamentals are the main determinants of spatial patterns, any negative or positive shock will have only temporary effects. For example, Davis & Weinstein (2002, 2008), Bosker et al. (2007) and Miguel & Roland (2011) use the war time bombing of Japan, Germany and Vietnam respectively, to test whether these shocks had long-term effects on the relative ranking of cities that were disproportionately destroyed. As explained by Davis & Weinstein (2002, 2008), the fact that cities like Hiroshima and Nagasaki recovered their population and industries in the long run strongly argues in favor of the first theory. There is only one spatial equilibrium, and regional policy interventions, no matter how large they are, cannot undo it in favor of another equilibrium. Yet the literature has also shown how large (or in some cases, even small) localized shocks can have permanent effects (Bosker et al., 2007, 2008; Bleakley & Lin, 2012; Dell, 2012). These studies demonstrate strong evidence in favor of the emergence of multiple spatial equilibria, which implies that regional policy interventions may impact regional development.

In parallel to the economic geography literature, the development literature has debated whether the underdevelopment of various parts of the world (e.g., Sub-Saharan Africa) is better explained by physical geography or history, and institutions in particular. Several studies have shown that natural geographic features have a strong influence on long-term development, in both developed and developing countries (Gallup, Sachs & Mellinger, 1999; Rappaport & Sachs, 2003; Sachs, 2003). Conversely, various articles have shown how institutional shocks may shape the patterns of development, for a given physical geography (Acemoglu, Johnson & Robinson, 2001, 2002; Banerjee & Iyer, 2005; Nunn, 2008; Dell, 2010). Obviously, the effects of a specific physical geography will depend on the technologies used in the economy, and technological progress may itself be conditioned by the quality of institutions. Therefore, the spatial patterns of development result from the interaction of physical geography, technological progress and institutions.

While war bombings constitute negative shocks (destructions) to economic geography, investments in transportation infrastructure(constructions) constitute positive shocks to economic geography. If the locational fundamentals theory is accurate, we should not observe any effect of new investments in transportation infrastructure, given the past infrastructure. On the contrary, if the increasing returns theory is right, any infrastructure investment should have significant effects. Fujita & Mori (1996), Behrens (2007), Redding, Sturm & Wolf (2011), Bleakley & Lin (2012) and Jedwab & Moradi (2013) have all shown how such investments may have permanent effects. Obviously, the impact of a new investment will depend on the previously used transportation technologies and infrastructure. The less efficient the old network is relative to the new one, the larger this impact will be.

This raises several questions. First, the impact of a historical shock is likely to depend on the context in which this shock takes place. To what extent is the shock large enough to durably affect the location of economic activity? Little is known about the extent of path dependence in the specific context of developing countries. These countries are often poorer, agrarian and more rural. Locational fundamentals could play a much larger role in these cases. Spatial patterns are stable, due to the continued effects of time-invariant geographical endowments (i.e., serial correlation). Conversely, these countries are less developed, so any spatial shock, such as a new investment in transportation infrastructure, would have a larger effect. This impact would be permanent if there are increasing returns in production. Spatial patterns are stable, due to path dependence. Second, the mechanisms of path dependence may also differ from what is observed in developed countries. As explained by Bleakley & Lin (2012), path dependence could be due to *sunk investments* or the *coordination problem* of contemporary factors. Given that fixed costs are a source of increasing returns, *sunk investments* (e.g., houses, schools, hospitals and roads) could account for urban persistence. Given returns-to-scale in production, factors need to be co-located in the same locations. There is a *coordination problem* as it is not obvious which locations should have the contemporary factors. Sunk investments may be less important in developing countries than in developed countries, as the materials used to build houses, schools, hospitals and roads are of relatively poorer quality. However, there are fewer sunk investments per capita in poorer countries, so that the locations that benefited from an early start may enjoy their initial advantage for a longer period than in developed countries. The coordination problem could also be less important, if the usual forces of agglomeration (e.g., industrial agglomeration effects or human capital accumulation) are more limited in these countries. At the same time, the fact that these countries are poorer imply that many coordination failures have not been solved yet. Therefore, the extent and forces of path dependence could differ in developing countries, which has implications for regional policy. Unfortunately, these countries are not only poor, but they also suffer from data shortage. This limits our understanding of the relationship between historical shocks and the location of economic activity.

To address these difficulties, we use a natural experiment and a new data set on railroads and city growth at a fine spatial level over one century in Kenya. All of Kenyan railroad lines were built before independence. Although profitable, railroads collapsed in the post-independence period. Yet they had long-term effects on urban and development patterns. In this paper, we document that colonial railroads shaped the economic geography of the country, and use this setting to show how and why historical shocks can have permanent effects in developing countries. The Kenyan setting is attractive for four reasons.

First, as in the rest of Africa, transportation costs in Kenya were extremely high one century ago. Kenya lacked navigable waterways. Draft animals were not used due to the Tsetse fly transmitting trypanosomiasis. There were only a few tracks, that did not become roads before the 1930s. Kenya only exported goods that were headloaded on short distances, or slaves who walked longer distances. Economic change was limited to the coast. The hinterland was poor, and devoid of cities. The British colonizer sought to build ways of penetration to the hinterland, mostly to ensure military domination. The British built the main line from Mombasa, on the coast, to Lake Victoria, in the west, in 1896-1901. Although the line crossed the country from east to west, the ultimate objective was to link the coast to Uganda (across Lake Victoria), where the Nile originates. The railway was even called the "Uganda Railway", although it was not technically located in Uganda. The British were then able to travel from Cairo to the eastern coast of Africa, which was a clear asset at the time of the scramble for Africa. The line (and the branch lines that were subsequently built) actually went through areas that were lowly-populated, for historical reasons. We show that the decrease in internal trade costs had a strong (and mainly unexpected) effect on the settlement of British farmers, who established cities from where they could manage their coffee farms and specialize in urban production activities. African population also increased along the lines, in the surroundings of the cities, as coffee cultivation required unskilled labor.

Second, since we follow a simple difference-in-difference strategy whereby we compare connected and non-connected locations of around 20x20km over time, we provide evidence that the placement of the lines and the timing of rail building were exogenous to the population and coffee booms. Even if placement was not exogenous, Kenya's history gives us various identification strategies we exploit to confirm our effects are causal. We find no effects for a set of placebo lines that consist of various explorer routes and branch lines that were planned but not built. Besides, as we have data at a fine spatial level, we can include many district fixed effects, ethnic group fixed effects, or even a quartic polynomial in longitude and latitude to identify our effects by comparing connected locations with only contiguous locations that were not connected. These strategies all give similar results.

Third, Kenya's rail sector declined post-independence, due to mismanagement and lack of maintenance in the rail sector, as well as massive investments in roads. Besides, most of the White farmers and workers left the areas and were leaving for the capital city, Nairobi, or the United Kingdom, after independence. Lastly, coffee production has significantly declined in the old producing areas due low international prices and the exhaustion of coffee soils after several decades of cultivation. As a result, locations along the railroad lines have lost their initial advantage in terms of transportation, human capital (the White settlers) and commercial agriculture. Yet today, locations along the railroad lines remain relatively more urban and economically developed today.

Fourth, we were able to create a new panel data set at a fine spatial level over one century (473 locations of more or less 20x20km, in 1901-2009). In particular, we use new data on rail building, cities, population densities, and economic change more generally. As argued by Lucas (1988), cities are the main engines of growth. Urbanization is our primary measure of development, in line with the literature (Acemoglu, Johnson & Robinson, 2002; Dittmar, 2011). We have also collected large amounts of data on colonial and post-colonial infrastructure. We can thus study the mechanisms by which railroad cities persisted post-1963.

What we find is striking. In the colonial period (1901-1963), the fall in trade costs attracted many White farmers and workers to the locations along the railroad lines. They created coffee farms where they hired African labor. They also established cities that served as trading stations for coffee production, and administrative and commercial centers. Therefore, population increased and cities emerged along the lines. At independence, locations along the railroad lines were more economically developed, and they still were in 1999. This effect is not explained by changes in transportation technology, as measured by roads today, but is due to path dependence. Indeed, the effects of colonial railroads on long-run development are explained by colonial urbanization. While colonial sunk investments matter, we show that railroad cities mostly persisted because their early emergence served as a mechanism to coordinate contemporary investments in the subsequent period.

In addition to the literature on path dependence, our findings advance the literature on transportation technology and growth. Trade costs were extremely high in Africa one century ago. They are still very high today (Atkin & Donaldson, 2012). Research has suggested that the lack of intra and intercontinental trade integration is a determining factor of African underdevelopment (Rodrik, 1998; Johnson, Ostry & Subramanian, 2007). In this regard, poor infrastructure is often mentioned as the main obstacle to trade expansion in Africa (Radelet & Sachs, 1998; Buys, Deichmann & Wheeler, 2010). Transportation infrastructure can indeed facilitate the circulation of goods, people and ideas. Railroads have boosted exports in Kenya, in line with the literature on transportation and trade (Michaels, 2008; Duranton & Turner, 2012; Donaldson, 2013; Faber, 2013; Donaldson & Hornbeck, 2013). They have encouraged the movement of workers and firms, in line with the literature on transportation and population and employment growth (Baum-Snow, 2007; Atack et al., 2010; Banerjee, Duflo & Qian, 2012; Ghani, Goswami & Kerr, 2012). They have promoted the diffusion of innovations, here the establishment of cities and the adoption of a new crop, in line with the literature on information and communication technology and development (Jensen, 2007; Aker, 2010; Dittmar, 2011). Additionally, they can trigger an equilibrium in which cities emerge to facilitate the accumulation of factors, and thus have long-term effects on economic growth (Bleakley & Lin, 2012; Jedwab & Moradi, 2013). Few papers have studied the causal impact of transportation in Africa, with the exception of Storeygard (2012) who finds that trade costs negatively affects remote regions. Burgess et al. (2013*a*) show how road building is driven by political considerations (ethnic favoritism) instead of economic considerations, which lowers the returns to such investments.[1]

Our focus on colonial transportation is also associated with the literature on the impact of colonization on development. We innovate in three ways. First, the literature has mostly focused on the impact of colonial institutions (Acemoglu, Johnson & Robinson, 2001, 2002; Banerjee & Iyer, 2005; Feyrer & Sacerdote, 2009; Dell, 2010; Iyer, 2010; Bruhn & Gallego, 2012), while the effects of colonial investments have been overlooked. Second, the few studies that examined colonial

[1]Other studies on Africa have relied on cross-country regressions (Buys, Deichmann & Wheeler, 2010), or panel data models using GMM and simulated regressions using cross-sectional data for one country (Dercon et al., 2008; Jacoby & Minten, 2009). Our methodology differs in that our natural experiment gives us an exogenous variation in transportation infrastructure.

investments highlighted the role of human capital (Glaeser et al., 2004; Huillery, 2009). However, the effects of investments in transportation infrastructure may have been as large (or even larger). In Kenya, the European farmers and workers – the owners of (human) capital during the colonial period – initially settled along the railroad lines. When they left after independence, cities persisted despite the loss of human capital, as railroads served as a mechanism to coordinate contemporary investments (e.g., in human capital) after independence. This shows how transportation infrastructure may have shaped the accumulation of human capital, and not vice-versa. Lastly, we use African panel data at a fine spatial level over almost one century.[2]

The paper is organized as follows. Section 2 presents the historical background of rail building and economic development in Kenya and the data used. Section 3 explains the methodology and shows the results for the colonial period. Section 4 studies why these effects persist over time. Section 5 discusses the nature of path dependence in poor countries and what the implications are for public policy.

2. RAILROADS AND PATH DEPENDENCE: BACKGROUND AND DATA

We discuss the historical background and the data we use in our analysis. The Online Data Appendix contains more details on how we construct the data.

2.1 New Data on Kenya, 1900-2009

We compiled a new data set. The data set is built on the 1962 Population Census - Kenya's first proper Census that also includes the African population. The units of observation are 478 so called 'locations' – third level administrative units, with an average size of 20x20 km (Figure 1). Using GIS we merged population data from the 1979, 1989 and 1999 Population Censuses.[3] We used various documents to recreate the history of rail construction. We know when each line was finished and opened. We also located lines that were planned but not built. For each line, we created dummies equal to one if the Euclidean distance of the cell centroid to the line is 0-10, 10-20, 20-30, 30-40 or 40-50 km. Our main analysis focuses on the rail network in 1930. We proceed similarly to merge the GIS road database from Burgess et al. (2013*a*). Settlement patterns shortly before the coming of the railroad 1901 are digitised from Soja (1968). Physical geography variables were taken from a several sources (see online appendix for details). The location of police stations, health centres, secondary schools, and post offices in 1960 were extracted and digitised from the *Atlas of Kenya* (Survey of Kenya, 1959). From the same source we obtained explorer routes. Point locations of primary and secondary schools, as well as health centres were accessed through the open government portal.

[2] We also contribute to the literature on the historical roots of African underdevelopment: Nunn (2008), Nunn & Wantchekon (2011), Michalopoulos & Papaioannou (2011), Nunn & Puga (2012), Michalopoulos & Papaioannou (2012) and Heldring & Robinson (2012).

[3] 1979-1999 geo-referenced Census locations are accessible online through the open government portal at https://opendata.go.ke/.

2.2 Railroad Lines Built

Infrastructure investments are typically endogenous, driven by the economic potential that justifies them. Hence, a simple comparison of connected and non-connected locations is likely to overstate the output created by the railroad. Railroad construction in Kenya provides us with a natural experiment to identify the effect of reduced trade costs on development.

The British established the East African Protectorate in, what is now Kenya. The introduction of the railroad hailed in the modern transportation age for Kenya, which previously had been a transit territory for Ugandan trade (Soja, 1968). As such, railroad construction to Uganda via Kenya provides us with a natural experiment to identify the causal effect of transportation infrastructure on economic change and path dependence. This summary draws on Hill (1950), King & van Zwanenberg (1975) and Otte & Neilson (2006). High transport costs had hindered long distance trade. Lake Victoria, some 450 miles inland from the Indian Ocean, was the only navigable waterway. Draft animals had not been adopted due to the trypanosomiasis transmitting Tsetse fly. Headloading was the main, and very costly means of transport before the introduction of railways. Only high value goods such as ivory were carried through the hinterland and slaves that could be walked. In ca. 1880, the western caravan route through German Tanganyika from Zanzibar to Lake Victoria was a longer distance than the eastern route via the British protectorate to Mombasa. The former was more popular, with merchant safety a concern on the eastern route traversing the tribal Masai territory (Memon, 1973).[4] Railroads catalyzed a transportation revolution.

Main line: The railroad was initially named after its original destination being Uganda. First surveys for a route were conducted by the Imperial British East Africa Corporation (IBEAC), a Royal Charter Company tasked with the administration of the British protectorate in East Africa. Suffering financial losses, the IBEAC ceded control over Uganda to the British Government in 1896 and the Kenya-Uganda mainline was constructed between 1896-1901. The railroad was built for three principal reasons. Firstly, for strategic and geopolitical reasons. The line shielded the Upper Nile valley against encroaching European powers, specifically the German push into Uganda. This was was perceived as a threat to British interests, control over Egypt and trade through the Suez (Otte & Neilson, 2006). Secondly, the Uganda protectorate was seen to hold vast wealth with further trade potential. Linking Kisumu at Lake Victoria to Mombasa at the coast would open up Uganda by reducing trade costs and improving safety for merchants. Thirdly, it had a deemed civilizing mission, bringing Christianity and the abolishment of slavery (Wrigley, 1959). The construction was debated fiercely within the British parliament. Critics doubted the usefulness of the railroad "from nowhere to utterly nowhere".

Branch lines: The Ugandan railroad established the general urban pattern of Kenya. The line produced its own nodes superseding the old caravan ones. Soja (1968) explains that the equal distribution of urban centers at key points along the main route reflects the weak influence of local economic factors in the initial urban growth. The

[4]Beachey (1967) explains that by 1885 as many as 2,000 porters would travel in a single caravan for safety. "The ivory caravans developed a life of their own, and the supply of their needs led to a system somewhat similar to that of ship chandlering"

interior nodes increased in size and importance as feeder lines were constructed, connecting them with major European farming areas and smaller administrative centers. Administration of the protectorate moved from Machakos to Nairobi, which was established because of its location on the last stretch of level ground before the Highlands. Nairobi came to dominate the entire country as a result of its railroad function. From the Mombasa, the railroad proceeded west with equidistant sidings approximately every 100 to 120 miles. The Mogadi branch tapping the lake's soda reserves was built in 1915 while the Highland branch lines were built between 1913 and 1930 with a distance of 50 miles was maintained. The branch line north from Nairobi to Thika was built during 1913 and extended to Nanyuki via Fort Hall and Nyeri in 1930. Further extensions from Naivasha to Thomson Falls followed in 1929, while the Nakuru, Eldoret and Kitale branches were completed in 1926. The final western branch to Butere followed in 1930. The branch lines are potentially endogenous. An explicit appraisal of the economic potential was done (Colony and Protectorate of Kenya, 1926). In practice, assessment was difficult because of vested interests of the European community and the inherent uncertainty of such investments. In 1930 all branch lines made a loss (Kenya and Uganda Railways and Harbours, 1931). We address endogeneity concerns by comparing the branch and main lines. If the branch lines were endogenous, we would find stronger effects for the branch lines, which we do not. Transportation costs decreased massively. Figures cited in Hill (1950) imply a 1902 freight rate of 11 shillings per ton mile for head porterage as compared to 0.09 shillings per ton mile by railroad on the Mombasa-Kisumu route. The caravan route from Zanzibar to Lake Victoria German road was ca. 3.5s per ton mile (Memon, 1973).[5]

Placebo lines: To address endogeneity concerns, we use a placebo strategy. The economic potential of the placebo lines was similar (if not better). They therefore qualify as a good coutnerfactual. First, we use explorer routes as placebos. The routes followed caravan routes from the Coast to Lake Victoria. Several factors influenced the course of the routes such as safety, distance, provision of water and food. Second, we use branch lines that were proposed in 1922 but not built (Colony and Protectorate of Kenya, 1926). The British East Africa Railway Commission considered six branch line extensions between 1921 and 1924. The viability of the extensions were presented to the Protectorate Legislative Council in 1926. The branch lines included (1) an extension of the mainline from Nakuru to Sergot, (2) Kericho to Sotik, (3) Thika to El Donyo Sabuk, (4) Machakos to a mainline junction approximately 30 mile before Nairobi, (5) Nyeri to Nayuki, and finally (6) Gilgil to Thomson Falls. The latter two lines, (5) and (6), were constructed in 1930. Despite vested colonial interests, the former extensions failed to materialize for either economic viability or cost in construction.

2.3 Economic development

There were only two cities with a population in excess of 5,000 in 1901: Mombasa and Lamu. This increased to 19 and 111 in 1962 and 1999 respectively. During

[5]Kenya's head porterage rates were higher than in other African countries, e.g. Ghana (5s ton mile), Nigeria and Sierra Leone: 2.5s ton mile, whereas rail freight rates were comparatively lower (0.80s, 0.19s and 0.27s per ton mile in Ghana, Nigeria and Sierra Leone ca. 1910 respectively) (Chaves, Engerman & Robinson, 2012; Jedwab & Moradi, 2013).

the nineteenth century the main form of production was small-scale subsistence agriculture. Raw materials, for example ivory were the primary exports until the introduction of the railways. The peculiarities of railroad placement led to the curious situation that the railroad traversed a sparsely settled territory with no freight to transport. European settlement was encouraged to generate economic development, justifying the railroad infrastructure and geopolitical imperative of “effective occupation”. Moreover, through the introduction of an agricultural export industry, the railroads could be made to be profitable. By 1915, 8,242 square miles of land had been alienated on behalf of about one thousand white settlers. Agriculture and specifically the establishment of export crops was the engine of Kenya’s development (King & van Zwanenberg, 1975).

Coffee had become the premier export to Europe. Despite its proximity to Ethiopia, thought to be the origin of coffee, the crop was only introduced by missionaries in 1893. It took 30 years before coffee was widely grown by Europeans, making Kenya a large exporter as early as 1930. Kenya developed three major coffee-growing regions in the high plateaus around Mt. Kenya, Kisii, Nyanza, Bungoma, Nakuru, Kericho and the Rwenzori Mountains and Aberdare Range. These areas were restricted to European settlers on a 999 year lease. The region came to be known as the White Highlands, and European or white immigration increased (Morgan, 1963). A scheme to attract new, ’ex-soldier settlers’ after World War One was devised. The scheme partially failed as many of the allocated farms were uncultivable or lacking in water. Nevertheless there was a steady stream of new settlers during the 1920’s. By the 1930’s, European farmers made up only a small proportion of the white population; the majority of Europeans were employed in a range of services which emerged to support the settlers. Demand for African labour grew as they supplied the dominant portion of agricultural labour while discriminatory practices against Africans prevented farmland ownership in the Highlands. “Native Reserves” grew in close proximity to the large European landowners. The 1934 Kenya Land Commission endorsed the segregated system. African migrant peasants “squatted” on European farms in exchange for land tenure through a system akin to “corvee labour”. Until 1940, with only a few exceptions, government services and overseas loans were directed towards the settlers. Even in the early 1950’s, Africans were almost entirely kept out of commercial agriculture in Kenya. Data on European agricultural production is available on an annual basis from 1923 early agricultural Censuses recorded. The first Agricultural Census that included Africans took place in 1960-61. It might be noted that the only railway built to serve an area of African settlement, that from Kisumu to Butere, has had no greater effect on the location of economic activities than those in Uganda (O’Connor, 1964, 1965).

In the early 1930’s prices of the major export products of coffee, sisal and maize, were on the rise. Despite the inflow of more settlers combined with further price increases, resulted in half of the total settlers dependent on maize. Export prices dropped drastically in 1930’s and most of these small-scale maize farmers incurred heavy debts. The entire structure of white farming very nearly collapsed during this period. A desperate attempt through agencies of the colonial government, aid amounting to £1,000,000 was injected to rescue white agriculture. King & van Zwanenberg (1975) note that it is quite probable white settlement would not have survived without aid. They explain that up to 1940 the settlers had farmed inefficiently, lacking the technology to utilize the land fully. However WWII created conditions necessary for the modernization of the settler agriculture. Under wartime conditions, the introduction of marketing boards offered white farmers fixed prices for volume purchases of export crops. The farmers were assured of both profits and sales, preventing small scale African agriculture from competing against the European in terms of efficiency. After Kenya's independence changes were introduced in the White Highlands. However large scale landholders merely changed from European to African. Only 5% of the White Highlands was taken over by small-scale farmers with most of the coffee, tea and sisal plantations remaining intact.

3. RAILROADS AND DEVELOPMENT AT INDEPENDENCE

In this section we show that railroads led to economic change during the colonial period. We test if connected locations experienced population and urban growth. We explain the various strategies we implement to obtain causal effects.

3.1 Main econometric specification

The main hypothesis we test is whether rail connectivity drove population growth during the colonial period. We follow a simple strategy where we compare the European, urban and total populations of connected and non-connected locations (l) in 1962, one year before independence:

$$zPop_{l,1962} = \alpha + Rail_{l,1962}\beta + \omega_p + X_l\zeta + \nu_{l,1962} \quad (1)$$

where $zPop$ is the standard score of European / urban / total population of location l. $Rail_{c,t}$ are dummies capturing rail connectivity in 1962 (but all the lines were built by 1931): being 0-10, 10-20, 20-30, 30-40 or 40-50 km away from a line. The dummies are equal to zero in 1901. We expect rail connectivity to have a positive and significant effect on population ($\beta > 0$). We include province fixed effects ω_p and a set of controls X_l to account for pre-existing settlement patterns. We have a cross-section of 473 locations. However, in the case of European and urban population, which were nil in 1901, results should be interpreted as long-differenced estimations for the period 1901-1962.[6] Our main analysis is performed on a sample of non-arid locations excluding the areas are unsuitable for agriculture and mainly inhabited by pastoralists. We believe the restricted sample of 403 non-arid locations gives more conservative estimates. Results are robust to removing the restrictions. We express all population numbers in z-scores. This let us abstract from natural growth and facilitates cross-comparisons. Moreover, we add the location's area (sq km) on the right hand side as one of the controls. This let us conveniently interpret effects as densities. Infrastructure investments are typically endogenous: places with highest return will be connected. We argued in section 2.2 that placement of railroads was not endogenous to economic potential and population. We now describe our strategies to ensure our estimates are causal.

[6] A dummy variable controls for cities that existed at the coast in 1901 are. See next subsection.

3.2 Exogeneity assumptions, controls, and identification

Firstly, we include controls to account for potentially contaminating factors. We account for pre-railroad transportation costs proxied by a “coastal location” dummy and the distance to the coast (km). We control for population and economic patterns that prevailed around 1890 by including the area of “major settled groups” (sq km), the area of “pastoralists” (sq km), an “area of isolated hunter and gatherer groups” dummy and a "city in 1901” dummy. Additionally, we add physical geography variables that shaped pre-existing population patterns and also determined the potential for European settlement such as the average altitude (m), annual rainfall (mm) between 1950 and 2000 and the share of arid soils (%). A “White Highlands” dummy directly accounts for the alienated agricultural lands that was reserved for settlers of European origin. We control for political economy by including a “provincial capital in 1962” dummy. Mombasa, Nairobi and Kisumu are important nodes that we dropped from the analysis - even though the latter two cities were often attributed to the railroad (Soja, 1968; Kapila, 2009).

Secondly, we follow a placebo strategy. Our placebo lines were routes from Mombasa to Lake Victoria used by European explorers and the caravan trade (Figure 3). These routes transversed known, relatively safe areas and populations that engaged in market exchange.[7] We also use branch lines proposed in the 1920s but never built, either because they were deemed unprofitable - as unprofitable as actual branch lines later turned out to be, or because the Great Depression dried up government investments. Table 1 shows the comparison of treatment and control groups in terms of observables.[8] Compared to the full control group, treated locations have a lower share of arid soils, nevertheless though higher rainfall. They are less likely to go through areas where African sedentary populations settled and are more likely to be in the White Highlands. This confirms that railroads went through a sparsely settled territory - the historical buffer zone between Kenya’s major ethnic groups - hardly the most likely candidate for agricultural and city growth. Differences between treatment and placebo locations follow similar lines.

Thirdly, endogeneity concerns particularly apply to the placement of branch lines. Branch lines were built to support European agriculture. They were built later when an urban system had emerged. We thus compare branch and main lines. Endogeneity is not a concern if we find similar effects for both.

Fourthly, since we have data at the third administrative unit level, which on average are only 20x20 km in size, we compare neighboring locations that are unlikely to differ in terms of unobservables. Locations less than 50 km from the lines are all similar in terms of observables.[9] If the placement is truly exogenous, the effect should strongly decrease as we move away from the line, which is exactly what we find. As an additional test we also include district fixed effects (N=41). Identifica-

[7] The railroad crossed a sparsely populated zone where several ethnic groups met. These later became the "White Highlands". We would expect catch-up of locations in this buffer zone if only because of imperial peace and population growth. Note that we control for “White Highlands”, so identification comes from locations within the White highlands.

[8] See Appendix Table 1 for tests that there are no spurious effects for each of the seven placebo lines in 1962.

[9] We regress each control on the rail dummies using the 40-50 km cells as the omitted group. There are no significant differences.

tion comes from variation between very similar locations within districts.

Fifthly, we apply a spatial regression discontinuity design mirroring the work of Dell (2010) and Michalopoulos & Papaioannou (2012). We include fourth-order polynomial of the longitude and latitude of the location's centroid (Lee & Lemieux, 2010). Hence, *Rail* identifies any sudden discontinuities from a very flexible spatial pattern such as that existing in pre-railroad age Kenya.

3.3 Main results

Table 2 shows the main results. We find a strong effect of rail connectivity on European population, but this effect decreases as we move away from the rail line and is zero after 10 km (column 1). Urban population follows the same pattern (column 2). The rail effect on urban population strongly decreases and becomes non-significant when including European population highlighting the settlement of Europeans just along the railroad lines as the main driver of urbanization (column 3). There is a strong effect of rail on population density up to 20 km (columns 4 & 5). This is consistent with increased farming opportunities along the railroad lines attracting African labour.[10] Indeed, there were only 50,000 Europeans at independence. The increase in population along the railroad lines must have been due to Africans moving to these areas.

As discussed in section 2.3, the railroad lines permitted and encouraged the settlement of Europeans along the railroad lines. Many of them did specialize in the cultivation of coffee and other cash crops for European markets. Rather unfortunately, we do not have good data at a fine spatial level on the production of coffee during the colonial period. First, we use district panel data on coffee production from 1922-32 to see if the construction of various branch lines during that period has a positive effect of the cultivation of coffee. Districts are the spatial unit above locations. We run panel regressions by including district and year fixed effects, as well as province-year fixed effects to account for time-variant heterogeneity at the province level. This means that we compare connected districts with neighboring non-connected districts of the same province over time. No matter the specification, we find that being connected increases the total area under coffee cultivation by about 1,000 acres (given a mean of 3,700 acres). We also use data on European coffee cultivation at the location level in 1962. We do not have data on African coffee production. We run the same regression as model (1), except the dependent variable is the location's share of agricultural land under coffee cultivation. We find a strong effect until 10 km (1.6**, given a standard deviation of 5.6). This confirms that many Europeans settled along the railroad lines to establish coffee plantations, for which they necessarily had to hire African labor.

3.4 Alternative identification strategies and robustness

Results of our identification strategies are reported in Table 3. Column (1) replicates the main results of Table 2. For the sake of simplicity, we use 0-10 and 0-20 km railroad dummies, because effects beyond were not significant. In column (2), we

[10] Migration was not always welfare enhancing, because of the multiple discriminatory interventions in the African labour market. For example, a high hut tax combined with wages fixed at low levels increased labour supply (Mosley, 1983; Deininger & Binswanger, 1995).

test that there are no effects when using the placebo lines instead. In column (3), we verify that the estimated effect for the branch lines does not exceed the one for the main line. In columns (4) and (5), we include 41 district fixed effects and 21 ethnic group fixed effects respectively. In column (6), we include first, second, third and fourth order polynomials of the longitude and latitude of the location, as in a spatial regression discontinuity design. In column (7), the placebo locations are the control group. All these strategies give similar results as in Table 2.

Results are also robust to a battery of robustness checks (Table 4). In column 1, we dropped the coastal province. In column 2, we used the full sample of locations including the ones in the sparsely populated, arid parts of Kenya. In column 3, we included distance to each of the nodes to account for possible network effects in the urban pattern. In column 4, we dropped the controls. This indeed has the biggest impact, increasing the estimated rail effect on European and urban population while decreasing the effect on population density. In column (5), we controlled for whether the location is within 10 km from a paved or improved road in 1963, demonstrating that railroads do not proxy other infrastructure investments. Results are also not sensitive to using logs instead of z-scores of the population (column 6). Finally, in column 7, we report Conley standard errors to correct for possible spatial correlation, with little effect.

3.5 Additional results

We showed that railroads attracted European settlers and created cities. We now examine other outcomes at independence. This is important, because it clarifies what other factors the railroads brought, which could then cause path dependence. We examine whether locations that were connected by rail also received more health and education investments (proxied by the presence of health centres and secondary schools), public services linked to enforcement of property rights (police stations), communication and other transportation facilities (post offices, paved and improved roads). We use the same set of controls as previously. Outcome variables were close to 0 for the year 1901. The cross-sectional regressions should therefore be interpreted as long-differenced estimations for the period 1901-1962. In Panel A, Table 5, we regress each measure of historical factors on the rail dummies. We find that there are more health centres and police stations in the immediate, 10 km vicinity of the railroad (column 1 and 3). There are significantly more secondary schools and post offices up to 30km, but their presence was decreasing with distance to the railroad (column 2 and 4). This pattern points to a diffusion of services. This is even more the case with paved roads that typically feeded and followed the rail network (column 5). Improved roads, in contrast, are completely unrelated suggesting that they were built in peripheral and core locations alike (column 6). In Panel B, Table 5, we report results conditioned on population (log total population) in 1962. Coefficients thereby obtain a per capita meaning. This is to test whether locations accumulated factors beyond what is expected by higher population densities due to the railroad, that is whether railroad locations were qualitatively different from locations of similar size. We find that this was the case with respect to secondary schools (column 2), post offices (column 4) and paved roads (column 5).

4. PATH DEPENDENCE IN DEVELOPMENT

In this section we document the decline of railroads and study their effects on long-run economic development. Railroad locations are more developed today, although they have lost their initial advantage in terms of transportation. We use our novel data set to examine the channels of path dependence.

4.1 Decline of Railroads and Exodus of Europeans

The rail network reached its maximum length in 1975, with 1,695 miles of track, and has stagnated since (Mitchell 2007). Freight volumes, however, declined continuously since the late 1970s, from 4 million tons of goods to around 2 million tons in the 1990s. What caused the collapse of rail? The World Bank (2005: 17) described problems of underinvestments and management issues in the parastatal Kenya Railways Corporation (KRC): "The railways suffer from aging tracks and rolling stock, not enough resources for maintenance", poor service and excessive and unpredictable transit times. Moreover, Kenyan governments massively invested in the road network. Kenya's length of paved roads increased from 1,405 km in 1965 to 6,635 km in 2002 (Burgess et al., 2013*b*).[11] Road outcompeted rail. In 2002, KRC only had a 15 to 20 percent market share in goods transport, though railroads had a notably larger market share in the transportation of wheat (51%) and coffee (50%) (The World Bank, 2005).

The number of Europeans in Kenya grew exponentially from 500 in 1901 to 55,759 in 1962. After independence, Kenyan Europeans left in large numbers. Land was distributed to African farmers, who were continuing coffee cultivation. But when world market price dropped in the 1980s, coffee production declined sharply (Figure 4). In other words, locations lost the very factors that gave raise to the urban pattern. The fact that numerous roads were built in the rest of the country implies that railroad locations lost their initial advantage in terms of transportation, and other factors must account for any urban persistence.

4.2 Economic Change Post-1962

Has the level effect on economic development in 1962 narrowed over time? We test this hypothesis by studying the relative growth of connected cells after 1962. We run the following model for 403 non-arid locations l for the year 1999:

$$zUPop_{l,1999} = \alpha' + Rail_{l,1962}\beta' + X_l\zeta' + \upsilon_{l,1999} \qquad (2)$$

with $zUPop_{l,1999}$ being the standard score of urban population in 1999. We standardize the urban variable to account for demographic growth post-1962. Kenya's population was multiplied by 3 in 40 years, and the size of its cities increased as a result of natural increase and advances in urban housing and transportation technologies. Results are shown in Table 6. Column 1 shows that railroad locations are more urbanized in 1999. Interestingly, the rail 0-10 km coefficient for urban population in 1999 is identical to the one for 1962 (column 2, Table 2), implying that the effect on the locations' ranking is stable over time. In column 2, we include European population in 1962, which again is statistically significant and reduces β' for $Rail$ 0-10 km by about a half. In column 3 we add urban population in 1962.

[11] Paved roads made about 13% of the total road network 1990-2000 (WDI).

The coefficient is highly significant and renders Rail and European Population insignificant. Hence, locations are more urbanized today primarily because they were more urbanized in the past. The previous analysis does not allow to study the dynamics of path dependence between the two years. We run the following model for 403 non-arid locations l and years $t = [1962, 1969, 1979, 1989, 1999, 2009]$:

$$zUPop_{l,t} = \alpha'' + zUPop_{l,t-1}\theta + Rail_{l,1962}\beta''_t + X_l\zeta'' + \upsilon_{l,t} \quad (3)$$

with $zUPop_{l,t}$ being the standard score of urban population in year t. As we control for the standard score of urban population in the previous period ($zUPop_{l,t-1}$), β''_t captures the additional effect of the railroad for each period. We have data on the urban population from 1901 to 2009. We do not have data on total population in 1901, 1969 and 2009. For the year 1901, we nevertheless control for historical settlement patterns using the various controls listed in Table 1. For the year 1979, the previous period is 1962. Figure 4 displays the effects β''_t for both the urban and total populations. The effects were large in 1962 (vs. 1901), as already shown in the column (1) of Table 6. The effects then decreased over time, and were not significant after 1979, which implies that the level effects completely stabilized in the 1980s. Interestingly, the effects are stronger for total population post-1962. Indeed, the various ethnic groups around the Rift Valley kept colonizing the land around the railroad lines after independence. Land expansion decelerated in the 1980s, probably as a result of low agricultural prices (including coffee prices).

4.3 Channels of Path Dependence

In Kenya, railroad cities persisted post-1962. The effects were unchanged between 1962 and 1999, although rail traffic collapsed in the 1980s. What could explain urban persistence? Bleakley & Lin (2012), who examine the long-term effects of portage sites on population patterns in the U.S., contrast the respective roles of historical and contemporary factors. First, (colonial) *sunk investments* could induce people to stay at these locations. If schools and hospitals are expensive to build, people are less mobile and initial advantages have long-run effects. The long-term effects of historical factors will depend on how fast sunk capital depreciates. The decline of railroads dates from only 20 years ago. Railroad cells are thus likely to be over-supplied with such factors, at least in the short run, and it may take time before population moves to other locations. Second, if there are returns-to-scale in production, factors need to be co-located in the same locations. There is a *coordination problem* as it is not obvious which locations should have the contemporary factors. In this case, it makes sense to locate factors in locations that are already developed, for example the railroad cells. The location of contemporary factors (including people) today then depends on past population density, without it being explained by historical factors. We study how the railroad effects on urban population today vary as we control for the various channels of path dependence, i.e. the historical factors and contemporary factors.

Historical Factors. In section 3.5, we showed that connected cells had better (non-railroad) infrastructure at independence (1962). The locations along the railroad lines had better access to health, educational, institutional, communication and transportation infrastructure. These effects were not entirely explained by rising population densities by 1962. If these historical factors had an independent long-term effect on urban population today, including them in the regression of column (X) should capture some of the effect of urban population in 1962 on urban population in 1999. We actually find that their inclusion reduces the coefficient of urban population in 1962 by 15% (from 0.82*** to 0.69***, columns (3) and (4) of Table 6). Thus, 15% of urban persistence post-1962 can be explained by sunk investments. These results are in line with the Table III of Bleakley & Lin (2012), where the inclusion of historical factors appears to reduce the long-term effects of portage in the U.S. by similar magnitudes. While historical factors matter to explain urban patterns today, they are not the main channel of path dependence. For example, even if railroads collapsed in the 1980s, and were replaced by roads at nearby sites, we find that changes in transportation technology do not explain why the location along the railroad lines are more developed now.

Contemporary Factors. We verify that railroad cells have higher densities of contemporary factors today (1999). We run the same regression as model (X), except we use as a dependent variable the density of various contemporary factors. Results are displayed in Table 7. In Panel A of columns (1)-(10), we show that the inhabitants of railroad locations have better access to health, educational and transportation infrastructure, even if the effects are not always significant at 10%. For example, railroad locations are characterized by a higher number of health clinics (column (4)), and higher probabilities of having a secondary school (column (9)) or being crossed by a paved road (column (11)). The point estimates are strongly reduced when controlling for population today (Panel B). In other words, cells that are more populated today also have higher densities of contemporary factors. These results validate the coordination failure hypothesis. This is also what was suggested by Figure 4, which showed that high densities at one period leads to high densities the next period, and, similarly, in the next periods. In column (5) of Table 6, we find that controlling for contemporary factors does not modify the relationship between urban population in 1962 and urban population in 1999 (the coefficient remains unchanged at 0.69***). These results point to the following story: railroad locations have higher densities today, because people co-locate where there are more people in the previous period, and other contemporary factors "follow" people. There were then more people in the previous period because of the initial population effect during the colonial period and repeated co-location decisions.

To conclude, if historical factors matter to explain urban patterns today, railroad cities persisted over time mostly because they solved the coordination failure of contemporary factors before independence, and for each subsequent period then.

4.4 Economic Development Today

We have shown that railroad locations have higher densities of contemporary factors today. However, these positive effects are somewhat reduced once we control for population densities. In per capita terms, railroad locations are not necessarily better endowed in factors than non-railroad locations of similar densities. However, these locations could still be wealthier than the other locations. We run the same regression as model (X) except we now use various contemporary measures of economic development as the dependent variable and we simultaneously control for urban population in 1999. We first use the poverty rate in 1999. Column (6) of Table 6 confirms that railroad locations are relatively less poor today, especially until 10 km (-4.5%, given a mean of 51%). We also use satellite data on night lights as an alternative measure of development, in line with Henderson, Storeygard & Weil (2012). Our dependent variable is average light intensity for each cell in 2000-01. Column (7) shows that the rail effects are positive, and significant until 10 km. Henderson, Storeygard & Weil (2012) shows that the elasticity of *ln(lights/area)* to *ln(GDP)* is 0.3. We modify our dependent variable to be in line with their paper, and we find the following railroad effects on *ln(lights/area)* when controlling for *ln(population)*: 2.2*** for 0-10 km, 1.6*** for 10-20 km, 1.4*** for 20-30 km and 1** for 30-40 km (not shown).[12] These locations are thus 66%, 48%, 42% and 30% wealthier per capita than other cells. Why are railroad locations wealthier than non-railroad locations? Since we are already controlling for population density, this difference is not due to a higher density of (observable) contemporary factors. But there could be unobservable contemporary factors (e.g., unobservable skills or specific economic activities) that were repeatedly co-located along the railroads.

5. DISCUSSION

The economic geography and development literatures have argued over whether locational fundamentals (i.e., the geographical endowments of various locations) or increasing returns (i.e., localized historical shocks) are the main determinants for the distribution of economic activity across space. If locational fundamentals are the main determinants of spatial patterns, any shock will have only temporary effects. If increasing returns are the main determinants of spatial patterns, shocks can have permanent effects. This has implications for regional policy. In particular, little is known about the extent and forces of path dependence in developing countries. The respective contributions of locational fundamentals and increasing returns could be different in poor countries, since they are more rural and agricultural (and thus less urban and industrial). Likewise, the respective contributions of sunk investments and coordination failures could vary with the level of development.

We used a natural experiment and a new data set on railroads and city growth at a fine spatial level over one century in Kenya. All of Kenyan railroad lines were built before independence. Although profitable, railroads collapsed in the post-independence period. Yet we showed that rail construction had a persistent impact on Kenya's urban system: locations close to the railroad are richer today. Our results suggest that railroads served as a mechanism to solve a coordination problem. Given returns-to-scale in production, factors need to be co-located in the same locations. There is a coordination problem as it is not obvious which locations should have the contemporary factors. Then it makes sense to locate factors in locations that are already developed. While colonial sunk investments matter, we show that railroad cities mostly persisted because they solved the coordination problem early on. Positive shocks may have large permanent effects in poor regions, by creating cities where no mechanisms of coordination (and few cities) existed before.

Our findings suggest that historical shocks may have large permanent effects in de-

[12] In column (7) of Table 6, we did not use the same outcome as in Henderson, Storeygard & Weil (2012), as there are locations for which no light is observed by satellite. Using the log of night light intensity mechanically drops these observations.

veloping countries. Since these countries are poorer, it may be easier to undo an existing equilibrium in favor of another equilibrium, despite the fact that physical geography is relatively important in these relatively more agricultural economies. The mechanisms of path dependence appear to be the same as in developed countries, as coordination failures mainly explain why spatial patterns may persist over time. Sunk investments are somewhat less important, for the mere reason that these factors are "mobile" in the longer run. Lastly, our results do not imply that any regional policy intervention will strongly affect regional development. They will have only limited effects if all the coordination failures have already been solved. Thus, our results are in line with both theories of the location of economic activities – locational fundamentals and increasing returns –, and we argue that the effects of localized shocks is likely to depend on the context in which they take place.

REFERENCES

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369–1401.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2002. "Reversal Of Fortune: Geography And Institutions In The Making Of The Modern World Income Distribution." *The Quarterly Journal of Economics*, 117(4): 1231–1294.

Aker, Jenny C. 2010. "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger." *American Economic Journal: Applied Economics*, 2(3): 46–59.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert Margo. 2010. "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest 1850-60." *Social Science History*, 34: 171–197.

Atkin, David, and Dave Donaldson. 2012. "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs." Unpublished manuscript, Yale University, Department of Economics.

Banerjee, Abhijit, and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review*, 95(4): 1190–1213.

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

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

Beachey, R. W. 1967. "The East African Ivory Trade in the Nineteenth Century." *The Journal of African History*, 8(2): 269–290.

Behrens, Kristian. 2007. "On the location and lock-in of cities: Geography vs transportation technology." *Regional Science and Urban Economics*, 37(1): 22–45.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *The Quarterly Journal of Economics*, 127(2): 587–644.

Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm. 2007. "Looking for multiple equilibria when geography matters: German city growth and the WWII shock." *Journal of Urban Economics*, 61(1): 152–169.

Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm. 2008. "A century of shocks: The evolution of the German city size distribution 1925-1999." *Regional*

Science and Urban Economics, 38(4): 330–347.

Bruhn, Miriam, and Francisco A. Gallego. 2012. "Good, Bad, and Ugly Colonial Activities: Do They Matter for Economic Development?" *The Review of Economics and Statistics*, 94(2): 433–461.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2013*a*. "The Value of Democracy: Evidence from Road Building in Kenya." National Bureau of Economic Research Working Paper 19398.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padro-i Miguel. 2013*b*. "The Value of Democracy: Evidence from Road Building in Kenya." Unpublished manuscript, Department of Economics, London School of Economics.

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.

Chaves, Isaías, Stanley L. Engerman, and James A. Robinson. 2012. "Reinventing the Wheel: The Economic Benefits of Wheeled Transportation in Early Colonial British West Africa." Unpublished manuscript, Department of Political Science, Stanford University.

Colony and Protectorate of Kenya. 1926. *Suggested Branch Railway Lines.* Nairobi, Kenya: Government Printer.

Davis, Donald R., and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5): 1269–1289.

Davis, Donald R., and David E. Weinstein. 2008. "A Search For Multiple Equilibria In Urban Industrial Structure." *Journal of Regional Science*, 48(1): 29–65.

Deininger, Klaus, and Hans P Binswanger. 1995. "Rent Seeking and the Development of Large-Scale Agriculture in Kenya, South Africa, and Zimbabwe." *Economic Development and Cultural Change*, 43(3): 493–522.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica*, 78(6): 1863–1903.

Dell, Melissa. 2012. "Path Dependence in Development: Evidence from the Mexican Revolution."

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehan. 2008. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." International Food Policy Research Institute. IFPRI discussion papers 840.

Dittmar, Jeremiah E. 2011. "Information Technology and Economic Change: The Impact of The Printing Press." *The Quarterly Journal of Economics*, 126(3): 1133–1172.

Donaldson, Dave. 2013. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review, Forthcoming*.

Donaldson, Dave, and Richard Hornbeck. 2013. "Railroads and American Economic Growth: A "Market Access" Approach." National Bureau of Economic Research, Inc NBER Working Papers 19213.

Duranton, Gilles, and Matthew A. Turner. 2012. "Urban Growth and Transportation." *Review of Economic Studies*, 79(4): 1407–1440.

Faber, Benjamin. 2013. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." Unpublished manuscript, London School of Economics, Department of Economics.

Feyrer, James, and Bruce Sacerdote. 2009. "Colonialism and Modern Income: Islands as Natural Experiments." *The Review of Economics and Statistics*, 91(2): 245–262.

Fujita, Masahisa, and Tomoya Mori. 1996. "The role of ports in the making of major cities:

Self-agglomeration and hub-effect." *Journal of Development Economics*, 49(1): 93–120.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew Mellinger. 1999. "Geography and Economic Development." Center for International Development at Harvard University CID Working Papers 1.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr. 2012. "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." National Bureau of Economic Research, Inc NBER Working Papers 18524.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9(3): 271–303.

Heldring, Leander, and James A. Robinson. 2012. "Colonialism and Economic Development in Africa." National Bureau of Economic Research, Inc NBER Working Papers 18566.

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

Hill, Mervyn F. 1950. *Permanent way.* Nairobi, Kenya: East African Railways and Harbours.

Huillery, Elise. 2009. "History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa." *American Economic Journal: Applied*, 1: 176–215.

Iyer, Lakshmi. 2010. "Direct versus Indirect Colonial Rule in India: Long-Term Consequences." *The Review of Economics and Statistics*, 92(4): 693–713.

Jacoby, Hanan G., and Bart Minten. 2009. "On Measuring the Benefits of Lower Transport Costs." *Journal of Development Economics*, 89(1): 28–38.

Jedwab, Remi, and Alexander Moradi. 2013. "Transportation Technology and Economic Change: Evidence from Colonial Railroads and City Growth in Africa." Unpublished manuscript, Department of Economics, George Washington University.

Jensen, Robert. 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *The Quarterly Journal of Economics*, 122(3): 879–924.

Johnson, Simon, Jonathan David Ostry, and Arvind Subramanian. 2007. "The Prospects for Sustained Growth in Africa: Benchmarking the Constraints." International Monetary Fund. IMF Working Papers 07/52.

Kapila, Neera Kent. 2009. *Race, Rail, and Society: Roots of Modern Kenya.* Nairobi, Kenya: East African Educational Publishers.

Kenya and Uganda Railways and Harbours. 1931. *General Manager's Bulletin No 2.* Nairobi, Kenya: Office of the General Manager of Railways and Harbours.

King, Anne, and R. M. A. van Zwanenberg. 1975. *An Economic History of Kenya and Uganda, 1800-1970.* Basingstoke, UK: Palgrave Macmillan.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281–355.

Lucas, Robert Jr. 1988. "On the mechanics of economic development." *Journal of Monetary Economics*, 22(1): 3–42.

Memon, P. A. 1973. "The spatial dynamics of trade and urban development in Kenya during the early colonial period up to 1915." *Nairobi Institute for Development Studies Working Paper*, 78.

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

Michalopoulos, Stelios, and Elias Papaioannou. 2011. "The Long-Run Effects of the Scramble for Africa." CEPR Discussion Papers 8676.

Michalopoulos, Stelios, and Elias Papaioannou. 2012. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica, Forthcoming*.

Miguel, Edward, and Gérard Roland. 2011. "The long-run impact of bombing Vietnam." *Journal of Development Economics*, 96(1): 1–15.

Morgan, W. T. W. 1963. "The "White Highlands" of Kenya." *The Geographical Journal*, 129(2): 140–155.

Mosley, Paul. 1983. *The Settler Economies: Studies in the Economic History of Kenya and Southern Rhodesia 1900-1963.* Cambridge: Cambridge University Press.

Nunn, Nathan. 2008. "The Long-term Effects of Africa's Slave Trades." *The Quarterly Journal of Economics*, 123(1): 139–176.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics*, 94(1): 20–36.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review*, 101(7): 3221–52.

O'Connor, Anthony M. 1964. "New Railway Construction and the Pattern of Economic Development in East Africa." *Transaction: Institute of British Geographers*, 36: 21–30.

O'Connor, Anthony M. 1965. *Railways and development in Uganda: a study in economic geography.* Nairobi, Kenya: East African Institute of Social Research by Oxford University Press (1965).

Otte, T. G., and K. Neilson. 2006. *Railways and International Politics : Paths of Empire, 1848-1945.* London, UK: Routledge Military History and Policy series (Book 24).

Radelet, Steven, and Jeffrey Sachs. 1998. "Shipping Costs, Manufactured Exports, and Economic Growth." Paper resented at the American Economics Association Meeting.

Rappaport, Jordan, and Jeffrey D Sachs. 2003. "The United States as a Coastal Nation." *Journal of Economic Growth*, 8(1): 5–46.

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf. 2011. "History and Industry Location: Evidence from German Airports." *The Review of Economics and Statistics*, 93(3): 814–831.

Rodrik, Dani. 1998. "Trade Policy and Economic Performance in Sub-Saharan Africa." National Bureau of Economic Research, Inc. NBER Working Papers 6562.

Sachs, Jeffrey D. 2003. "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income." National Bureau of Economic Research Working Paper 9490.

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

Storeygard, Adam. 2012. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in sub-Saharan Africa." Unpublished manuscript, Department of Economics, Brown University.

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 and spelling (1st ed.).* Nairobi, Kenya: Survey of Kenya.

The World Bank. 2005. *Kenya: Issues in Trade Logistics.* Washington, DC: World Bank.

Wrigley, C. C. 1959. *Crops and wealth in Uganda; a short agrarian history.* Kampala, Uganda: East African Institute of Social Research.

TABLE 1: OBSERVABLES FOR TREATED VS. CONTROL LOCATIONS IN 1901

Variable of Interest:	**Railroad Dummy, 0-20 km**			
Group of Control Cells:	**(1) All Control Locations**		**(2) Placebo Locations**	
Dependent Variable:	Coeff.	SE	Coeff.	SE
Share of arid soils (%)	-0.00**	(0.00)	-0.00	(0.00)
"Coastal location" dummy	-0.03	(0.03)	-0.01	(0.02)
Distance to the coast (km)	-6.75	(6.46)	-15.87**	(6.70)
Average annual rainfall (mm)	-61.1*	(31.3)	-123.0***	(44.5)
Altitude: mean (m)	24.7	(39.1)	65.2	(49.0)
Area (sq km)	-0.16	(0.11)	0.14	(0.11)
"City in 1901" dummy	-0.01*	(0.01)	0.00	(0.00)
"Provincial capital in 1962" dummy	0.00	(0.01)	-0.00	(0.02)
Area of "major settled groups" (sq km)	-0.31***	(0.11)	-0.38***	(0.14)
Area of "pastoralists" (sq km)	-0.17	(0.11)	0.04	(0.10)
"Isolated groups" dummy	0.02	(0.05)	0.19***	(0.05)
"White Highlands" dummy	0.23***	(0.04)	0.23***	(0.05)
Number of treated locations:	162		162	
Number of control locations:	241		69	

Notes: OLS regressions using data on 12 outcomes for 403 non-arid locations in 1901. These are the main controlling variables we use in our empirical analysis. This table tests whether the treated and control locations are significantly different in terms of observable characteristics around 1901, for various groups of control locations. We regress each control variable on a dummy equal to one if the location is less than 20 km from a (colonial) railroad line, while simultaneously including province fixed effects (N = 8). Robust standard errors: * p<0.10, ** p<0.05, *** p<0.01. There are 12 different regressions for each column. There are 162 treated locations. In column (1), all control locations are included (N = 241). In column (2), the control locations are the locations less than 20 km from a placebo line (N = 69). See Online Data Appendix for data sources.

TABLE 2: COLONIAL RAILROADS AND POPULATION GROWTH, 1901-1962

Dependent Variable:	**Number of Inhabitants in 1962 (Z-Score)**				
	European Pop.	**Urban Pop.**		**Total Pop.**	
	(1)	(2)	(3)	(4)	(5)
Railroad Dummy, 0-10 km	0.54***	0.43**	0.22	0.32**	0.21
	(0.20)	(0.20)	(0.14)	(0.14)	(0.14)
Railroad Dummy, 10-20 km	0.07	0.10	0.07	0.31**	0.30**
	(0.13)	(0.13)	(0.12)	(0.14)	(0.14)
Railroad Dummy, 20-30 km	0.12	0.01	-0.04	0.07	0.05
	(0.11)	(0.11)	(0.10)	(0.16)	(0.16)
Railroad Dummy, 30-40 km	-0.02	-0.18	-0.17*	-0.08	-0.08
	(0.11)	(0.11)	(0.09)	(0.19)	(0.19)
Railroad Dummy, 40-50 km	-0.05	-0.15*	-0.13*	0.07	0.08
	(0.09)	(0.09)	(0.08)	(0.13)	(0.13)
Number of Europeans			0.39**		0.21***
(Z-Score, 1962)			(0.20)		(0.07)
Province Fixed Effects	Y	Y	Y	Y	Y
Location Controls	Y	Y	Y	Y	Y
Number of Observations	403	403	403	403	403
Adj. R-squared	0.26	0.34	0.45	0.42	0.45

Notes: OLS 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. In column (1), the dependent variable is the standard score of European population of location l in 1962. In columns (2)-(3), the dependent variable is the standard score of the population of location l that resides in localities whose population is superior to 2,000 inhabitants in 1962. As the European and urban populations were almost nil in 1901, the cross-sectional regressions of columns (1)-(3) should be interpreted as long-differenced estimations for the period 1901-1962. In columns (4)-(5), the dependent variable is the standard score of the total population of location l in 1962. All regressions include province fixed effects (N = 8), and controls at the location level to account for pre-existing settlement patterns: share of arid soils (%), "coastal location" dummy, distance to the coast (km), average annual rainfall (mm), altitude (m), area (sq km), "city in 1901" dummy, "provincial capital in 1962" dummy, area of "major settled groups" (sq km), area of "pastoralists" (sq km), "isolated groups" dummy, and "white Highlands" dummy. Mombasa, Nairobi and Kisumu are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 3: ALTERNATIVE IDENTIFICATION STRATEGIES, 1901-1962

Strategy:	*Baseline Results (see Table 1)* (1)	*Placebo Treat-ment Effects* (2)	*Effects Main vs. Branch Lines* (3)	*Including District Fixed Effects* (4)	*Including Ethnic Fixed Effects* (5)	*Spatial RDD (Long Lat)* (6)	*Control Group: Placebo Locations* (7)
Dependent Variable:	**Number of European Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-10 km	0.49***	0.05	0.40*	0.51***	0.46**	0.46***	0.43**
	(0.18)	(0.13)	(0.23)	(0.19)	(0.18)	(0.18)	(0.20)
Main 1963, 0-10 km			0.20				
			(0.41)				
Dependent Variable:	**Number of Urban Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-10 km	0.41**	0.18	0.41*	0.40**	0.40***	0.41**	0.38*
	(0.18)	(0.12)	(0.25)	(0.17)	(0.15)	(0.18)	(0.21)
Main 1963, 0-10 km			-0.00				
			(0.37)				
Dependent Variable:	**Total Number of Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-20 km	0.29***	0.08	0.27**	0.35***	0.18*	0.24**	0.31**
	(0.09)	(0.07)	(0.11)	(0.10)	(0.10)	(0.09)	(0.12)
Main 1963, 0-20 km			0.05				
			(0.12)				
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y
Number of Obs.	403	403	403	403	403	403	231

Notes: OLS 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. Column (1) replicates the main results of Table 2, using 0-10 and 0-20 km railroad dummies for the sake of simplicity. In column (2), we test that there are no effects when using the placebo lines instead. In column (3), we verify that the effects are not lower for the main line. In columns (4) and (5), we include 41 district fixed effects and 21 ethnic group fixed effects respectively. In column (6), we include first, second, third and fourth order polynomials of the longitude and latitude of the location, as in a spatial regression discontinuity design. In column (7), the placebo locations are the control group. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 4: ROBUSTNESS CHECKS, 1901-1962

Robustness Test	*Dropping the Coastal Province* (1)	*Using the Full Sample* (2)	*Include Distances to Each Node* (3)	*Dropping the Controls* (4)	*Controlling for Roads 1963* (5)	*Logs instead of Z-Scores* (6)	*Conley Standard Errors (50 km)* (7)
Dependent Variable:	**Number of European Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-10 km	0.34**	0.46***	0.45***	0.64***	0.49***	0.76***	0.47**
	(0.16)	(0.17)	(0.17)	(0.21)	(0.18)	(0.25)	(0.21)
Dependent Variable:	**Number of Urban Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-10 km	0.47**	0.39**	0.41**	0.52**	0.41**	0.70*	0.30*
	(0.21)	(0.17)	(0.18)	(0.25)	(0.16)	(0.36)	(0.16)
Dependent Variable:	**Total Number of Inhabitants in 1962 (Z-Score)**						
Rail 1963, 0-20 km	0.27**	0.25***	0.26***	0.16*	0.29***	0.79***	0.192*
	(0.11)	(0.08)	(0.09)	(0.09)	(0.10)	(0.25)	(0.11)
Province Fixed Effects	Y	Y	Y	Y	Y	Y	Y
Location Controls	Y	Y	Y	N	Y	Y	Y
Number of Obs.	344	470	403	403	403	403	403

Notes: OLS 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. The variables of interest are the 0-10 and 0-20 km railroad dummies for the sake of simplicity. In column (1), we drop the coastal province and focus on the non-arid locations west of Nairobi. In column (2), we use the full sample. In column (3), we include the Euclidean distances (km) to Mombasa, Nairobi and Kisumu. In column (4), we drop the controls. In column (5), we control for whether the location is within 10 km from a paved or improved road in 1963. In column (6), we use the log of European / urban / total population as the dependent variable. In column (7), standard errors are corrected for spatial autocorrelation using the approach of Conley (1999), with a distance cut-off of 50 km. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 5: COLONIAL RAILROADS AND HISTORICAL FACTORS, 1901-1962

Dependent Variable:	**Hospital or Clinic Dummy**	**Secondary School Dummy**	**Police Station Dummy**	**Post Office Dummy**	**Paved Road Dummy**	**Improved Road Dummy**
	(2)	(1)	(3)	(4)	(5)	(6)
Panel A: ***Railroads and historical factors (1962)***						
Railroad Dummy, 0-10 km	0.14*	0.28***	0.14*	0.35***	0.62***	0.11
Railroad Dummy, 10-20 km	0.05	0.15**	0.11	0.25***	0.36***	-0.10
Railroad Dummy, 20-30 km	0.09	0.11*	0.02	0.13*	0.18***	-0.10
Railroad Dummy, 30-40 km	-0.01	0.11	-0.01	0.03	0.16**	0.02
Railroad Dummy, 40-50 km	-0.02	0.06	-0.02	0.15	0.18***	-0.01
Panel B: ***Railroads and historical factors (1962), conditioned on historical population (1962)***						
Railroad Dummy, 0-10 km	0.11	0.26***	0.11	0.34***	0.61***	0.09
Railroad Dummy, 10-20 km	0.02	0.13**	0.07	0.22***	0.35***	-0.11
Railroad Dummy, 20-30 km	0.09	0.11*	0.02	0.13*	0.18***	-0.10
Railroad Dummy, 30-40 km	-0.01	0.11*	-0.01	0.03	0.16**	0.02
Railroad Dummy, 40-50 km	-0.02	0.05	-0.02	0.14	0.18***	-0.01
Mean of the Variable	0.28	0.15	0.30	0.33	0.28	0.30
Province FE, Controls	Y	Y	Y	Y	Y	Y

Notes: OLS regressions using infrastructure data on 403 non-arid locations for the year 1962. Robust standard errors are not reported, for the sake of space; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the railroad dummies on six measures of historical factors in 1962. As these measures were close to 0 for the year 1901, the cross-sectional regressions of columns (1)-(6) should be interpreted as long-differenced estimations for the period 1901-1962. In Panel A, we regress each measure of historical factors on the rail dummies. In Panel B, we control for historical population (log total population) in 1962. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 6: COLONIAL RAILROADS AND ECONOMIC DEVELOPMENT, 1999

Dependent Variable:	**Col. (1)-(5): Urban Population in 1999 (Number of Inhabitants, Z-Score)**					**Poverty Rate (%) 1999**	**Night Lights 2000**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Railroad Dummy, 0-10 km	0.43**	0.23	0.05	-0.16	-0.18*	-4.47**	0.89*
	(0.20)	(0.15)	(0.08)	(0.10)	(0.10)	(1.91)	(0.48)
Railroad Dummy, 10-20 km	-0.05	-0.08	-0.14*	-0.24**	-0.22**	-2.85	0.34
	(0.13)	(0.13)	(0.08)	(0.10)	(0.10)	(1.75)	(0.33)
Railroad Dummy, 20-30 km	-0.14	-0.18**	-0.16**	-0.21**	-0.18**	-2.45	0.28
	(0.09)	(0.09)	(0.07)	(0.08)	(0.07)	(1.62)	(0.27)
Railroad Dummy, 30-40 km	-0.17	-0.16*	-0.02	-0.04	-0.01	-0.66	0.44
	(0.12)	(0.09)	(0.08)	(0.07)	(0.08)	(1.73)	(0.35)
Railroad Dummy, 40-50 km	-0.15	-0.13	-0.02	-0.02	-0.04	0.99	-0.08
	(0.10)	(0.08)	(0.06)	(0.05)	(0.05)	(1.98)	(0.29)
European Pop. 1962 (Z)		0.37**	0.06	0.07	0.06	0.27	-0.09
		(0.18)	(0.07)	(0.06)	(0.06)	(1.40)	(0.50)
Urban Pop. 1962 (Z)			0.82***	0.69***	0.69***	0.05	0.53
			(0.11)	(0.08)	(0.08)	(0.41)	(0.39)
Urban Pop. 1999 (Z)						-0.68	1.20*
						(0.93)	(0.66)
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y
Historical Factors 1962	N	N	N	Y	Y	Y	Y
Contemporary Factors 1999	N	N	N	N	Y	Y	Y
Adj. R-squared	0.27	0.37	0.74	0.81	0.83	0.75	0.58

Notes: OLS panel regressions using population data on 403 non-arid locations for the year 1999. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In columns (1)-(5), the dependent variable is the standard score of the population of location l that resides in localities whose population is superior to 2,000 inhabitants in 1999. In column (4), we control for historical factors in 1962 (see footnote X in the text). In column (5), we control for contemporary factors in 1999 (see footnote X in the text). In columns (6), the dependent variable is the poverty rate (%) of location l in 1999. In column (7), the dependent variable is mean night light density of location l in 2000-01. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 7: COLONIAL RAILROADS AND CONTEMPORARY FACTORS, 1999

Dependent Variable:	**Hospital Dummy**	**Number Hospitals**	**Health Clinic Dummy**	**Number Health Clinics**	**Health Dispensary Dummy**	**Number Health Dispensaries**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: ***Railroads and contemporary factors (1999)***						
Railroad Dummy, 0-10 km	0.07	0.22	0.09	0.90**	0.08**	0.34
Railroad Dummy, 10-20 km	-0.02	-0.13	0.11	0.60*	0.05	-0.87
Railroad Dummy, 20-30 km	-0.08	-0.15	0.01	0.44	0.02	-1.00
Railroad Dummy, 30-40 km	-0.14	-0.32**	0.05	0.04	-0.00	-2.93***
Railroad Dummy, 40-50 km	-0.06	-0.02	0.13*	0.21	-0.01	-1.68*
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999)***						
Railroad Dummy, 0-10 km	0.00	0.12	0.01	0.59	0.05	-0.82
Railroad Dummy, 10-20 km	-0.09	-0.24*	0.03	0.29	0.02	-2.05***
Railroad Dummy, 20-30 km	-0.11	-0.20	-0.02	0.31	0.00	-1.53*
Railroad Dummy, 30-40 km	-0.08	-0.22	0.12*	0.32	0.02	-1.85**
Railroad Dummy, 40-50 km	-0.07	-0.04	0.11	0.17	-0.08	-1.87**
Mean of the Variable	0.43	0.60	0.71	1.87	0.71	7.20
Dependent Variable:	**Primary School Dummy**	**Number Primary Schools**	**Secondary School Dummy**	**Number Secondary Schools**	**Paved Road Dummy**	**Improved Road Dummy**
	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: ***Railroads and contemporary factors (1999)***						
Railroad Dummy, 0-10 km	0.01	6.03	0.07**	2.16	0.44***	-0.01
Railroad Dummy, 10-20 km	0.01	-3.57	0.04	1.16	0.27***	-0.11
Railroad Dummy, 20-30 km	0.01	-11.34	0.02	-2.91	0.22***	-0.06
Railroad Dummy, 30-40 km	-0.02	-18.11*	-0.08	-4.72*	0.10	-0.01
Railroad Dummy, 40-50 km	0.00	-6.00	0.07**	0.01	0.07	0.08
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999)***						
Railroad Dummy, 0-10 km	0.00	-6.79	0.02	-0.99	0.41***	-0.04
Railroad Dummy, 10-20 km	0.00	-16.56***	-0.01	-2.03	0.25***	-0.14*
Railroad Dummy, 20-30 km	0.01	-17.13***	0.00	-4.33**	0.21***	-0.07
Railroad Dummy, 30-40 km	-0.02	-6.15	-0.03	-1.78	0.12	0.02
Railroad Dummy, 40-50 km	-0.00	-8.16	0.06**	-0.51	0.07	0.07
Mean of the Variable	0.99	66.5	0.96	14.5	0.68	0.42
Province FE, Controls	Y	Y	Y	Y	Y	Y

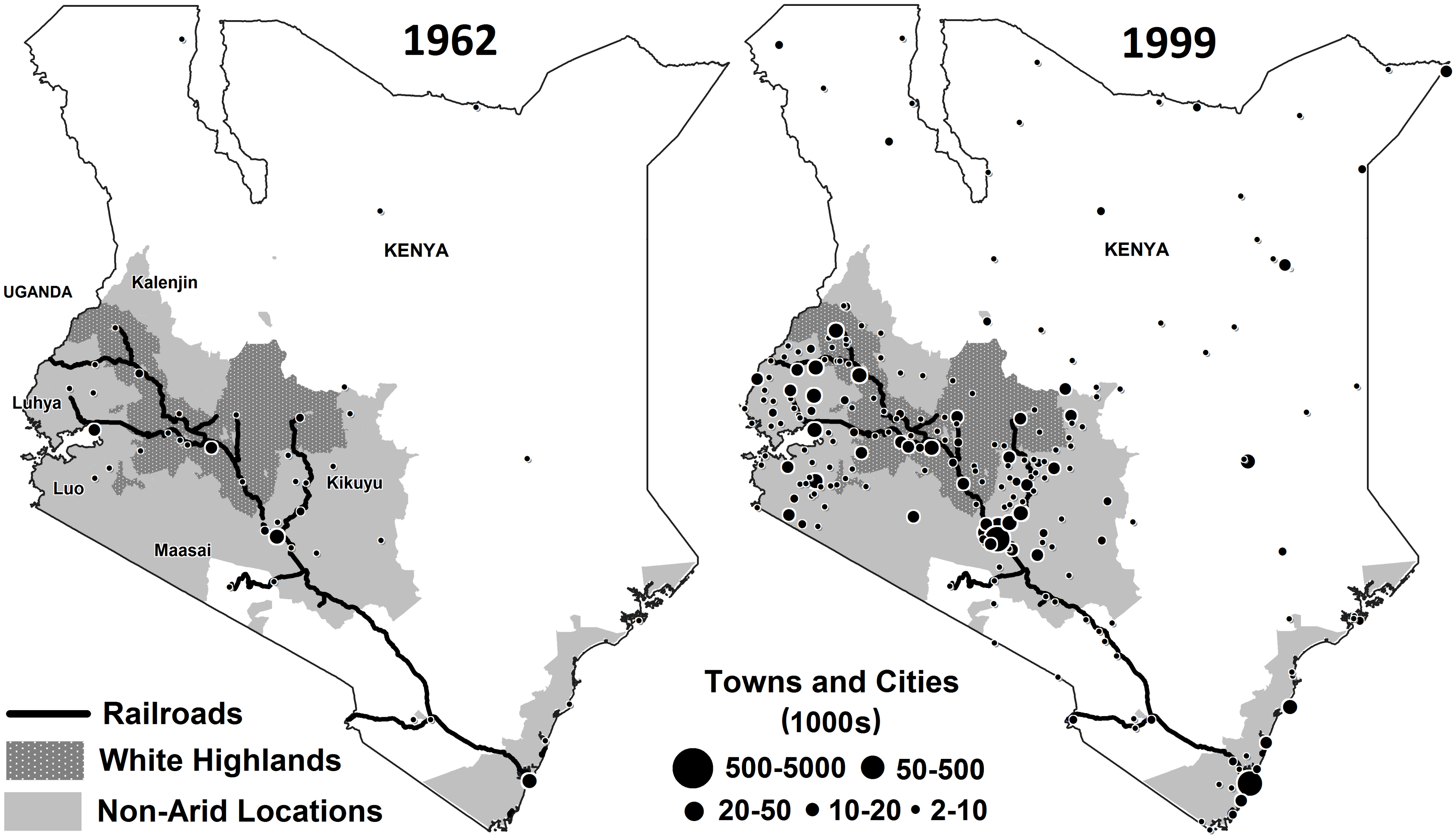
Notes: OLS regressions using infrastructure data on 403 non-arid locations for the year 1962. Robust standard errors are not reported, for the sake of space; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the railroad dummies on twelve measures of contemporary factors in 1999. As these measures were close to 0 for the year 1901, the cross-sectional regressions of columns (1)-(12) should be interpreted as long-differenced estimations for the period 1901-1999. In Panel A, we regress each measure of contemporary factors on the rail dummies. In Panel B, we control for contemporary population (log total population) in 1999. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

APPENDIX TABLE 1: (NON-)EFFECTS FOR PLACEBO LINES, 1901-1962

Type of Placebo Line:	All Lines	Col. (2)-(4): Explorer Routes			Col. (5)-(8): Planned But Never Built			
Placebo Line:		Thomson 1883	Fischer 1885	McKinnon Sclater 1897	Eldoret Sergoit 1926	Mochakos 1926	Kericho Sotik 1926	Thika Ol Donyo Sabuk 1926
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	**Number of European Inhabitants in 1962 (Z-Score)**							
Dummy Placebo, 0-10 km	0.05	0.24	-0.07	-0.02	1.36	0.15	-0.06	0.08
	(0.13)	(0.24)	(0.11)	(0.14)	(1.48)	(0.41)	(0.21)	(0.78)
Dependent Variable:	**Number of Urban Inhabitants in 1962 (Z-Score)**							
Dummy Placebo, 0-10 km	0.18	-0.01	0.10	0.15	3.40	0.85	0.08	2.41
	(0.12)	(0.12)	(0.17)	(0.17)	(2.79)	(0.85)	(0.24)	(1.95)
Dependent Variable:	**Total Number of Inhabitants in 1962 (Z-Score)**							
Dummy Placebo, 0-20 km	0.07	0.09	0.03	0.01	-0.12	0.34	0.20	0.68***
	(0.07)	(0.08)	(0.09)	(0.09)	(0.29)	(0.36)	(0.41)	(0.24)
Province Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Location Controls	Y	Y	Y	Y	Y	Y	Y	Y
Number of Obs.	403	403	403	403	403	403	403	403

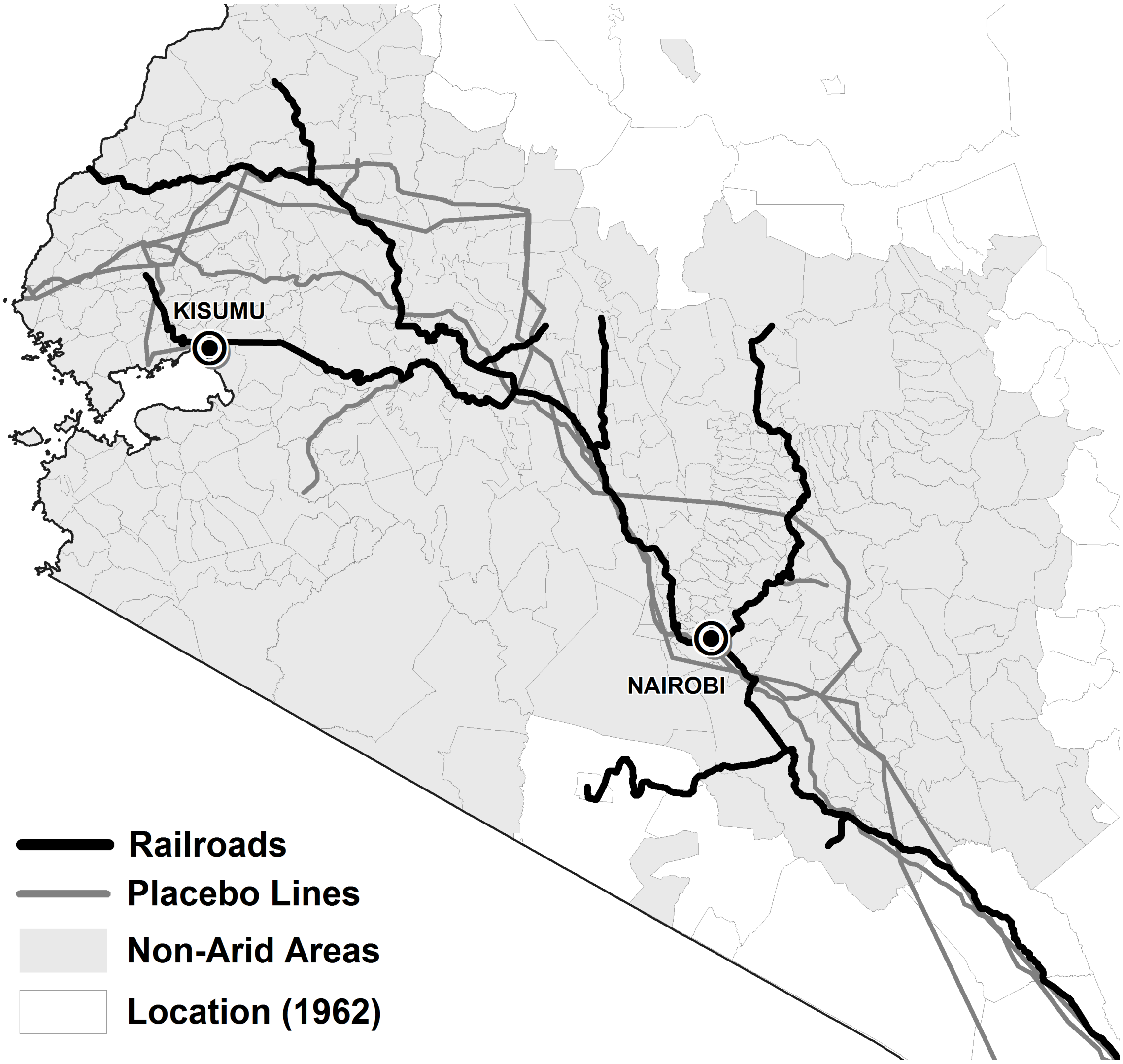
Notes: OLS panel regressions using population data on 403 non-arid locations for the year 1962. Robust standard errors in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The variables of interest are the 0-10 and 0-20 km railroad dummies for the sake of simplicity. This table tests that there are no spurious effects for seven placebo lines in 1962. In column (1), we consider all placebo lines. In columns (2)-(4), we consider three explorer routes as placebo lines of the main line. In columns (5)-(8), we consider four (branch) lines that were planned but not built. There is only one significant effect, when considering total population for Thika-Ol Donyo Sabuk (1926). This effect could be driven by the fact that the placebo line strongly intersects with the actual railroad lines, for which the effect is positive. Reassuringly, this effect becomes -0.59 (not significant) when excluding the locations within 20 km from an actual railroad line. The locations along this placebo line are thus not "different" from other control locations, in 1962. All regressions include province fixed effects (N = 8), and the same controls at the location level as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

Figure 1: Railroads, White Settlement and City Growth in Kenya, 1962-1999



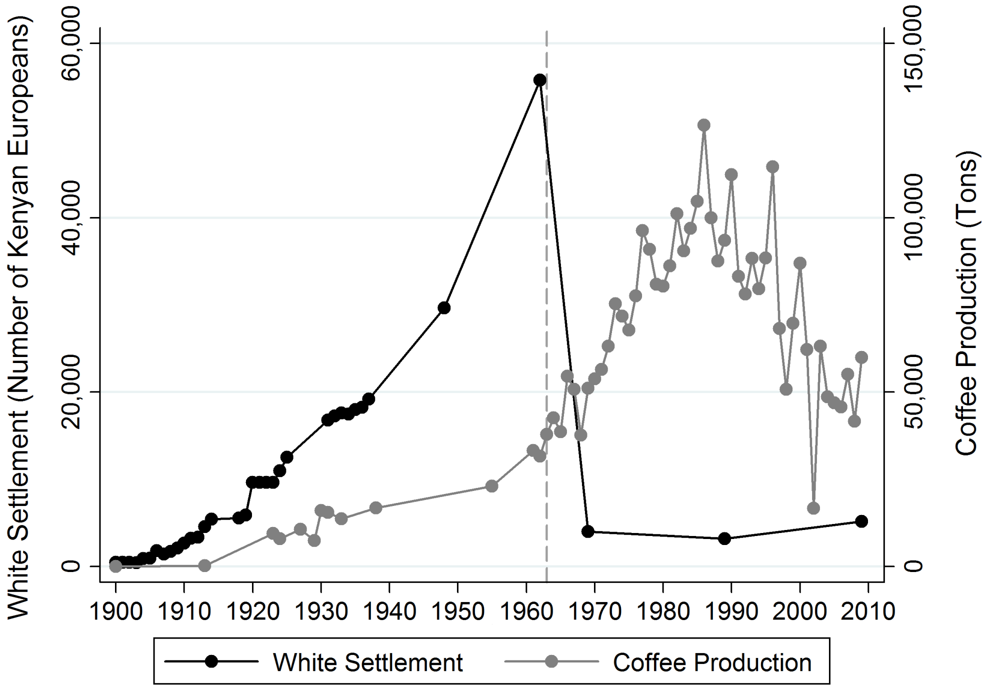
Notes: The map shows railroads and cities in Kenya in 1962 and 1999. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria in the west was built in 1896-1901 to link Uganda to the coast. The non-arid areas are "locations" where arid soils account for less than 10% of the total area (N = 403 out of 473 locations of Kenya in 1962). The White Highlands were areas where Europeans settled in considerable numbers during the colonial period. These areas previously served as a buffer zone between the ethnic groups of the Rift Valley: the Maasais, the Luos, the Luhyas, the Kalenjins and the Kikuyus. Cities are localities where population is superior to 2,000 inhabitants in 1962 (N = 35) and 1999 (N = 185). We do not have data for localities below the 2,000 population threshold. There were only 5 cities in 1901, and all of them except Nairobi (established in 1899) were on the coast. See Online Data Appendix for data sources.

Figure 3: Railroads and Placebo Lines in the Non-Arid Areas of Kenya



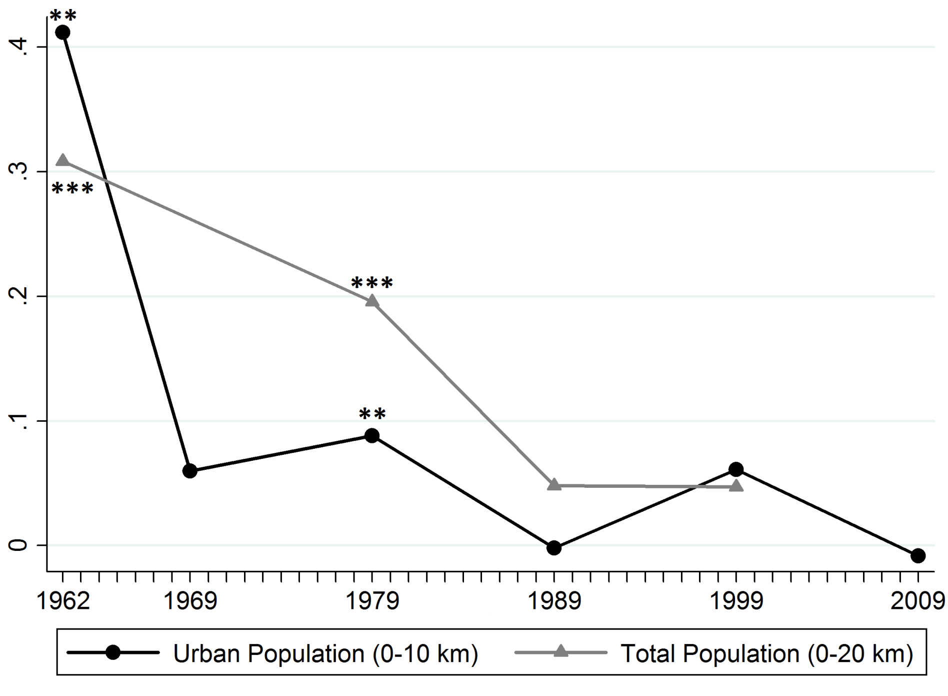
Notes: The map shows the railroad lines and the placebo lines in Kenya. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria in the west was built in 1896-1901 to link Uganda to the coast. The branch lines were built between 1926 and 1926. The map also shows seven placebo lines: the Thomson route (1883), the Fischer route (1885), the McKinnon-Sclater route (1890), Nakuru-Sergoit (1926), Kericho-Sotik (1926), Thika-Donyo Sabuk (1926) and Mochakos (1926). Thomson, Fischer and Sclatter were all explorers in the late 19th century. These explorer routes could have been considered as an alternative itinerary for the main line. The four other placebo lines were branch lines that were planned but not built. The non-arid areas are "locations" where arid soils account for less than 10% of the total area (N = 403 out of 473 locations of Kenya in 1962). The map displays the (non-coastal) non-arid locations of Kenya, to the west of Nairobi. See Online Data Appendix for data sources.

Figure 3: White Settlement and Coffee Production, 1900-2009



Notes: The graph shows the number of "Kenyan Europeans" and coffee production for selected years in 1900-2009. Data is interpolated for missing years. The vertical dashed line indicates the year 1963, when Kenyan gained independence. Most Kenyan Europeans left the country after 1963, as they had to acquire Kenyan citizenship to stay. Coffee production declined from 1986 as a result of falling world coffee prices and exhausted coffee soils. See Online Data Appendix for data sources.

Figure 4: Effects of Railroads on Population for Each Period, Controlling for Population in the Previous Period, 1962-2009



Notes: The graph shows the effects of repeated cross-sectional regressions (see equation (3)) where the dependent variable is urban (total) population at period t and the variable of interest is a railroad dummy equal to one if the location is less that 10 (20) km from a railroad line, while simultaneously controlling for urban (total) population at period t-1. The previous period is 1901 for the year 1962. We cannot control for past total population in 1962, as there was no exhaustive census before. We include various measures of pre-existing settlement patterns. We do not have total population data in 1969 and 2009. The previous period is then 1962 for the year 1979, for this dependent variable.