CENTRE FOR THE STUDY OF AFRICAN ECONOMIES

CSAE Working Paper WPS/2014-04

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

Remi Jedwab† and Edward Kerby‡ and Alexander Moradi§

This Version: January 2014

Abstract: Little is known about the extent and forces of urban path dependence in developing countries. Railroad construction in colonial Kenya provides a natural experiment to study the emergence and persistence of this spatial equilibrium. Using new data at a fine spatial level over one century shows that colonial railroads causally determined the location of European settlers, which in turn decided the location of the main cities of the country at independence. Railroads declined and settlers left after independence, yet cities persisted. Their early emergence served as a mechanism to coordinate investments in the post-independence period, yielding evidence for how path dependence influences development.

Keywords: Path Dependence; Urbanisation; Transportation; Colonialism
JEL classification: R11; R12; R40; O18; O33; N97

*We are grateful to David Albouy, Hoyt Bleakley, François Bourguignon, James Fenske, Richard Hornbeck, Stephen Smith, Adam Storeygard, Anthony Yezer and seminar audiences at the African Economic History Network Meeting (The Graduate Institute in Geneva), ASSA meetings (Philadelphia), Cape Town, the Economic History Association Meeting (Washington D.C.), George Mason-George Washington Economic History and Development Workshop, George Washington, London School of Economics, Maryland, the Royal Economic Society conference (Cambridge), Stellenbosch, Sussex and Wits for helpful comments. We are grateful to Dominic Bao for excellent research assistance. This paper was previously circulated as "*Out of Africa*: Colonial Railroads, White Settlement and Path Dependence in Kenya". We gratefully acknowledge the support of the Institute for International Economic Policy and the Elliott School of International Affairs at GWU.

†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).

§Corresponding Author: 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.

UNIVERSITY OF OXFORD

Centre for the Study of African Economies
Department of Economics · University of Oxford · Manor Road Building · Oxford OX1 3UQ
T: +44 (0)1865 271084 · F: +44 (0)1865 281447 · E: csae.enquiries@economics.ox.ac.uk · W: www.csae.ox.ac.uk

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 existing literature on path dependence has been divided. It has argued over whether locational fundamentals (i.e., the geographical endowments of various locations) or increasing returns (i.e., localised 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. Davis & Weinstein (2002, 2008), Bosker et al. (2007, 2008) 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. Davis & Weinstein (2002, 2008) argue that the fact that Hiroshima and Nagasaki recovered their population and industries supports the first theory. There is only one spatial equilibrium, and regional policy interventions cannot undo it in favour of another one. Yet the literature has also shown how localised events can have permanent effects (Krugman, 1991*a*; Bosker et al., 2007, 2008; Redding, Sturm & Wolf, 2011; Bleakley & Lin, 2012; Holmes & Lee, 2012; Duranton & Maystadt, 2013; Michaels & Rauch, 2013). These studies suggest the existence of multiple spatial equilibria, which implies that regional policies 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 is better explained by either physical geography or history; in particular historical institutions. 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; Beeson, DeJong & Troesken, 2001; Bosker & Buringh, 2010; Maloney & Caicedo, 2012; Motamed, Florax & Matsers, 2013). Conversely, various articles have demonstrated that historical shocks may shape the patterns of development for a given physical geography (Acemoglu, Johnson & Robinson, 2001, 2002; Nunn, 2008; Dell, 2010, 2012). The question remains - how does history and path dependence interact with the process of economic development? To shed light on this, conceptualise a poor, overwhelmingly rural country. Locational fundamentals and increasing returns can explain why an urbanisation equilibrium emerges, i.e. a stable network of cities, and why it can therefore persist.

The *emergence* of an equilibrium can be due to locational fundamentals, such as being situated on a fertile plain or the proximity to a coastline or a river, etc. Natural advantages are often the main determinants of spatial economic patterns at the beginning of the development process (Beeson, DeJong & Troesken, 2001; Davis & Weinstein, 2002; Bosker & Buringh, 2010; Bleakley & Lin, 2012). A historical shock can also draw people together in a few locations (Nunn & Puga, 2012; Michaels & Rauch, 2013). People then stay if there are strong local increasing returns.

The *persistence* of an equilibrium could be due to *serial correlation*, if locational fundamentals have a continuous effect on spatial economic patterns, or *path dependence*, if increasing returns solidify the early advantages of the existing cities (Krugman, 1991*c*; Davis & Weinstein, 2002). There are various tests for the hypothesis of path dependence. First, wars are sufficiently large, negative shocks that we should expect would have permanent effects. If decimated cities recover their population, locational fundamentals must be the main determinants of spatial development (Davis & Weinstein, 2002). Second, the fact that cities arising as a result of a temporary advantage persist after this advantage is eroded provides evidence for increasing returns (Redding, Sturm & Wolf, 2011; Bleakley & Lin, 2012). Krugman (1991*b*) has contrasted the role of "history" and "expectations" in explaining increasing returns. Cities are costly to build. Since fixed costs are a source of increasing returns, *sunk investments* ("history" in the Krugmanian sense) may account for path dependence. Second, given returns-to-scale in production, factors must be co-located in the same locations. There is a *coordination problem* as it is not obvious which locations should have the factors. The newly established cities solve coordination failures for the first period, and continue for each subsequent period.

The *optimality* of an equilibrium depends on whether it is possible to increase welfare by selecting another equilibrium. If there are multiple spatial equilibria, a persistent equilibrium may not be the one that would have been optimal given the distribution of locational fundamentals. Nunn & Puga (2012) and Michaels & Rauch (2013) show how a historical shock may trap people in suboptimal locations.

While more empirical evidence is needed on the mechanisms by which an equilibrium can emerge and persist, few natural experiments have been forthcoming that allow these mechanisms to be investigated. The study of how path dependence emerges also implies that adequate data for a sufficiently long enough period is available. Moreover, studies of persistence require specific data on the channels of the path dependence. As a result of this data scarcity, little is known about the nature of path dependence in the specific context of developing countries. The extent (the relative importances of locational fundamentals and increasing returns) and forces (the relative importances of sunk investments and coordination failures) of path dependence could be different for poorer and more rural countries. Our research question is all the more important because cities are essential in the process of development (Henderson, 2010; Duranton, 2013). Unfortunately, these countries are not only poor, but suffer from a paucity in data, limiting our understanding of the relationship between history, path dependence and development.

To address these difficulties, we use a natural experiment and an extensive new data set on railroads, European settlement and city growth at a fine spatial level over one century (473 locations, *approx.* 16 x 16 km, from 1901 - 2009). All of Kenya's railroads were built during the colonial era. Our results find that they had a strong causal impact on the location of European settlers. Economic development in the European areas in turn determined the location of Kenya's main cities at independence. Railroads subsequently fell into disrepair and most settlers departed in the immediate post-independence period. Yet colonial railroads had long-term, permanent effects on urban and development patterns. We use this setting to show how a historical shock can create an urban equilibrium, which then persists as a result of path dependence. The Kenyan example is compelling for three reasons:[1]

First, we provide evidence that the placement of the railroad was exogenous to the population growth. Transport costs in Kenya were extremely high one century ago. Kenya lacked navigable waterways and draft animals were not used due to the Tsetse fly. There were only a few tracks, that were not developed into roads before 1945. Kenya only exported high-value goods which were headloaded, or slaves who were walked to ports. Economic change was restricted to the coast. For geo-strategic reasons in 1901, the British built a railroad from Mombasa, on the coast, to Lake Victoria, in the west. Although the line crossed Kenya from east to west, the ultimate objective was to link the coast to Uganda (across Lake Victoria), where the Nile originates; the control of which was regarded as a vital asset in the Scramble for Africa. Our results show that the railroad (and its branch lines) had a strong, unexpected impact on European settlement, establishing cities from where they managed their farms and specialised in urban activities. The African population also increased along the lines, working mainly for the settlers.

Second, even if the placement was not exogenous, Kenya's history yields various identification strategies we exploit to confirm our effects are causal. These alternative strategies find no effects for a set of "placebo lines" that consist of branch lines that were planned but not built and alternative explorer routes from the coast to the west of the country. What's more, we have data at a fine spatial level allowing for the inclusion of ethnic group fixed effects, district fixed effects, and 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 contribute similar results. Lastly, even if the main line was exogenous, the branch lines were potentially endogenous, as the coloniser sought to connect specific areas with high economic potential. Endogeneity is only a concern if we find larger effects for the branch lines, which we do not.

Third, Kenya's railroads declined in the immediate post-independence period, mainly due to mismanagement and lack of maintenance in the sector, but also as extensive investments in sealed roads escalated. While few locations of the hinterland were connected to the coast in the colonial period (1895-1963), almost all locations are now connected, and an entirely different spatial distribution could have emerged post-1963. Subsequently Europeans were required by the first president after independence to choose their citizenship. Their refusal to give up the British citizenship resulted in an exodus from Kenya. Coffee was then the main export of the colony, and the engine of growth in the European areas. Coffee production collapsed in the 1980s, owing to lower international prices. Locations along the railroad lines thus lost their initial advantage in terms of transportation, human capital and agriculture. Yet today, these locations remain relatively more urban and developed. This effect is not explained by changes in transport technology, as measured by roads today, but is due to path dependence. Thus the long-run effects of colonial railroads are explained by colonial urbanisation. We use historical data on infrastructure and find that, while colonial sunk investments matter, railroad cities mainly persisted

[1] As argued by Lucas (1988), cities are the main engines of growth. Urban growth is our primary measure of economic development, consistent with the existing literature (Bairoch, 1988; Acemoglu, Johnson & Robinson, 2002; Bosker & Buringh, 2010; Dittmar, 2011).

because their early emergence served as a mechanism to coordinate investments in subsequent periods. This shows the stability of the equilibrium if, for example, regional policies cannot affect expectations as much as infrastructure.

This study makes a valuable contribution to the literature. For example, the Kenyan natural experiment allows us to examine the causal emergence and persistence of an urban equilibrium. We study the long-term effects of a temporary man-made advantage, in the form of an investment in transport infrastructure. Fujita & Mori (1996), Behrens (2007), Redding, Sturm & Wolf (2011) and Bleakley & Lin (2012) have also examined how such investments may have permanent spatial effects. We find strong evidence for path dependence, in line with the pre-eminent literature, but in contradiction with Davis & Weinstein (2002, 2008) and Miguel & Roland (2011). Our data set also serves to disentangle the mechanisms of path dependence. As in Bleakley & Lin (2012), sunk investments appear not to be the main force of path dependence, but rather coordination problems of contemporary factors.

Our focus then changes to the question of a developing country, for which the consequences, extent and forces of path dependence could differ. First, studying an initially unurbanised country allows us to analyse the emergence of the equilibrium, whereas many studies have only tested if an equilibrium persists in the face of a shock. Kenya provides this case, as it had only four cities before colonisation. Second, locational fundamentals should play a larger role in agrarian countries. However, any spatial shock could have substantial effects in rural countries. These effects will be permanent if there are strong increasing returns. Industrial agglomeration effects and human capital externalities may nevertheless be more limited in such contexts. In Kenya, the employment share of industry is low (*approx.* 5%), as most urban workers are employed in the non-tradable service sectors. Kenya also has one of the lowest literacy rates in the world. Third, sunk investments may be less important in poor countries, as the materials used to build cities are cheaper. Yet, since there are fewer investments per capita, locations benefiting from an early start may enjoy their initial advantage for a longer period. The coordination problem is then less important if the expected forces of agglomeration are more limited. The fact that these countries are poorer may nonetheless imply that many coordination problems may as yet not been solved, which constrains capital accumulation. Spatial investments then permanently solve coordination problems.

Our findings also advance the literature on transport technology and growth. Trade costs were extremely high in Africa one century ago. They are still very high today (Atkin & Donaldson, 2012). Poor infrastructure is often cited as the main obstacle to trade expansion and growth in Africa (Rodrik, 1998; Buys, Deichmann & Wheeler, 2010). Transport infrastructure can indeed facilitate the circulation of goods, people and ideas. Railroads boosted exports in Kenya, consistent with the literature on transportation and trade (Michaels, 2008; Duranton & Turner, 2012; Donaldson, 2013; Faber, 2013; Donaldson & Hornbeck, 2013). Investment has encouraged the movement of workers and firms, in line with the literature on transportation, population and employment growth (Baum-Snow, 2007; Atack et al., 2010; Banerjee, Duflo & Qian, 2012; Ghani, Goswami & Kerr, 2012). It has promoted the diffusion of innovations, here the establishment of cities and the adoption of new crops, consistent with literature on information and communication technology and development (Jensen, 2007; Dittmar, 2011). Investment can trigger an equilibrium in which cities emerge to facilitate the accumulation of factors, and thus promote long-term development (Bleakley & Lin, 2012; Jedwab & Moradi, 2013). However the literature examining the causal impact of transportation in Africa is sparser, with the exception of Storeygard (2012). Burgess et al. (2013) find that road building is driven by political prerogatives rather than economic considerations.[2]

Lastly, our focus on colonial transportation is associated with the literature on the impact of colonisation 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; Dell, 2010; Iyer, 2010; Albouy, 2012), while the effects of colonial investments have been overlooked. Second, the studies that examined colonial investments highlighted the role of human capital (Glaeser et al., 2004; Huillery, 2009; Wantchekon, Novta & Klasnja, 2013). The effects of investments in transport infrastructure may have been as large (or even larger). Finally, we use African panel data at a fine spatial level over one century.[3]

The paper is organised as follows. Section 2 offers a conceptual framework to guide the empirical analysis. Section 3 presents the historical background and the data used. Section 4 shows the emergence of the urban equilibrium. Section 5 studies the persistence of this equilibrium. Section 6 concludes.

2. CONCEPTUAL FRAMEWORK

In this section, we offer a framework to analyse the relationship between history, path dependence and development. We reinterpret the literature to describe how an urban equilibrium, i.e. a network of cities, can emerge (termed *"urban emergence"*) and why it can persist (termed *"urban persistence"*). We focus on the respective roles of locational fundamentals and increasing returns in shaping economic geography, and explain how our natural experiment fits into this framework.

2.1 Locational Fundamentals vs. Increasing Returns

Assume an economy with many locations characterised by increasing returns. The population P of each location l at period t is a positive function of its locational fundamentals G_l ($f'_G > 0$) and population in the previous period $P_{l,t-1}$ ($f'_P > 0$):

$$P_{l,t} = f(G_l; P_{l,t-1}) \quad (1)$$

Increasing returns ($f'_P > 0$) can give rise to multiple spatial equilibria. Then, a temporary advantage due to G or P has a persistent effect by creating a stimulus shifting local population density to a higher equilibrium. Initially, when a country is poor, population densities are low, and no urban equilibrium (a set of high Ps) exists. The fact that no cities emerged could result from a bad geography (G) or a low economic return (f'_G) to that geography given the available technologies.

2.2 Urban Emergence

The emergence of an urban equilibrium can be due to locational fundamentals or increasing returns. Locational fundamentals are "first nature" or "natural" advan-

[2] Other studies on transport infrastructure in 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).

[3] We also contribute to the literature on the historical roots of African underdevelopment: Nunn (2008), Huillery (2009), Fenske (2010, 2013), Nunn & Wantchekon (2011), Michalopoulos & Papaioannou (2011, 2012), Nunn & Puga (2012), Heldring & Robinson (2012), Jedwab & Moradi (2013), Bonfatti & Poelhekke (2013) and Wantchekon, Novta & Klasnja (2013).

tages, such as the proximity to a coast or a river, being situated in a fertile plain, etc. We know that the rise of complex civilisations was conditioned by their geographical environment (Bairoch, 1988; Diamond, 2011; Motamed, Florax & Matsers, 2013). Likewise, the location of cities in Europe and the U.S. was strongly influenced by natural characteristics (Krugman, 1991*a*; Beeson, DeJong & Troesken, 2001; Bosker & Buringh, 2010; Bleakley & Lin, 2012). Increasing returns are permitted by "second nature" or "man-made" advantages, such as the existence of a city or being connected to man-made transport networks. An equilibrium can emerge without good locational fundamentals if a historical shock draws people together in a few locations. If there are strong increasing returns (f_P' is large enough), the locations will keep growing. Colonisation provides such a shock, as colonial powers disproportionately invested in a few locations. For instance, the construction of the colonial railroad gave an early advantage to selected locations. Nunn & Puga (2012) and Maloney & Caicedo (2012) also use colonisation as a spatial shock.

2.3 Urban Persistence

An urban equilibrium that persists over time could be a result of multiple factors. It could be due to serial correlation, the fact that locational fundamentals have a continuous effect on spatial economic patterns, or path dependence, the fact that increasing returns solidify the initial advantages of the existing cities. For example, many cities in developed countries are located along rivers or the coastlines, two natural advantages in terms of transport, even though trade costs have decreased considerably over time. This persistence is still explained by serial correlation if these locational fundamentals are now valued as consumption amenities (i.e. waterfront real estate) instead of production amenities (i.e. harbours) (Rappaport & Sachs, 2003). If these natural advantages no longer matter, increasing returns must account for persistence (Krugman, 1991*c*; Bleakley & Lin, 2012).

There are various tests for the path dependence hypothesis. First, if serial correlation explains persistence, a shock should have no permanent effect. Various studies have examined the long-term effects of conflict on population and development. For example, the war time bombing of Japan, Germany and Vietnam provide such a test. The fact that many destroyed cities recovered their population and industry indicates that serial correlation accounts for urban persistence (Davis & Weinstein, 2002, 2008; Miguel & Roland, 2011). In contrast, Bosker et al. (2007, 2008) and Redding, Sturm & Wolf (2011) for Germany, and Dell (2012) for Mexico, find a permanent effect of conflict on development. Second, if path dependence explains persistence, a geographical shock causing a change in locational fundamentals should have no permanent effect, as increasing returns will "protect" the affected locations. In the past, deforestation and prolonged droughts have undone various urban equilibria, such as the urban networks of the Fertile Crescent, the Maya Civilization and the Mali Empire (Bairoch, 1988; Diamond, 2011). Increasing returns are more important in the modern period, and the loss of a natural advantage may have lower effects today (Bosker & Buringh, 2010; Bleakley & Lin, 2012). Lastly, whether we use changes in G or P to test the path dependence hypothesis, the persistence of an equilibrium will depend on the unpredictability, magnitude and duration of these changes. For the shock in G or P to have a permanent effect, we need its effect to dominate the positive effect of the other variable (P and G respectively).

There are two main mechanisms of path dependence (Krugman, 1991*b*; Bleakley & Lin, 2012). Firstly, given that fixed costs are a source of increasing returns, *sunk* *investments* (e.g., buildings, schools, hospitals and roads) could account for path dependence. Secondly, 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. Past population P_{t-1} could matter ($f_P' > 0$) because it proxies for past investments, or because it sends a signal to all factors that they can simultaneously co-locate in this specific location. It will be important later on to disentangle the effects of both these channels. Indeed, to what extent an equilibrium will persist is likely to depend on which force dominates. If sunk investments are the main channel of path dependence, an equilibrium can be undone by strongly investing in infrastructure in another location (depending on how expensive the necessary investments are). Expectations may prove more difficult to adjust, as it requires forceful or collective agreement in relocation. Therefore, even if non-labour factors were perfectly mobile, an equilibrium would still persist for the mere reason that expectations can be self-fulfilling.

2.4 Urban Optimality

If there are multiple spatial equilibria, the existing equilibrium may be ex-post suboptimal. For example, if locational fundamentals change over time, an equilibrium that was created as a result of these fundamentals at period t-1 may be suboptimal at period t given the new fundamentals. However, increasing returns will magnify the advantage obtained in the previous period. Similarly, an equilibrium arising from a historical shock is suboptimal if it does not match the distribution of fundamentals. The created cities will persist thanks to increasing returns, although these locations were not a priori supposed to be so developed. In both cases, a social planner would be able to increase overall welfare by spatially reallocating people to the fundamentally best locations. Nunn & Puga (2012) find that various African populations permanently moved to rugged areas to protect themselves against raids during the slave trade. As ruggedness has a negative impact on trade and development, this spatial distribution is suboptimal today. Likewise, Michaels & Rauch (2013) show that many French towns were originally founded along the roads of the Roman Empire. These towns persisted in the Middle Ages, even when improvements to water transport should have made it more profitable to have these towns located along navigable waterways. There are two caveats to this analysis. First, natural advantages that condition the emergence of a city, may have no effect on the growth of that city once it is created. With increasing returns and technological change, cities emancipate themselves from their surrounding geographical environment (e.g., Dubai). Second, we need to take into account the number and size of cities, instead of just comparing the fundamentals of existing and "counterfactual" cities. Kenya had few cities one century ago. Without the colonial railroad, many cities would have only emerged a few decades later. Given increasing returns, the fact that these cities emerged earlier had positive effects on the accumulation process of the whole economy. Thus, even if the railroad cities had worse fundamentals than a set of counterfactual cities, the railroad has increased urban growth.

2.5 Colonial Railroads in Kenya as a Natural Experiment

Railroad construction in colonial Kenya permits us to study the emergence and persistence of an equilibrium: (i) *Urban emergence*: Kenya had few cities before colonization (low P), probably as a result of its bad geography (low G) and poor technology (low f_G') then. The railroad disproportionately favoured locations that became the main cities at independence ($\Delta P > 0$). (ii) *Urban persistence*: The railroad locations lost their initial advantage post-independence, which should have led to relative declines in their population ($\Delta P < 0$). Yet we find that these locations are relatively more developed today (P remains high). Does our natural experiment provide a lower test of the path dependence hypothesis than the rest of the literature? Firstly, as in prior literature, war bombing doesn't provide the ultimate test. While suffering population loss and building damage, many sunk investments such as road networks, or the reputation of a city (that contributes to solving the coordination problem), may have not been affected. Bosker et al. (2008) show that the evidence for multiple equilibria is weaker when not taking into account network effects. Besides this, we do not observe a pure market equilibrium in the post-war period if the governments of these countries massively reinvested in the destroyed cities (Miguel & Roland, 2011). Secondly, studies on one sector only may lack external validity. For instance, Redding, Sturm & Wolf (2011) focus their analysis on the airline industry, for which strong increasing returns were clearly expected. Results could have differed for other sectors. Lastly, while conflicts constitute negative shocks to economic geography, investments in transport infrastructure constitute positive shocks. Changes in transport technology may then considerably erode the initial advantage of the locations using the old technology. Bleakley & Lin (2012) show how portage sites have become major cities in the 19th century U.S. as water transport was important then. Although portage declined post-1880, the portage locations remain more developed today. We show how a man-made advantage (railroad construction) has created an equilibrium that persisted after this advantage was eroded (in the post-independence period). Our test may not be as "neat" as Bleakley & Lin (2012), since the relative decline of railroads only started fifty years ago. Yet, this decline was a large, unexpected shock. As the perfect test does not exist, we believe this provides a valuable, robust contribution to the literature.

3. RAILROADS AND PATH DEPENDENCE: BACKGROUND AND DATA

In this section we discuss the historical background and data used in our analysis. The Online Data Appendix contains more details on how we construct the dataset. The following summaries draw on Hill (1950), Morgan (1963), Hazlewood (1964), Soja (1968), King & van Zwanenberg (1975), Hornsby (2013) and Burgess et al. (2013), as well as data that we compiled ourselves from various official documents.

3.1 New Data on Kenya, 1895-2010

In order to analyse the effect of rail building on development, we construct a new data set of 473 locations, the third level administrative units (with a median area of 256 sq km, *approx.* locations of 16 x 16 km), for the following years: 1901 (six years after the establishment of the East Africa Protectorate), 1962 (one year before independence), 1969, 1979, 1989, 1999 and 2009 (the census years). We obtain the layout of rail lines in GIS from *Digital Chart of the World*. We then use various documents to recreate the history of rail construction. We know when the main line and each branch line was completed. Our analysis focuses on the rail network in 1930, as thereafter it did not change. We then identified explorer routes that provide a good counterfactual for the mainline (from the coast to Lake Victoria). We also located branch lines that were planned but not built. For each rail or "placebo" line (i.e., each explorer route and branch line), we create dummies equal to one if the Euclidean distance of the location centroid to the line is 0-10, 10-20, 20-30 or

30-40 km. We proceed similarly to merge the GIS road database from Burgess et al. (2013). We then use census gazetteers to reconstruct a GIS database of localities above 2,000 inhabitants. The number of these localities increased from 5 in 1901 to 42 in 1962 and 247 in 2009. Since our analysis is at the location level, we use GIS to construct the urban population for each location-year observation. While we have exhaustive urban data for all the years mentioned above, we only have georeferenced population data in 1962, 1979, 1989 and 1999. Rather unfortunately, the population census of 1962 was the first census for which the African population was recorded. To proxy for African population in 1901, we digitised a map of historical settlement patterns that shows the location of the main sedentary and pastoralist groups in the 19th century. From census data, we obtain the number of Europeans and Asians for each location in 1962. We also digitised and geospatialised voter registries of the Europeans in 1919 and 1933, which we use as proxies for the number of Europeans the same years. The voter registries also contain information on the occupation of these settlers. Locations then do not have the same area, so we will control for location area in the regressions. Lastly, we have data on commercial agriculture (e.g., coffee and wheat cultivation) in 1962. We also have data on infrastructure provision (e.g., schools and hospitals) and economic development (e.g., poverty rates and satellite night-lights) at the location level in 1962 and 2000.

3.2 The Railroad Age in Kenya, 1895-1963

Infrastructure investments are typically endogenous, driven by the economic potential that justifies them. Hence, a simple comparison of connected and non-connected locations could overstate the output created by the railroad. The British coloniser established the East African Protectorate in what is now Kenya, Tanzania and Uganda in 1895. Improving transport infrastructure in the region was key; to permit military domination and boost trade, historically constrained by high transport costs. Uganda was the most advanced and economically promising colony. The British thus wanted to link Uganda to the coast. Kenya, which lies in-between, was simply a transit territory. The main railroad of Kenya, from Mombasa on the coast to Lake Victoria in the west (see Figure 1), was even called the "Uganda Railway". Thus, rail construction to Uganda via Kenya provides an ideal natural experiment to identify the causal effects of transport infrastructure on economic change.

Trade costs before the rail: Transport costs were extremely high at the turn of the century. Draft animals had not been adopted due to the Tsetse fly trypanosomiasis transmitting, and Kenya lacked navigable waterways. Headloading was the main, and very costly means of transport prior to the railways. Only high value goods were carried through the hinterland, while slaves could be walked. The main caravan route from Lake Victoria to the coast did not pass through Kenya but rather Tanzania in the South, since Zanzibar was the main port of Eastern Africa. The McKinnon-Sclater road, an ox cart track from Mombasa to the Uganda border, was established in 1890 by the Imperial British East Africa Corporation (IBEAC), a royal charter company tasked with the administration of the Protectorate. However, the road did not reach Uganda before 1896, the year when rail construction began. Moreover, the road did not meaningfully reduce trade costs, as the journey was slow and difficult. Railroads catalysed a transport revolution.

Main line: Suffering financial losses, the IBEAC ceded control over Uganda and Kenya to the British Government in 1896 and the mainline (from Mombasa to Lake Victoria, see Figures 1 and 2) was constructed between 1896-1901. It was built for three principal reasons. Firstly, for strategic and geopolitical reasons. The line shielded the region against competing European powers, by allowing the fast transportation of troops. Lake Victoria is also the source of the Nile River, and the British thought that by linking Uganda to the coast they could unify all their colonies in Northern, Eastern and Southern Africa (Appendix Figure 1 shows the map of the "Cape to Cairo Railway", a plan to unify British Africa from south to north by rail). Secondly, Uganda was seen to hold vast wealth with further trade potential. Linking Lake Victoria to the coast would open up Uganda by reducing trade costs. Thirdly, it had a deemed civilising mission. The construction was debated fiercely within the British parliament. Critics doubted the usefulness of the railroad "from nowhere to utterly nowhere". Since Kenya was a transit country, the itinerary of the line was chosen in order to minimise its construction costs, i.e. the rail distance from Mombasa (the largest city of Kenya then, with 15,000 inhabitants) to Lake Victoria. The line was first built to Nairobi, an uninhabited swamp that traces its urban origin as a railroads depot in 1899 but later became the railroads headquarters and the country's capital. Nairobi was chosen because it was a water hole that could supply the rail construction workers with water. The line then went to Kisumu on Lake Victoria, via Lake Nakuru, another source of water along the route.[4]

Branch lines: The mainline established the general urban pattern of Kenya. Soja (1968) explains that the equal distribution of urban centres at key points along the main route reflects the weak influence of local factors in the initial urban growth. The interior nodes increased in size and importance as various branch railroad lines were constructed (see Figures 1 and 2): Thika (1913), Magadi (1915), Taveta (1918), Eldoret (1926), Kitale (1926), Solai (1926), Tororo (1928), Thomson Falls (1929), Nanyuki (1930) and Butere (1930). No railroad was built post-1930. While the placement of the mainline could be considered as exogenous to future population growth *within* Kenya, the branch lines were potentially endogenous, as the colonial government sought to connect areas of high economic potential. The branch lines were actually not profitable, which could question the ability of the government to appraise the economic potential of various areas at the time.

Reduced trade costs: Transport costs tremendously decreased in the hinterland. Figures cited in Hill (1950) imply a 1902 freight rate of 11 shillings (s) per ton mile for head porterage as compared to 0.09s per ton mile by railroad on the main route, hence a reduction of transportation costs of at least 1,000%. Trade costs along the Mc Kinnon-Sclater ox cart road were probably not much lower than for head porterage. When they were not infected with trypanosomiasis and attacked by lions, oxen could only be used for some sections of the road, on which they advanced slowly. The road fell into disuse as soon as the railroad was built. The caravan route from Lake Victoria to Zanzibar was then ca. 3.5s per ton mile, but the route was also much longer than going directly to Mombasa.[5] Pre-railroad trade costs were thus prohibitively high. For example, the cultivation of export crops would never have

[4]While the proximity to these water holes influenced the placement of the railroad, it had no impact on future agriculture or population growth. Firstly, the water hole in Nairobi was not adequate for commercial agriculture, whereas Lake Nakuru became a National Park, which restricted the use of its water. Secondly, agriculture in Kenya is principally rain-fed and does not use irrigation. Thirdly, we will show that the results are robust to controlling for the distances to the nodes.

[5]Head porterage rates were higher in Kenya than in other African countries ca. 1910 (11s per ton mile), e.g. Ghana (5s per ton mile), Malawi (3s), Nigeria and Sierra Leone (2.5s). Rail freight rates were then lower, at 0.09s per ton mile vs. 0.80s, 0.19s and 0.27s per ton mile in Ghana, Nigeria and Sierra Leone respectively (Chaves, Engerman & Robinson, 2012; Jedwab & Moradi, 2013).

boomed without modern transport technology. Coffee could only be produced west of Nairobi. Given the export price in Mombasa (1,300s per ton in the early 1910s), the distance from Nairobi to Mombasa (311 miles) and head porterage costs (11s per ton mile), coffee production was not profitable at all ex-ante (even if production costs were nil). Then, given the railroad distance between the two locations (318 miles) and railroad costs (0.09s per ton mile), coffee production was profitable as long as production costs did not reach 98% of the export price. Railroads were thus essential to the economic colonisation of the hinterland.

Colonial roads: At first roads were complementary to the rail as they were feeders to it. The colonial government sought to protect the railroad by under-investing in roads. The existing roads were of poor quality until the 1950s. For example, the bitumenisation of the road between Mombasa and Nairobi only started in 1945. Roads later became serious competitors for the rail. Even if no railroad had been built, roads would have certainly permitted the economic boom brought by the railroad, but not before post-WW2. This makes a difference of forty years. However, our objective is not to compare the respective impacts of railroads and roads. We focus on the "railroad age" (1895-1963) because it provides us with a natural experiment to identify the impact of modern transport technology (vs. no transport technology at all) on economic development and path dependence.

Placebo lines: To address endogeneity concerns, we have identified various routes that provide a good counterfactual for the railroad lines. We use these "placebo lines" as a placebo check of our identification strategy. Figure 2 shows the geographic location of these placebo lines, while the Online Data Appendix extensively describes each of these routes. First, several explorer routes (from the coast to Lake Victoria) provide a good counterfactual for the mainline. Various segments of these routes could have been alternatively selected to become a segment of the mainline. The explorer routes then traversed areas with better locational fundamentals. Several factors influenced the course of the routes such as safety, distance and the provision of water and food. They often went through more densely populated areas. As we will show later, the economic potential of these "placebo lines" was better than for the mainline. Comparing the growth patterns of the locations along the railroad lines and the placebo lines should thus lead to a downward bias. Second, we use various branch lines that were proposed in 1926 but not built. As described in the Online Data Appendix, despite vested colonial interests, these extensions failed to materialise for either economic viability or cost in construction.

3.3 Patterns of Economic Development in Kenya, 1895-2010

Kenya was one of the poorest countries in the late 19th century. Slaves, hides, rubber and ivory were its main exports. With the exception of Nairobi, founded in 1898, there were only four localities with a population in excess of 2,000 inhabitants in 1901, all of which were coastal port towns. The hinterland was devoid of cities, and the various non-coastal tribes, such as the Kikuyus, the Kalenjins, the Luhyas, the Luos and the Maasais (see Figure 1), were very poor. These ethnic groups were geographically separated by the Rift Valley, that served as a buffer zone between them. The creation of the Protectorate in 1895 established peace over the whole country. Yet economic activity was constrained by high trade costs.

The construction of the railroad to Uganda dramatically changed the economy of Kenya, and made it one of the richest African countries at independence. The peculiarities of railroad placement led to the curious situation that the railroad traversed sparsely settled areas with no (Kenyan) freight to transport. European settlement was later encouraged to create an agricultural export industry, which would increase railroad traffic and make the railroads profitable. Land was alienated and offered to European settlers, whose number increased to almost 60,000 inhabitants at independence (see Figure 3). There were two groups of settlers. The farmers settled in the countryside, whereas the civil and railway servants, merchants and professionals settled in the cities. According to the 1919 Voter Registry, 50% of the settlers were farmers. The European census then indicates that this share decreased to 25% while the share of civil and railway servants increased to 40% in 1948. The shares of merchants, professionals and missionaries remain stable.

The export of cash crops was the engine of Kenya's development. In the early colonial period, coffee and maize accounted for 40% and 20% of exports respectively. Coffee was introduced by missionaries in 1893. It took 30 years before coffee was widely grown by Europeans. In the later period, maize exports declined to 5%, while tea exports increased to 10%. The main food crop for Europeans was wheat, whose production expanded. Then, it was not until the Swynnerton Plan of 1954 that African farmers were all allowed to grow coffee. Kenya then imported manufactured goods and fuels. Most goods were transported by rail, and railroad traffic rose dramatically as shown by two different measures in Figure 3. The first measure – *tons per km per million of population* – is the volume of goods transported by rail per capita, measured in metric tons per km of rail network and per million of population. Total traffic could be a function of population, hence the need to account for population growth. The second measure – *million ton km* – is the volume of goods transported by railway, measured in metric tons times kilometres traveled.

The number of localities with a population in excess of 2,000 inhabitants increased to 42 at independence, as Europeans established cities serving various functions. First, they lived in towns from where they managed their farms. Second, the cities were trading stations through which export crops were transported to the coast and imported goods were dispatched from the coast. Asians who emigrated to Kenya to build the railroad also established themselves as merchants in these cities. Third, part of the surplus generated by agricultural exports was spent on locally produced urban goods and services. Lastly, many towns became administrative seats. While soldiers and policemen accounted for 1% of European settlement in these areas, the number of bureaucrats increased to respond to the needs of the farmers. The areas where Europeans settled to grow crops or specialise in urban activities came to be known as the White Highlands. Demand for African labour grew in these areas as they supplied the dominant portion of agricultural and urban labour. African migrant peasants squatted on European farms in exchange for land tenure. Their inbound migration to towns was limited by the "kipande" system which required all workers to obtain a registration certificate from their employer.

Agricultural and migration restrictions were lifted after Kenya gained independence in 1963. The cultivation of European crops diffused to the former African areas, where cities thrived (see Figure 1). The number of towns increased to 247 in 2009. Many Africans also migrated to the estates and cities of the White Highlands. At independence, Europeans were required by the new president Kenyatta to choose Kenyan citizenship. Most of them refused to give up their British citizenship, resulting in an exodus (see Figure 3). Kenya's railroads also fell largely out of use in the 1970s, due to mismanagement and lack of maintenance in the rail sector, as well as considerable road investments in the immediate post-independence period (see Figure 3). Kenya's economy remained specialised in the export of cash crops, but coffee production continuously declined from the mid-1980s, due to low international prices, while tea, mainly grown in the African areas far away from the rail, has become Kenya's main export crop, accounting for 25% of total exports.

The econometric analysis in the following section will show that there were indeed causal effects of the railroads on European settlement, agriculture, urbanisation, and African migration during the colonial period. We will then use the facts that railroads collapsed, settlers left and coffee production declined post-indepence, to test how colonial railroads had lasting long-term effects on Kenya's development.

4. RAILROADS AND ECONOMIC DEVELOPMENT AT INDEPENDENCE

In this section we show that railroads led to economic change during the colonial period (1895-1963). We test if connected locations experienced population and urban growth. We explain the various strategies implemented to obtain causal effects. Additional evidence on the various mechanisms by which railroads spurred population and urban growth is contributed to the analysis.

4.1 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, total, African and Asian populations of connected and non-connected locations (l) in 1962, one year before independence (1963):

$$zPop_{l,1962} = \alpha + Rail_{l,1962}\beta + \omega_p + X_l\zeta + \nu_{l,1962} \quad (2)$$

where $zPop_{l,1962}$ is the standard score of European / urban / total / African / Asian population of location l in 1962. $Rail_{l,1962}$ are dummies capturing rail connectivity in 1962: being 0-10, 10-20, 20-30 or 30-40 km away from a line. The dummies would have been equal to zero in 1901. We expect rail connectivity to have a positive effect on population ($\beta > 0$). We include eight province fixed effects ω_p and a set of controls X_l to account for pre-existing settlement patterns and potentially contaminating factors. We have a cross-section of 473 locations. However, in the case of European, Asian and urban population, which were close to nil in 1901, results should be interpreted as long-differenced estimations for the period 1901-1962. Our main analysis is performed on a sample of locations excluding the areas that are unsuitable for agriculture (see Figures 1 and 2). The 403 non-arid locations belong to the South. If we use the full sample, we run the risk of comparing the southern and northern parts of Kenya, whose geography and history differ. If unobservable factors correlated with the railroad explain why the South was historically more developed than the North, excluding the northern locations should give us more conservative estimates, as it ensures that we are comparing like with like. We will show that our results hold when using the full sample. To make sure that our results are not driven by the three main nodes of the railroad, we drop Mombasa, Nairobi and Kisumu, though the latter two are clearly railroad locations. Identification is thus derived from being connected in between those nodes.

We express all population numbers in standard (z-)scores. Standardising the population variables has two advantages. First, non-standardised values differ greatly (e.g., total vs. European population), so the coefficients β cannot be readily compared across outcomes. Standardised values facilitates the interpretation of the coefficients. Second, total population has been growing over time, so the coefficients will increase for later periods, unless we standardise the variables. Logs could perform the same role. However, there are many locations with a European or urban population equal to 0. Using logs would drop these observations. We will thus use standard scores in the main analysis, but show that the results hold when employing logs too. 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 population densities. We now describe the tests performed to ensure our effects are causal.

4.2 Exogeneity Assumptions, Controls, and Identification

In our analysis, we include various controls at the location level (X_l). For dependent variables whose levels were close to 0 in 1901, the cross-sectional regressions in 1962 should be interpreted as long-differenced estimations for the period 1901-1962. For the other variables (e.g., African population), it is important to control for historical settlement patterns, as there was no exhaustive census before 1962. We add various demography, physical geography and economic geography variables, hoping they capture the initial distribution of population (in 1901). If the variables are adequate controls for the baseline levels, these cross-sectional regressions may also be interpreted as long-differenced estimations. Besides, these factors, and the geographical locational fundamentals in particular, may have also determined the potential for European settlement and driven urban and African population growth post-1901, hence the need for their inclusion in the regressions.

We add physical geography variables such as the share of arid soils (%), the shares of soils (%) suitable for agriculture, coffee or tea, the mean and standard deviation of altitude (m) and average annual rainfall (mm). We include economic geography variables such as a "coastal location" dummy if the location borders the sea, the Euclidean distance to the coast (km), area (sq km), and a "provincial capital" dummy equal to one if the location contains a provincial capital. Lastly, the additional measures of historical settlement are: (i) a "city in 1901" dummy (there were only three non-nodal cities in 1901), and (ii) the area shares (%) of "major settled groups" and "pastoralist groups", and a "isolated groups" dummy.

We test if connected locations and non-connected locations initially differed, using the variables above. Even if we control for these factors in our analysis, a significant difference could mean that the line placement was endogenous. However, since this is not a randomised experiment, initial differences are likely and expected. We regress each control on a dummy equal to one if the location is less than 20 km from the rail, while adding province fixed effects. The results in column (1) of Table 1 show that treated locations are less arid and less rugged (see "Altitude: standard deviation"). This could lead to an upward bias. Yet the coefficient of aridity is small, and the coefficient of ruggedness is strongly reduced when dropping the 18 outlying observations of the 4,000 m high Mounts Kenya, Elgon and Kinangop. Treated locations have less rainfall, are less suitable for tea, and were historically less populated (see "Area share of major settled groups"). This could lead to a downward bias, if we expect fast population and urban growth in denser areas ($f_p' > 0$), or an upward bias, if the railroad locations grew faster because they were lowly-populated ($f_p' < 0$). Even if placement was not exogenous, we exploit various identification strategies to ensure our effects are causal.

Firstly, since we have data for 16 x 16 km locations, neighbouring locations may not differ in terms of unobservables. The location's median area is 256 sq km, approximately that of the Boston metropolitan area. When comparing locations less than 20 km and the locations between 20 and 40 km from a railroad (column (2)), we find that the closer locations are less rugged, which could lead to an upward bias. They have less rainfall, and they are less suitable for tea and initially less dense, which could lead to a downward bias and give more conservative estimates. If these locations do not differ in terms of unobservables, and if the placement is exogenous, the effect should strongly decrease as we move away from the line. As additional tests, we will also include: (i) 21 "ethnic group" fixed effects, as we know the main group of each location before colonisation, in order to compare the effects for neighbouring locations belonging to the same ethnic homelands, (ii) 35 district fixed effects (using the boundaries of 1962), in order to compare the effects for neighbouring locations belonging to a same administrative district, and (iii) a fourth-order polynomial of the longitude and latitude of the location's centroid, in order to flexibly control for demography, and physical and economic geography. The three types of variables control for heterogeneity at a fine spatial level, and identification then comes from variations in the treatment between similar locations.[6]

Secondly, we can compare the connected locations with locations that would have been connected if the placebo lines had been "counterfactually" built. As discussed above, we expect the placebo locations to have better locational fundamentals than the railroad locations. When compared to the placebo locations (column (3)), the railroad locations are indeed of a higher elevation, less suitable for coffee and tea, larger, and historically less populated. This should lead to a downward bias and give more conservative estimates. We will test that: (i) no spurious effects are found for the placebo lines (whether for each of them, or all of them altogether), (ii) the main effects are robust to using only the placebo locations as a control group.

Thirdly, endogeneity concerns particularly apply to the placement of branch lines. Branch lines were built to support European agriculture and settlement in specific areas. They were built later, when an urban system had already emerged. We thus compare branch and main lines in our econometric analysis. We actually find that the branch and main lines do not significantly differ when using the (observable) locational fundamentals described above (not shown). Endogeneity is less a concern if we find similar effects for both, or relatively larger effects for the main line.

While we do not prefer one identification strategy in particular, and while none of these strategies is perfect, we will show that they all support relatively similar results providing comfort that our effects are causal.

4.3 Main Results at Independence

Table 2 shows the main results for population growth, assuming that these cross-sectional regressions for the year 1962 can be interpreted as long-differenced estimations between 1901 and 1962. We find a strong effect of rail connectivity on

[6]The "ethnic group fixed effects" strategy mirrors the work of Michalopoulos & Papaioannou (2011, 2012), who use the fixed effects to study the long-run impacts of exogenous spatial variations in institutions *within* a same ethnic group in Africa. We use the same digitized ethnic map of Murdock to construct our ethnic group fixed effects (see Online Data Appendix). District fixed effects should perform the same role. The "fourth-order polynomial of the longitude and latitude" strategy echoes the work of Dell (2010), who utilises a spatial regression discontinuity to examine the long-run impacts of the *mita*, an extensive forced mining labour system in effect in colonial Peru and Bolivia.

European and urban population growth, but these effects decrease as we move away from the railroad and are zero after 30 km and 10 km respectively (columns (1)-(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 a significant driver of urbanisation (column (3)). Obviously, the effect of European settlement on urban growth is not necessarily causal (0.37**), as the growth of these European cities could also have attracted more European settlers. Yet we think that the correlation is interesting *per se*, as it shows the strong interaction between European settlement and city growth.

There is a strong effect of the railroad on population density up to 20 km (columns (4)-(5)). This is consistent with increased opportunities along the railroad lines attracting African labour. Indeed, there were only 60,000 Europeans at independence. The rising population along the railroad lines must have been due to Africans moving to and being born in these areas during the colonial period. This is confirmed by using African population as the dependent variable (columns (6)-(7)). Including the number of Europeans in the same location captures some of the railroad effects on African population. The remaining effects could be due to Africans settling in the hinterland of the cities to work on European farms (i.e., the neighbouring locations). We also find a strong effect of the railroad on the Asian population, which disappears when adding the number of Europeans (columns (8)-(9)). The Asians only settled in the cities, as they were not allowed to acquire land. The coefficient of correlation between the Asian and urban population was thus 0.90.

Table 3 investigates the further effects of the railroad on European settlement, since the other outcomes are highly correlated with it. Again, we do not claim that urban, total, African and Asian population growth was uniquely and causally determined by where Europeans settled. Yet European settlement, which was causally influenced by rail construction, had a profound impact on the economic geography of the colony. Column (1) of Table 3 replicates the main results on European settlement from column (1) of Table 2. In columns (2)-(3), we find that the railroad effects are partially explained by the fact that the railroad (completed in 1930) had already attracted many settlers by 1933. The railroad had further effects post-1933. As we do not have data on the number of Europeans for each location in 1933, we use the number of European voters for the same year instead, our dependent variable in columns (3)-(5). In column (4), we show that the location of the European farmers explains European settlement. The number of European non-farmers was thus higher where there were more European farmers (not shown). In column (5), we show that most of European settlers in 1933 lived in locations that were already settled in 1919. In columns (6)-(11), we show that the cultivation of European crops (coffee, maize and wheat) has expanded along the railroad lines. The effects are then reduced when controlling for the number of European farmers (columns (7), (9) and (11)). An interpretation of the positive rail effects in columns (9) and (11) could be that maize- and wheat-producing European farms were larger along the lines. We do not find any railroad effect for tea (not shown), as the areas suitable for tea were far from the railroad.[7] The commercialisation of agriculture, whether

[7] We also use district panel data from 1922-32 to see if the construction of branch lines during that period had positive effects on European agriculture. We have data for 22 districts of the non-arid areas and 11 years, hence 242 observations. We run panel regressions by adding district and year fixed effects, as well as province-year fixed effects to account for time-variant heterogeneity at the province level. As these districts are large, we believe that this panel analysis is not as informative as

because railroads reduced trade costs or encouraged the diffusion of crop production techniques, was thus one the mechanisms by which the railroad contributed to economic growth in these areas. Another mechanism could have been the establishment of the "imperial peace" over these areas. As shown in the next section, the railroad effects will also be robust to the inclusion of 35 district fixed effects, which should capture any spatial heterogeneity in terms of pacification.

4.4 Alternative Identification Strategies and Robustness

Table 4 displays the results when we implement the identification strategies. Column (1) replicates our main results from Table 2 (columns (1), (2) and (4)). For the sake of simplicity, we focus on the 0-30 km dummy for European settlement (Panel A), the 0-10 km dummy for urban population (Panel B) and the 0-20 km dummy for total population (Panel C), as there are no effects beyond these distances.

Spatial Discontinuities. The fact that locations are small implies that neighbouring locations may not differ in terms of unobservables. Including 21 ethnic group fixed effects (35 district fixed effects) allows us to control for spatial heterogeneity across ethnic groups (districts). The railroad effects are then identified by comparing treated locations with neighbouring untreated locations. Columns (2) and (3) show that the results hold when adding the fixed effects. We can also include a fourth-order polynomial of the longitude and latitude of the location's centroid, in order to flexibly control for demography, and physical and economic geography. The railroad effects are then identified from spatial discontinuities in the treatment (i.e., the fact that treated locations are more developed than otherwise predicted by general spatial patterns). Column (4) confirms that the effects are similar then.

Branch Lines vs. Main Lines. Even if the placement of the branch lines was potentially endogenous, the main line was built for exogenous reasons. Endogeneity is less a concern if we find similar effects for both lines (or larger effects for the main line). For both European settlement and urban population, we actually find larger point estimates for the main line (column (5)). The European settlement effect is not significant for the branch lines, which indicates that Europeans mostly settled along the main line. For total population, the point estimates are higher for the branch lines, but not significantly so (column (6)). Therefore, we cannot reject the hypothesis that the railroad effects are not relatively lower for the main line.[8]

Placebo Regressions. In columns (6) and (7), we test that there are no spurious effects for the placebo lines. For each dependent variable, we create a placebo treatment dummy equal to one if the location is less than X km from a placebo line (X = 30, 10 and 20 km respectively). The placebo treatment effects are high and significant for European population (panel A of column (6)) and total population (panel C). One issue here is that some of the placebo lines intersect with the area of influence (e.g., 0-20 km) of the existing railroad lines, so that there may be a

---

the cross-sectional analyses using locations. We nevertheless find positive effects of being connected to the rail network on coffee, maize and wheat cultivation (not shown, but available upon request). While we cannot be sure that these effects are causal, they are consistent with the results of Table 3.

[8]As we have data on the number of European voters for each location in 1919 and 1933, we verify that the branch lines that were built post-1919 had positive significant effects on European settlement during the 1919-1933 period (not shown, but available upon request). We also find that the newly built branch lines had no effects on European settlement before 1919 (not shown). In other words, the branch lines only have an effect after they are built. Thus, even if their placement was endogenous, we use the exogenous timing of their construction to identify causal effects.

correlation between the treatment and placebo dummies. We thus verify that there are no significant positive effects when dropping the railroad locations, i.e. the locations that are also less than X km from a railroad line. This allows us to compare the placebo locations with the other control locations, while suppressing the effects from the railroad lines. The placebo effects become nil for European and urban population (column (7)). The coefficient is high but not significant for total population. Indeed, as described in the Online Data Appendix, many of the explorer routes and proposed branch lines went through densely populated (African) areas, that have probably also experienced (non-urban) population growth before independence. The placebo locations may not be a good control group for this specific outcome.[9] In column (8), we use the placebo locations as a control group of the railroad locations. The effects are positive and significant for European and urban population, not for total population. In column (9), we drop the locations where the railroad and placebo lines intersect, i.e. the locations that are less than 10 km from both a railroad line and a placebo line. This allows us to compare the effects of the railroad and placebo lines when they followed clearly different paths (see Figure 2). The European settlement effect remains highly significant (panel A). The urban population effect is not significant at 10% (the p-value is 0.13), but the point estimate is unchanged (0.36 in col. (9) vs. 0.35* in col. (8)). The loss of significance must be due to the low numbers of observations (N = 109) and degrees of freedom given the controls (N = 88). There is no effect for total population, as the African population must have been growing along the placebo lines too.

Robustness checks. The results suggest that the railroad had strong causal effects on European settlement and urban population. The effects on total population also appear to be causal, but the placebo lines have also experienced an increase in population. Table 5 indicate that these results hold if we: (i) drop the coastal province to focus on the areas west of Nairobi (column (1)), (ii) use the full sample of locations including the ones in the arid parts of Kenya (column (2)), (iii) also control for the Euclidean distances to each railroad node (Mombasa, Nairobi and Kisumu) and their squares to account for spatial spillovers from these cities (column (3)), (iv) also include a dummy equal to one if the location belongs to the areas that became officially known as the "White Highlands" (column (4)),[10] (v) drop the controls (column (5)), which reduces the effects on total population, as we cannot control for initial settlement patterns then, (vi) control for whether the location is within 10 km from a paved or improved road at independence (column (6)), demonstrating that railroads do not proxy other transport infrastructure investments, (vii) use logs of the population variables as dependent variables (column (7)), (viii) replace the rail dummies by the Euclidean distance of the location's centroid to the rail and its

[9] Appendix Table 1 shows the placebo effects for each one of the twelve placebo lines individually. We test whether there are no spurious effects just along these placebo lines. That is why we only use a 0-10 km dummy for the three dependent variables. In Panels B, D and F, we drop the locations less than 10 km from a railroad line, in order to compare the placebo locations with the other control locations, while suppressing the immediate effects from the railroad lines. The coefficients are never significant in these panels, although the point estimates can be high for a few lines.

[10] The railroad contributed to determine the future location of the White Highlands, by permitting European migration to specific areas. We run a regression similar to model (1) except the dependent variable is a dummy equal to one if the location belongs to the White Highlands at independence. The coefficients of the four railroad dummies are: 0.27*** for 0-10 km, 0.25*** for 0-20 km, 0.17*** for 20-30 km and 0.08* for 30-40km. European migration to the White Highlands was later further encouraged by the British coloniser. By showing that the results hold *within* the White Highlands, we confirm that the railroad effects are mostly driven by the reduction in transport costs.

square (column (8)), or a dummy equal to one if the location is crossed by the rail (column (9)), and (ix) use Conley standard errors with a distance cut-off of 50 km (about three locations) to account for spatial autocorrelation (column (10)).

Other robustness checks. The results hold when constructing the rail dummies using the Euclidean distance to a rail station in the 1930s (not shown). Secondly, we obtain strong positive effects on settlement and urban growth at the extensive margin. For example, being within 10 km of the railroad increases the probability of receiving settlers and a city by almost 40% and 10% respectively (not shown). Then, conditional on receiving settlers or a city, the locations along the railroad are denser in terms of European or urban inhabitants (not shown). Thirdly, we test that the results are robust to using other population thresholds to define a locality as a city. We obtain similar results if we use 5,000 instead of 2,000: the coefficient of the 0-10 km rail dummy is the same as before, at 0.39*** (not shown). For the year 1969, we can use 500 as we also have data on localities between 500 and 2,000 inhabitants: the coefficient is 0.44*** (not shown). When using the urbanisation rate (%) in 1962, we also find a positive effect for the 0-10 km dummy, at 3.92* (not shown). The effect is less significant, since standardised rural population also increased along the lines: the coefficients of the 0-10 and 10-20 km rail dummies are 0.38*** and 0.31*** respectively (not shown). Fourthly, for the European and urban population, we can also follow a difference-in-difference strategy where we compare connected and non-connected locations over time. The European population of the 403 non-arid locations was nil, and three locations had a city in 1901. We include location and year fixed effects, and we drop the (time-invariant) controls. The results are even stronger now. For example, when the dependent variable is standardised urban population, the coefficient of the 0-10 km rail dummy is 0.75** (not shown) vs. 0.39*** in the cross-section (column (2) of Table 1). Fifthly, Conley standard errors do not account for spatial autocorrelation among the group of railroad locations. We cluster standard errors using nine groups: one group for the locations of the main line, seven groups for the locations of each branch line, and one group for untreated locations. The results hold when doing so (not shown). Lastly, we cannot estimate the effects for 67 arid locations only.

4.5 Additional Results on Economic Change at Independence

We now discuss several additional results on economic change.

Historical factors. We showed that railroads attracted settlers and created cities. We now examine other outcomes at independence. This is important, because it clarifies what other factors the railroad brought. We use the same model as model (2) to examine whether railroad locations have better infrastructure in 1962. We have data on the number of hospitals, clinics, dispensaries, secondary schools, provincial police headquarters, divisional police headquarters, police stations, police posts, post offices and postal agencies for each location. We also know the Euclidean distances of the location to a paved road or an improved road, and whether the location contained an industrial centre. No location had any of these types of infrastructure in 1900. The cross-sectional regressions in 1962 can be interpreted as long-differenced estimations between 1901 and 1962. As shown in Panel A of Table 6, there are strong positive effects on the probabilities of having a hospital or a clinic (column (1)), a secondary school (column (2)), a police station (column (3)), a post office (column (4)), a paved road (column (5)) or an industrial centre (column (6)). These effects decrease as we move away from the line. For the sake of space, we only show the results when using indicator variables, although similar results are obtained when using the number of each type of infrastructure instead. The effects are slightly reduced when controlling for the log of total population in 1962 (Panel B), and further reduced when also including the log of European population in 1962 (Panel C). We use logs because we expect a non-linear relationship between public good availability and population. The fact that the effects remain high in Panel B but not in Panel C shows that these places were better endowed in infrastructure *per African capita*, but not as much *per European capita*. This suggests that railroads increased European population density, and public goods (for Europeans) were created as a result. Another interpretation could be that the coloniser has invested in other public goods simultaneously as it was building the railroad. In that case, the railroad effects on population could be also explained by these investments. However, there were almost no other investments in 1901. There were only 500 settlers then, who all lived in Mombasa or Nairobi. A second interpretation could be that railroads influenced the placement of other public goods between 1901 and 1962. Population also increased as a result of the presence of these factors. Even if we believe that the coloniser would not have invested in public goods where there were no Europeans, we cannot rule out this possibility.[11]

General equilibrium effects. While settlers migrated from Great Britain, African population growth was due to in-migration from the African areas. If railroads reallocate labour across space, does overall welfare increase? People migrate because they expect a higher income at the destination location. In Kenya, the rail gave access to a new factor of production – land – that made people more productive, as it was used to grow export crops. The railroad also permitted the growth of cities, in a country that was non-urbanised before. This economic transformation allowed Kenya to become a relatively wealthy African country at independence. As explained in section 3.2, roads would not have permitted this transformation before the 1940s. The technology to bitumenise roads in tropical Africa was not invented before the 1920s, and no road was paved before 1945 in Kenya (Burgess et al., 2013; Jedwab & Moradi, 2013). Trade costs would have remained high in the hinterland. However, while the aggregate effects were clearly positive, they were potentially negative for various African tribes, because of the multiple discriminatory interventions in the African labour market (Deininger & Binswanger, 1995). A hut tax was imposed upon the Africans, Africans worked as indentured farmers for the Europeans, and African wages were fixed at low levels in the cities. Europeans were thus able to capture most of the rents of the colonial economy. Some of the surplus must have nonetheless accrued to the natives, since they experienced significant progress in nutrition and health, as shown by anthropometric measures (Moradi, 2009). Since our goal is not to study the effects of colonisation on development, but to show how some colonial investments have resulted in a specific urban equilibrium, we leave these issues aside for future research.

5. RAILROADS, PATH DEPENDENCE AND DEVELOPMENT

We now document the relative decline of railroads and study their effects on long-run development. The persistence of the equilibrium provides strong evidence for path dependence. We also use our novel data set to study its channels.

[11]The effects on paved roads remain high even when controlling for European settlement. Indeed, roads served as feeder roads to the railroad during the colonial period. However, the roads do not explain why the railroad locations are more developed, as discussed in subsection 4.4.

5.1 Relative Decline of Railroads and Exodus of Europeans

The rail network reached its maximum length in 1930, and has stagnated since. As shown in Figure 3, freight volumes were increasing until independence in 1963. Railroad traffic has then declined since (more exactly after 1956 or 1973 depending on the measure of rail traffic used). Traffic volumes are now much lower than what they used to be, although rail traffic has not entirely disappeared. Railroads now account for 5-10% of total traffic (in volume and value) vs. almost 100% in the early colonial period. What caused the relative decline of colonial railroads?

Gwilliam (2011) and Jedwab & Moradi (2013) describe how underinvestment and poor management in the rail sector during a period of considerable road investment produced a significant decline of the former in Africa. The World Bank (2005) focuses on the specific case of Kenya. First, political and economic instability had a damaging effect on past public investments.[12] By the 1970s, track, motive power, and rolling stock were in desperate physical condition, due to the lack of resources devoted to reinvestments and maintenance. There were also management issues, as the Kenya Railways Corporation (KRC) was overstaffed, which absorbed a significant share of its revenues. Service quality was poor, which reduced traffic and freight revenues, thus further delaying the maintenance and accelerating the decline of the network. Second, the first governments of Kenya massively invested in roads (The World Bank, 2005; Burgess et al., 2013). Roads were three times cheaper to build. Yet maintenance costs were much lower for railroads. Kenya's total length of paved and improved roads increased threefold between 1964 and 2002. Rent-seeking favoured large construction projects prone to embezzlement, such as the construction of new roads. Road building was then driven by political considerations instead of economic considerations, as the various presidents of the country disproportionately targeted their ethnic homelands. Maintaining the colonial rail network was of no use in this regard. Roads soon outcompeted the rail, and they now account for most of the goods and passengers traffic.

The railroad locations did not completely lose their access to transportation, as the railroad lines did not entirely collapse and were replaced by roads at nearby sites. Yet they clearly lost their early advantage in terms of transportation. Almost all of the 80 locations less than 10 km from a colonial railroad already had a paved or improved road in 1964. There were then 122 non-railroad locations with a paved or improved road in 1964, while this number has increased to 256 in 2002. This implies that only 69 locations have no access to a paved or improved road. However, even these 69 locations have access to an earthen road. Kenya not being as tropical as most African countries, most earthen roads are suitable for motor traffic throughout the year. Therefore, while few locations of the hinterland were connected (by rail) to the coast in the early colonial period, almost all locations are now connected (by road). This has considerably eroded the initial advantage of the railroad locations. Therefore, road decentralization could have permitted the emergence of an entirely different spatial distribution post-independence.

The number of Europeans grew exponentially from 500 in 1901 to almost 60,000 in 1962. Europeans were then asked by the new president Kenyatta to choose the Kenyan citizenship. Most of them refused to give up their British citizenship,

[12] This instability includes the difficult transition to independence in 1963, the exodus of the European and Asian populations, the abolition of democracy in 1969, the death of Kenyatta and his replacement by his vice-president Arap Moi in 1979, and the economic crisis in the 1980s and 1990s.

which was not well-received by the new African political elite (Hornsby, 2013). Interracial tensions grew, and the European settlers searched for a way out. The "Million-Acre Scheme" was initiated and funded by the British government in 1962-1967 to transfer 1.2 million acres of European-owned land to African small-scale farmers. By 1969, most of the European farmers had left Kenya (see Figure 3). The European settlers who accepted to acquire Kenyan citizenship then settled in Nairobi and Mombasa. The African farmers that replaced them in the hinterland remained specialised in coffee. Yet coffee production has dramatically declined since the mid-1980s, as a result of low international prices. Coffee only account for 10% of total exports now, while tea now accounts for 25% of total exports. However, tea cultivation is mainly concentrated in the Western areas of Kenya, which also explains why cities have been growing in the African areas around Lake Victoria (see Figure 1). The railroad locations have thus lost their initial advantage in terms of European settlement and, to a lesser extent, commercial agriculture. As the settlers represented the business elite during the colonial period, these areas must have experienced a loss in their stock of human and entrepreneurial capital.

5.2 Economic Change Post-1962

The railroad locations have lost their initial advantage in terms of transportation, European settlement and commercial agriculture. In these conditions, should we expect these locations to be relatively more developed today? We test this hypothesis by studying the relative growth of connected locations after 1962. We focus on urban growth in order to test the persistence of the urban equilibrium.

Cross-sectional analysis. We run the same regression for 403 non-arid locations l as model (2), except the dependent variable is now defined for the year 2009:

$$zUPop_{l,2009} = a + Rail_{l,1962}\gamma + \delta_p + X_l\pi + \upsilon_{l,2009} \quad (3)$$

with $zUPop_{l,2009}$ being the standard score of urban population in 2009. We include the same province fixed effects and controls as before, and we also drop the locations including the nodes. Results are reported in Table 7. Column (1) shows that railroad locations are more developed today. The point estimates are not very different from what we found for the year 1962 (column (2) of Table 2). This shows the stability of the urban equilibrium between 1962 and 2009. In column (2), we show that urban persistence is not explained by the fact that railroads were replaced by roads at nearby sites. Controlling for whether the locations is 10 km from a paved road, or an improved road, does not strongly modify the relationship between colonial railroads and urban density today. Being close to a paved road has a positive effect on urbanisation. We verify that the results hold when using other measures of road infrastructure in 2002 (not shown, but available upon request): (i) we also add dummies for being 10-20, 20-30 and 30-40 km from a paved, or an improved road, (ii) we replace the road dummies by the Euclidean distances of the location's centroid to a paved road, or an improved road, and their squares, (iii) we use dummies equal to one if the location is crossed by a paved road, or an improved road, and (iv) we control for the total lengths of paved and improved roads in the location. In column (3), we verify that these long-term effects are causal by comparing the railroad locations with the placebo locations. The results are unchanged.[13] The

[13]By definition, if the effects are causal in the short run (the colonial period), they should also be causal in the long run. We nonetheless test that these effects are robust to adding ethnic group or district fixed effects, or a fourth-order polynomial of the longitude and latitude of the location's

railroad effects then disappear when we control for the standard score of European settlement in 1962 (column (4)). But the effect of the number of Europeans is then reduced when we add the standard score of the urban population in 1962 (column (5)), thus controlling for the fact that the settlers established these cities during colonisation. The long-run effects are strongly explained by the urban population in 1962. A one standard deviation in urban population in 1962 is associated with a 0.70 standard deviation in urban population in 2009. Hence, railroad locations are more urbanised today because they were more urbanized in the past.[14]

Dynamic cross-sectional analysis. The previous cross-sectional analysis does not allow us to study the dynamics of path dependence between the two years 1962 and 2009. We now run the following repeated cross-sectional regressions for 403 non-arid locations l and six years $t = [1962, 1969, 1979, 1989, 1999, 2009]$:

$$zUPop_{l,t} = c + Rail_{l,1962}\kappa_t + zUPop_{l,t-1}\rho_t + \lambda_p + X_{l,t}\theta_t + \omega_{l,t} \quad (4)$$

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 ($zUPop_{l,t-1}$), κ_t captures the additional effect of the railroad for each period. We have data on the urban population from 1901 to 2009. We also run the same regressions using total population as the dependent variable, but we do not have data for the years 1901, 1969 and 2009. For the year 1901, we nevertheless control for historical settlement patterns using various controls listed in Table 1. For the year 1979, the previous period is 1962. There is spatial convergence (divergence) in the period $[t\text{-}1;t]$ if κ_t is lower (higher) than 0. If κ_t is equal to 0, the effect is unchanged during the period. Figure 4 displays the effects κ_t. The effects were large in 1962 (vs. 1901), as already shown in column (2) of Table 2. The effects were then close to 0 the following decades, with the exception of the year 1979 (vs. 1969) for which there is a positive and significant effect (0.08**). The population effects are small and not significant, with the exception of the year 1979. These effects are not smaller for the mere reason that we are studying periods of 10-20 years instead of 60 years, as already shown when studying the cross-section in 2009 (see column (5) of Table 7). The positive effects in 1979 could then be explained by the fact that migration restrictions were lifted in 1963. The various ethnic groups of the Rift Valley kept colonising the land around the railroad lines. After 1979, roads had been built or upgraded throughout the country, rail traffic was clearly declining, the settlers had left, and land expansion was decelerating as coffee production was falling.

5.3 Channels of Path Dependence

What explains path dependence? Colonial *sunk investments* could induce people to stay. 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 loss of the initial advantage dates back from 40 to 50 years, whether we consider 1973 or 1956 as the cut off year to identify path dependence. Railroad locations could still be over-supplied with such factors, which would attract more people. Second, if there are returns-to-scale in production, factors need to be co-located in the same locations. There is a *coordi-*

centroid, or comparing the branch and main lines (not shown but available upon request).

[14]We cannot use the urbanization rate as the dependent variable, as we do not have total population data in 2009. Then, we could use urbanization rates in 1999. However, many cities have been expanding beyond the boundaries of their location post-1962, which gives urban rates above 100%.

nation problem as it is not obvious which locations should have the factors. Then, it makes sense to locate factors in locations that are already developed. The location of factors (including people) today 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 channels of path dependence.

Historical factors. In section 4.5, we showed that connected locations had better health, educational, institutional, communication, transportation and industrial infrastructure at independence. If these historical factors had an independent effect on urban population today, adding them to the regression of model (3) should capture some of the effect of urban population in 1962 on urban population in 2009. In Table 6, we only showed the effects when using indicator variables for each type of infrastructure. For the path dependence analysis, we include both the indicator variables and the numbers of each subtype of infrastructure in the location, as we want to control as much as possible for sunk investments.[15] We actually find that their inclusion reduces the coefficient of urban population in 1962 by 25% (from 0.70*** to 0.53***, columns (5) and (6) of Table 7). There are two caveats to our analysis. First, we may not properly measure the historical factors at independence, which would make us underestimate the contribution of sunk investments. The coefficient of urban population in 1962 is almost unchanged at 0.50*** (not shown) if we also include the square, cube and fourth power of the numbers of each subtype of infrastructure. If the sunk investments hypothesis is correct, the urban coefficient should then decrease relatively more when we only include the most "expensive" assets rather than the "cheapest" ones. When we only include the most expensive asset of each type of infrastructure (e.g., paved roads for roads and hospitals for health infrastructure), the coefficient decreases to 0.55***. When we include the cheapest ones only, the coefficient only decreases to 0.64***. Our analysis is thus strongly biased if we omit other expensive assets that existed in 1962. Since there were no universities, airports or dams in the 403 non-nodal locations in 1962, we believe we properly capture the historical factors. Second, African in-migration to the European areas was controlled during colonisation. As a result, the railroad locations were better endowed in infrastructure *per African capita*. We thus verify that the results remain the same when also controlling for the African population in 1962. The coefficient of urban population in 1962 remains 0.53*** (not shown). To conclude, 25% of urban persistence can be explained by sunk investments. These results are in line with 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. Whereas historical factors matter to explain urban patterns today, they are not the main channel of path dependence. While cities are expensive to build, which may explain their persistence when they economically decline (Glaeser & Gyourko, 2005), history has shown that cities well endowed with infrastructure can nonetheless collapse (e.g., Detroit) while cities with initially little infrastructure can grow very fast (e.g., Houston). Building costs have relatively decreased over time (Bleakley & Lin, 2012). Likewise, in Kenya, many cities have grown at a fast pace in the African areas. Therefore, building costs, even if non-negligible, only partially explain why there is path dependence. It is then interesting to note that

[15] These variables are dummies equal to one if the location has a hospital or a clinic, a secondary school, any type of police station (police headquarters, stations and posts), a post office or postal agency, a paved road, an improved road, or an industrial centre, and the respective numbers of hospitals, clinics, dispensaries, secondary schools, provincial police headquarters, divisional police headquarters, police stations, police posts, post offices and postal agencies for each location.

the coefficients of the rail dummies are negative in column (6). Once we control for the positive long-term effects of population and factors in 1962, the railroad locations are "losing" relative to the rest of the country. This is an illustration of path dependence. Without increasing returns (in people and other factors), the railroad locations would have declined after the erosion of the early advantage.

Contemporary factors. We first verify that railroad locations have higher densities of factors today (around 2000). We run the same regression as model (3), except we use as a dependent variable the density of various contemporary factors.[16] Results are displayed in Table 8. In Panel A of columns (1)-(12), we show that the inhabitants of railroad locations have better access to health, educational and transport infrastructure. Railroad locations are characterised by a higher number of hospitals (column (2)), health clinics (column (4)), health dispensaries (column (6)), primary schools (column (8)) and secondary schools (column (10)). Railroad locations have also higher probabilities of being crossed by a paved road (column (11)). The negative effects for the "improved road dummy" is due to the fact that many roads along the lines were paved over time. The point estimates are strongly reduced when controlling for population today (Panel B). Locations 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 (7) of Table 7, we find that controlling for contemporary factors does not modify the relationship between the urban populations in 1962 and 2009 (the coefficient is almost unchanged). These results point to the following story: railroad locations have higher densities today, because people co-locate where there are more people in the past, and other factors "follow" people (instead of people following factors). There were then more people in the past because of the population effect during the colonial period and repeated co-location decisions afterwards. Since industrial agglomeration effects and human capital externalities must be limited in Kenya, capital and highly skilled workers may not be the co-located factors. Rather, the cities may serve to co-locate the relatively skilled workers of the service sector. Unfortunately, we do not have employment data to examine this hypothesis. Thus, although the cities of developed and developing countries may serve to co-locate distinct factors, the main channel of path dependence is the same. The relative fall of Detroit and rise of Houston could then be interpreted in terms of a switch in expectations with regard to the future growth of these cities. This shows the stability of the equilibrium if, for example, regional policies cannot affect expectations as much as infrastructure.

External validity. To what extent are our results not driven by the specificities of the Kenyan context? Firstly, the African governments of the post-independence period have been dominated by two ethnic groups: the Kikuyus and the Kalenjins (see Figure 1). As shown by Burgess et al. (2013), the areas that shared these ethnicities have received disproportionately more public investments. We should thus expect more urban persistence in the Kikuyu-Kalenjin (KK) areas. Conversely, these governments may have had a specific interest in protecting the colonial cities of the non-KK areas so as to better control their population. In that case, we should expect more urban persistence in the non-KK areas. The ethnic group fixed effects partially control for some of these effects. We then run the same regression as in column (7)

[16] We do not have data on factors in 1969, 1979 and 1989. However, disentangling the mechanisms of path dependence only requires data at independence and for the most recent period.

of Table 7, except we now also interact urban population in 1962 with a dummy equal to one if the location belongs to the KK areas. The independent effect of urban population in 1962 remains almost the same, at 0.47*** (not shown) vs. 0.54*** before. The interacted effect is 0.20, but is not significant. Therefore, even if ethnic favoritism may have protected cities in the KK areas, this effect was too small to explain path dependence. Secondly, we believe that the results that we have obtained for Kenya can be generalized to other countries in the developing world. Gwilliam (2011) describes how railroad traffic has collapsed in most African countries. Jedwab & Moradi (2013) show that the locations along the colonial railroad lines are still relatively more developed in Africa today. Then, many African countries have received significant numbers of European settlers during colonization, such as Algeria, Angola, Ivory Coast, Namibia, South Africa and Zimbabwe.

To conclude, if historical factors matter to explain urban patterns today, railroad cities mainly persisted because they solved the coordination problem of contemporary factors before independence, and for each subsequent period thereafter.

5.4 Economic Development Today

Railroad locations have higher densities of 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 per capita, if they are better endowed in non-observable factors that raise total factor productivity. We run the same regression as model (2) except we now use various contemporary measures of development as the dependent variable. If we simultaneously control for contemporary urban population, we restrict our comparison to the railroad and non-railroad cities that have a similar size. While European cities kept growing after independence, many cities grew at a fast pace in the African areas (see Figure 1). The European and African cities have two different origins. The fact that the former were founded earlier probably gave them an initial advantage, which could have made them wealthier in the long run.

We first use two poverty measures in 1999. In columns (1) and (2) of Table 9, the dependent variable is the poverty headcount ratio, the percentage of the population of location l living below the national poverty line (%). In columns (3) and (4), the dependent variable is the poverty gap, the mean shortfall from the poverty line of location l, expressed as a percentage of the poverty line (%). No matter the measure used, the locations closer to the railroad are less poor (column (1) and (3)). Columns (2) and (4) then show that we obtain similar results if we focus our analysis on cities of similar sizes, by controlling for the urban population the same year (1999). Railroad cities are wealthier than non-railroad cities *ceteris paribus*. These results hold if we use satellite data on night lights as an alternative measure of development, as in Henderson, Storeygard & Weil (2012). Our dependent variable in column (5) and (6) is average light intensity for each location in 2000-01. In column (7) and (8), the dependent variable is the log of average night light intensity per sq km. The issue with using this variable is that there are 134 locations for which no light bright enough is recorded, so that average light intensity is equal to 0. Using logs forces us to drop these observations, hence the need to also rely on a more simple measure of night light intensity as in columns (5) and (6). No matter the measure used, and the sample considered, we find that the railroad locations and cities are "brighter", and thus wealthier. Henderson, Storeygard & Weil (2012) show that the elasticity of *ln(lights/area)* to *ln(GDP)* is 0.3 for a sample of countries. Assuming there is the same relationship at the subnational level in Kenya, and using the same dependent variable, we find that the railroad locations could be 30-70% wealthier than the non-railroad locations (column (7)). The railroad cities could then be 20-35% wealthier than the non-railroad cities of similar sizes (column (8)). These results suggest that railroad locations are wealthier today. Therefore, there must be unobservable factors that were repeatedly co-located along the railroads, such as entrepreneurial, social or institutional capital. Colonial railroads may have also solved a coordination failure for these specific factors, and the initial advantage obtained as a result must have had positive effects one century later.

5.5 Discussion: An Optimal Urban Equilibrium?

We now discuss the optimality of the equilibrium. The following discussion is unavoidably speculative, as we cannot know what would have happened if the colonial railroad had not been built. We nonetheless think that it is important to hypothesise on whether the colonial urban equilibrium is optimal today.

First, had the coloniser not built railroads, we believe roads would have permitted the same economic transformation, but only a few decades later. The technology to bitumenise roads in tropical Africa was not invented before the 1920s, and no road was paved in Kenya before 1945 as paving costs were still high. Trade costs would have probably remained high in the hinterland, fewer cities would have emerged in the colonial period, and the country would have been less urbanised at independence and today (as cities take time to grow). Kenya was the most urbanised Eastern African country in 1960 and 2010 according to Geopolis (2013).

Second, what would have happened if the railroad had been built by an African government in a unified Kenya instead? Possibly fewer Europeans would have settled. This may have not prevented the commercialisation of agriculture though. For example, the colonial railroad had positive spatial effects on the cultivation of cocoa by Africans in Ghana (Jedwab & Moradi, 2013). However, it is likely that the African state would have chosen other locations for the placement of the railroad. The objective of the coloniser was to link Uganda (and thus Lake Victoria) to the coast. Kenya was simply a transit territory, so the line was built so as to minimise the Euclidean distance between the two, while also taking into account construction costs (e.g., the need to stop in Nairobi, a "place of cool waters"). An African railroad would have probably connected the most populated areas of the hinterland to the coast. This mimics what the first European explorers did, as they traversed densely populated African areas to find food and water *en route* to the west (see Figure 2). The comparison of the railroad and placebo locations has confirmed that the latter were historically more populated (see column (3) of Table 1). In terms of geographical fundamentals, the placebo locations were lower in altitude (by 111 m only though) and more suitable for coffee and tea cultivation. While these could be important assets for rural growth, they are unlikely to affect urban growth (once a city is formed). In the long run, given increasing returns and technological progress, physical geography should have no effect on urban economic activity. Thus, the railroad cities are not necessarily suboptimal today.

Third, if no railroad had been built, and roads eventually reduced trade costs for all locations, cities could have emerged in the areas that were historically populated by Africans prior to the colonial era. These populations must have been living in these areas because of the better fundamentals they provided. Without colonisation, development could have been limited to these areas, and population growth could have resulted in urban growth. In the 19th century, we define the "historically settled locations" as locations in which the area share of the "major settled groups" is above 75%. We regress the various controls on the 0-20 km rail dummy to test if the railroad and historically settled locations differ. We do not include province fixed effects, as the exercise is simply a comparative one. Column (4) of Table 1 displays the results. While the railroad locations were historically less populated (-44***), they are more urban (0.36***) and richer (-14.2***) today. In terms of fundamentals, they are less coastal (-0.08**), higher in altitude (405***) and more suitable for coffee (8.7**). This is not surprising as the rail was built to connect the (hilly) hinterland. Without the railroad, there would have been more cities restricted to the coast, a positive urban asset. Altitude could be a negative urban asset, if it increases trade costs (but there are roads now), or a positive urban asset, if they are less disease-ridden. While we cannot be sure whether the geography of these cities still matters today, it is clear that the colonial railroad has permitted a dramatic increase in the number and size of cities in the hinterland.

6. CONCLUSION: PATH DEPENDENCE AND DEVELOPMENT

In this paper, we have used a natural experiment and new data to study the emergence and persistence of a spatial equilibrium following a historical shock. The construction of the colonial railroad had a strong causal impact on European settlement patterns in Kenya. Economic development in the European areas in turn determined the location of main cities of the country at independence. These locations remain relatively more developed today, although they lost their initial advantage in terms of transportation and human capital (the settlers) in the post-independence period. The railroad cities mainly persisted because their early emergence served as a mechanism to coordinate investments in the colonial and subsequent periods.

Our results make the following contributions. First, we provide strong evidence for path dependence and the existence of multiple spatial equilibria. Second, we find that the "nature" of path dependence does not differ between developed and developing countries. Sunk investments only partially contribute to spatial persistence. Factors are not costlessly mobile, but not to the point of preventing the emergence of another equilibrium. Cities mainly persist because they solve the coordination failure of factors for the first period, when these cities emerged, and for each subsequent period thereafter. The factors that are co-located may differ between developed and developing countries, yet the mechanism is the same. Third, increasing returns suggest that regional policies can influence regional development. However, if spatial equilibria persist because they solve the coordination failure of factors for each period, regional policies may be less effective. Indeed, expectations could be more difficult to adjust than infrastructure stocks. Lastly, we study the causal impact of an investment in transport infrastructure. In poor and rural countries, transport infrastructure and the cities they contribute to create can serve as a mechanism to coordinate contemporary investments. More factors can be subsequently accumulated, which promotes long-term development. The spatial and economic impact was large in our context, because the previously deployed technology (headloading) was highly inefficient. The impact could have been lower in a richer country, with lower initial trade costs. This shows the importance of studying the extent and forces of path dependence in the specific context of developing countries.

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.

Albouy, David Y. 2012. "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment." *American Economic Review*, 102(6): 3059–76.

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.

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

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.

Beeson, Patricia E., David N. DeJong, and Werner Troesken. 2001. "Population growth in U.S. counties, 1840-1990." *Regional Science and Urban Economics*, 31(6): 669–699.

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.

Bonfatti, Roberto, and Steven Poelhekke. 2013. "From Mine to Coast; Transport infrastructure and the direction of trade in developing countries."

Bosker, Maarten, and Eltjo Buringh. 2010. "City seeds. Geography and the origins of the European city system." C.E.P.R. Discussion Papers CEPR Discussion Papers 8066.

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.

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

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.

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics*, 92(1): 1–45.

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." Unpublished manuscript, Department of Economics, Harvard University.

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.

Diamond, Jared. 2011. *Collapse: How Societies Choose to Fail or Succeed: Revised Edition.* London: Penguin Books.

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. 2013. "Growing through cities in developing countries." Unpublished manuscript, Department of Economics, University of Pennsylvania.

Duranton, Gilles, and Jean-Francois Maystadt. 2013. "The Development Push of Refugees: Evidence from Tanzania." Unpublished manuscript, Wharton School, University of Pennsylvania.

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.

Fenske, James. 2010. "Ecology, trade and states in pre-colonial Africa." University Library of Munich, Germany MPRA Paper 27203.

Fenske, James. 2013. "Does Land Abundance Explain African Institutions?" *The Economic Journal*, 123(573): 1363–1390.

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.

Geopolis. 2013. "Africapolis: L'urbanisation en Afrique Centrale et Orientale." Rapport general de l'etude pour l'Agence Francaise de Developpement.

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., and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy*, 113(2): 345–375.

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.

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

Hazlewood, A. 1964. *Rail and Road in East Africa: Transport Co-ordination in Under-Developed Countries.* Oxford: Basil Blackwell.

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. 2010. "Cities and Development." *Journal of Regional Science*, 50(1): 515–540.

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.

Holmes, Thomas J., and Sanghoon Lee. 2012. "Economies of Density versus Natural Advantage: Crop Choice on the Back Forty." *Review of Economics and Statistics*, 94(1): 1–19.

Hornsby, Charles. 2013. *Kenya: A History Since Independence.* London: I.B. Tauris.

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.

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

Krugman, Paul. 1991*a*. "History and Industry Location: The Case of the Manufacturing Belt." *American Economic Review*, 81(2): 80–83.

Krugman, Paul. 1991*b*. "History versus Expectations." *The Quarterly Journal of Economics*, 106(2): 651–67.

Krugman, Paul. 1991*c*. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 99(3): 483–99.

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

Maloney, William F., and Felipe Valencia Caicedo. 2012. "The persistence of (subnational) fortune : geography, agglomeration, and institutions in the new world." The World Bank Policy Research Working Paper Series 6187.

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.

Michaels, Guy, and Ferdinand Rauch. 2013. "Resetting the Urban Network: 117-2012." Centre for Economic Performance, LSE CEP Discussion Papers dp1248.

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.

Moradi, Alexander. 2009. "Towards an Objective Account of Nutrition and Health in Colonial Kenya: A Study of Stature in African Army Recruits and Civilians, 1880-1980." *The Journal of Economic History*, 69(03): 719–754.

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

Motamed, Mesbah, Raymond Florax, and William Matsers. 2013. "Agriculture, Transportation and the Timing of Urbanization: Global Analysis at the Grid Cell Level." Unpublished manuscript, Tufts University, Department of Economics.

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.

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.

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.

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

Wantchekon, Leonard, Natalija Novta, and Marko Klasnja. 2013. "Education and Human Capital Externalities: Evidence from Colonial Benin." Unpublished manuscript, Department of Politics, Princeton University.

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

Variable of Interest:			**Rail Dummy, 0-20 km**			
Group of Control Cells:			**(1) All**	**(2) 20-40 km**	**(3) Placebo**	**(4) Historical**
Dependent Variable:	Mean	SD	Coeff.	Coeff.	Coeff.	Coeff.
"Coastal location" dummy	0.06	0.25	-0.03	0.01	0.00	-0.08**
Distance to the coast (km)	443	221	-8	-8	-8	28
Average annual rainfall (mm)	1,184	332	-59*	-136***	-28	-44
Altitude: mean (m)	1,421	678	19	-30	111***	405***
Altitude: standard deviation (m)	126	110	-36***	-30**	6	0
Share of arid soils (%)	0.0	0.7	-0.1**	-0.0	-0.1	-0.0
Share soils suitable for agriculture (%)	79.1	36.9	-0.4	-3.3	-6.0	-0.8
Share soils suitable for coffee (%)	18.9	31.0	-0.7	0.1	-5.5*	8.7**
Share soils suitable for tea (%)	12.6	26.7	-11.9***	-15.4***	-10.1**	-5.7
Area (sq km)	358.5	495.4	-84.6	-58.3	95.8**	59.3**
"City in 1901" dummy	0.01	0.09	-0.01*	0.00	-0.02	-0.03*
"Provincial capital" dummy	0.01	0.10	0.00	-0.00	-0.01	-0.01
Area share of "major settled groups" (%)	56.4	43.4	-10.7**	-15.8***	-21.4***	-44.0***
Area share of "pastoralists" (%))	14.4	32.3	-1.0	4.2	4.9	14.9***
"Isolated groups" dummy	0.20	0.40	0.02	0.06	0.13**	0.21***
Urban Pop. (Z-Score, 2009)			0.60***	0.58***	0.45***	0.36***
Poverty Headcount Ratio (%, 1999)			-3.9***	-2.9**	-6.0***	-14.2***
Number of treated locations:			162	162	162	162
Number of control locations:			241	89	81	126

Notes: OLS regressions using data on 15 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 railroad line, while simultaneously including province fixed effects (N = 8). Robust standard errors: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. There are 15 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 between 20 and 40 km from a railroad line (N = 89). In column (3), the control locations are the locations less than 20 km from a placebo line (N = 81). In column (4), the control locations are the "historically settled locations" defined as the locations for which the area share of major settled groups is above 75% (N = 126). Province fixed effects are not included then. See Online Data Appendix for data sources.

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

Dependent Variable:	Columns (1)-(9): Number of Inhabitants in 1962 (Z-Score)								
	European Pop.	**Urban Pop.**		**Total Pop.**		**African Pop.**		**Asian Pop.**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rail Dummy, 0-10 km	0.81***	0.39***	0.09	0.43***	0.24*	0.41***	0.23*	0.54***	0.16
	(0.22)	(0.15)	(0.11)	(0.13)	(0.13)	(0.13)	(0.13)	(0.17)	(0.12)
Rail Dummy, 10-20 km	0.31***	0.11	-0.00	0.32***	0.24**	0.32***	0.25**	0.11	-0.04
	(0.09)	(0.10)	(0.10)	(0.12)	(0.12)	(0.12)	(0.12)	(0.08)	(0.09)
Rail Dummy, 20-30 km	0.28*	0.06	-0.04	0.16	0.09	0.15	0.09	0.13	-0.00
	(0.14)	(0.11)	(0.10)	(0.17)	(0.15)	(0.17)	(0.15)	(0.13)	(0.09)
Rail Dummy, 30-40 km	0.11	-0.14	-0.18*	0.03	0.01	0.03	0.01	-0.12	-0.18
	(0.10)	(0.11)	(0.09)	(0.16)	(0.15)	(0.16)	(0.16)	(0.13)	(0.11)
Number of Europeans			0.37**		0.24**		0.22**		0.47**
(Z-Score, 1962)			(0.18)		(0.11)		(0.10)		(0.19)
Province Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Location Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of Observations	403	403	403	403	403	403	403	403	403
R-squared	0.19	0.34	0.45	0.37	0.42	0.38	0.42	0.27	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 in excess of 2,000 inhabitants in 1962. In columns (4)-(5), the dependent variable is the standard score of the total population of location l in 1962. In columns (6)-(7), the dependent variable is the standard score of the African population of location l in 1962. In columns (8)-(9), the dependent variable is the standard score of the Asian population in location l in 1962. As the European, urban and Asian populations were almost nil in 1901, the cross-sectional regressions of columns (1), (2), (3), (8) and (9) should be interpreted as long-differenced estimations for the period 1901-1962. All regressions include province fixed effects (N = 8), and control 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), mean and standard deviation of altitude (m), area (sq km), shares of soils (%) suitable for agriculture, coffee or tea, "city in 1901" dummy, "provincial capital" dummy, area of "major settled groups" (%), area share of "pastoralists" (%) and "isolated groups" dummy. Mombasa, Nairobi and Kisumu are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 3: COLONIAL RAILROADS, EUROPEAN SETTLEMENT AND AGRICULTURE, 1901-1962

	Col. (1)-(5): Number of Inhabitants (Z-Score)					**Col. (6)-(11): Thousand Acres (Z-Score)**					
Dependent Variable:	**European Pop. 1962**		**European Voters 1933**			**European Coffee Cultivation 1962**		**European Maize Cultivation 1962**		**European Wheat Cultivation 1962**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Rail Dummy, 0-10 km	0.81***	0.58***	0.60***	-0.02	0.19	0.27*	0.03	0.49***	0.34*	0.38*	0.24
	(0.22)	(0.22)	(0.20)	(0.04)	(0.14)	(0.14)	(0.07)	(0.18)	(0.18)	(0.21)	(0.21)
Rail Dummy, 10-20 km	0.31***	0.24***	0.17**	-0.04	0.01	0.03	-0.04	0.51***	0.46***	0.27**	0.22**
	(0.09)	(0.08)	(0.07)	(0.03)	(0.08)	(0.11)	(0.12)	(0.14)	(0.14)	(0.11)	(0.11)
Rail Dummy, 20-30 km	0.28*	0.19*	0.23*	-0.03	0.15	0.08	-0.02	0.43	0.37	0.20	0.14
	(0.14)	(0.11)	(0.12)	(0.05)	(0.11)	(0.10)	(0.10)	(0.27)	(0.27)	(0.16)	(0.16)
Rail Dummy, 30-40 km	0.11	0.14*	-0.08	-0.03	0.01	0.06	0.08	0.03	0.05	-0.11	-0.10
	(0.10)	(0.08)	(0.09)	(0.03)	(0.05)	(0.06)	(0.06)	(0.10)	(0.09)	(0.09)	(0.08)
Number of European Voters		0.38***									
(Z-Score, 1933)		(0.13)									
Number of European Farmers				0.96***			0.36*		0.22**		0.22
(Z-Score, 1933)				(0.11)			(0.21)		(0.09)		(0.16)
Number of European Voters					0.70***						
(Z-Score, 1919)					(0.22)						
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of Observations	403	403	403	403	403	403	403	403	403	403	403
R-squared	0.19	0.34	0.45	0.37	0.42	0.38	0.42	0.19	0.22	0.27	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 columns (1)-(2), the dependent variable is the standard score of European population of location l in 1962. In column (3)-(5), the dependent variable is the standard score of the number of European voters of location l in 1933. Data is not available on the total number of Europeans at the location level for that year. In column (6)-(11), the dependent variables are the standard scores of the area of location l devoted to European coffee cultivation, maize cultivation and wheat cultivation (thousand acres) in 1962. As the European population and coffee, maize and wheat cultivation were almost nil in 1901, the cross-sectional regressions of columns (1)-(9) should be interpreted as long-differenced estimations for the period 1901-1962. 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 4: ALTERNATIVE IDENTIFICATION STRATEGIES, 1901-1962

Strategy:	*Baseline Results (see Table 2)*	*Including Ethnic Fixed Effects*	*Including District Fixed Effects*	*Fourth Order Polynomial Long Lat*	*Effects Main vs. Branch Lines*	*Placebo Treatment Effects*	*Drop if Rail 0-X km = 1*	*Control Group: Placebo Locations*	*Drop if Intersection*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A:	***Dependent Variable*: Number of European Inhabitants in 1962 (Z-Score)**								
Rail Dummy, 0-30 km	0.41***	0.54***	0.20*	0.38***		0.26**	-0.03	0.34***	0.20***
	(0.11)	(0.15)	(0.11)	(0.12)		(0.11)	(0.05)	(0.10)	(0.07)
Main Dummy, 0-30 km					0.54***				
					(0.17)				
Branch Dummy, 0-30 km					0.20				
					(0.15)				
Panel B:	***Dependent Variable*: Number of Urban Inhabitants in 1962 (Z-Score)**								
Rail Dummy, 0-10 km	0.35**	0.35**	0.25**	0.35**		0.17	0.03	0.35*	0.36
	(0.15)	(0.15)	(0.13)	(0.15)		(0.11)	(0.07)	(0.21)	(0.24)
Main Dummy, 0-10 km					0.47*				
					(0.25)				
Branch Dummy, 0-10 km					0.37*				
					(0.21)				
Panel C:	***Dependent Variable.*: Total Number of Inhabitants in 1962 (Z-Score)**								
Rail Dummy, 0-20 km	0.31***	0.33***	0.19**	0.23**		0.36***	0.22	0.08	0.02
	(0.08)	(0.10)	(0.10)	(0.09)		(0.09)	(0.14)	(0.14)	(0.14)
Main Dummy, 0-20 km					0.30**				
					(0.15)				
Branch Dummy, 0-20 km					0.39***				
					(0.11)				
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of Obs.	403	403	403	403	403	403	187;323;241	281;154;243	236;109;198

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. For the sake of simplicity its employs only the 0-30 km (European pop.), 0-10 km (urban pop.) and 0-20 km (total pop.) rail dummies. In columns (2) and (3), we include 21 ethnic group fixed effects and 35 district fixed effects respectively. In column (4), we include first, second, third and fourth order polynomials of the longitude and latitude of the location's centroid. In column (5), we verify that the effects are not lower for the main line. In columns (6) and (7), we test that there are no effects when using the placebo lines as an alternative. In column (7), we drop the locations less than X = 30 km (Panel A), 10 km (Panel B) and 20 km (Panel C) from a railroad line, in order to compare the placebo locations with the other control locations, while suppressing the effects from the railroad lines. In columns (8)-(9), the use the placebo locations are the control group: First we use the locations less than 30 km (Panel A), then 10 km (Panel B) and finally 20 km (Panel C) from a placebo line. In column (9), we drop the locations less than 10 km from both a railroad line and a placebo line, in order to compare the treated locations with the placebo locations for the segments where the railroad and placebo lines do not intersect. All regressions include province fixed effects (N = 8), and the same controls as in Table 2. Mombasa, Nairobi and Kisumu are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 5: ROBUSTNESS AND SPECIFICATION CHECKS, 1901-1962

Robustness/Specification Check:	*Dropping Coastal Province* (1)	*Using the Full Sample* (2)	*Include Distances to Nodes* (3)	*Controlling for White Highlands* (4)	*Dropping the Controls* (5)	*Controlling for Roads 1963* (6)	*Logs instead of Z-Scores* (7)	*Euclidean Distance to Rail* (8)	*Location Crossed by Rail* (9)	*Conley SEs (50 km)* (10)
Panel A:	***Dependent Variable*: Number of European Inhabitants in 1962 (Z-Score, but Logs in Column (7))**									
Rail Dummy, 0-30 km	0.35***	0.40***	0.24***	0.23**	0.41***	0.35***	1.36***			0.41***
	(0.09)	(0.11)	(0.10)	(0.11)	(0.08)	(0.11)	(0.22)			(0.11)
Distance to Rail (10 km)								-0.074***		
								(0.024)		
Sq. Distance to Rail (10 km)								0.001		
								(0.001)		
Crossed by Rail Dummy									0.542***	
									(0.16)	
Panel B:	***Dependent Variable*: Number of Urban Inhabitants in 1962 (Z-Score, but Logs in Column (7))**									
Rail Dummy, 0-10 km	0.41**	0.36**	0.35**	0.34**	0.52**	0.23*	0.66**			0.35**
	(0.17)	(0.15)	(0.15)	(0.16)	(0.15)	(0.12)	(0.32)			(0.13)
Distance to Rail (10 km)								-0.065**		
								(0.025)		
Sq. Distance to Rail (10 km)								0.003**		
								(0.001)		
Crossed by Rail Dummy									0.380**	
									(0.15)	
Panel C:	***Dependent Variable.*: Total Number of Inhabitants in 1962 (Z-Score, but Logs in Column (7))**									
Rail Dummy, 0-20 km	0.31***	0.25***	0.19**	0.26***	0.16*	0.24**	0.80***			0.31***
	(0.10)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)	(0.24)			(0.09)
Distance to Rail (10 km)								-0.077***		
								(0.023)		
Sq. Distance to Rail3 (10 km)								0.002**		
								(0.001)		
Crossed by Rail Dummy									0.464***	
									(0.13)	
Province FE, Controls	Y	Y	Y	Y	Y; N	Y	Y	Y	Y	Y

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. To focus the analysis, the variables of interest are the 0-30 km (European tot.), 0-10 km (urban tot.) and 0-20 km (total tot.) railroad dummies. In column (1), we drop the coastal province and focus on the non-arid locations west of Nairobi (N = 344). In column (2), we use the full sample (N = 470). In column (3), we add the Euclidean distances (km) to Mombasa, Nairobi and Kisumu and their squares. In column (4), we add a dummy variable equal to one if the location belongs to the White Highlands. In column (5), we drop the controls. In column (6), we control for locations within 10 km of a paved, improved or earthen road in 1963. In column (7), the log of (European/urban/total population + 1) is the dependent variable. In column (8), the variables of interest are the Euclidean distance (tens of km) of the location's centroid to the rail and its square. In column (9), the variable of interest is a dummy variable equal to one if the location is crossed by the rail. In column (10), 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, and the same controls at the location level as in Table 2. Mombasa, Nairobi and Kisumu are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 6: 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	Industrial Center Dummy
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: ***Railroads and historical factors (1962)***						
Rail Dummy, 0-10 km	0.29***	0.21***	0.15*	0.35***	0.57***	0.28***
Rail Dummy, 10-20 km	0.15**	0.09	0.13*	0.25***	0.30***	0.13**
Rail Dummy, 20-30 km	0.11*	0.14*	0.05	0.16**	0.14**	0.06
Rail Dummy, 30-40 km	0.10	0.03	-0.02	0.01	0.12*	0.00
Panel B: ***Railroads and historical factors (1962), conditioned on historical population (1962)***						
Rail Dummy, 0-10 km	0.27***	0.19**	0.11	0.33***	0.57***	0.26***
Rail Dummy, 10-20 km	0.13**	0.07	0.08	0.23***	0.29***	0.11**
Rail Dummy, 20-30 km	0.10*	0.14*	0.04	0.16**	0.14**	0.06
Rail Dummy, 30-40 km	0.10	0.03	-0.02	0.02	0.12*	0.00
Panel C: ***Railroads and historical factors (1962), also conditioned on European settlement (1962)***						
Rail Dummy, 0-10 km	0.18***	0.06	-0.08	0.15*	0.53***	0.17***
Rail Dummy, 10-20 km	0.05	-0.04	-0.06	0.09	0.26***	0.04
Rail Dummy, 20-30 km	0.05	0.07	-0.06	0.06	0.12**	0.01
Rail Dummy, 30-40 km	0.07	-0.02	-0.09	-0.05	0.11*	-0.03
Mean of the Variable	0.28	0.15	0.30	0.33	0.28	0.12
Province FE, Controls	Y	Y	Y	Y	Y	Y

Notes: OLS regressions using data on 403 non-arid locations for the year 1962. In the interest of space, robust SEs are not reported; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the rail dummies on six measures of historical factors in 1962. As the measures were close to 0 in 1901, the cross-sectional regressions could be interpreted as long-differenced estimations for the period 1901-1962. In Panel A, we regress each measure on the rail dummies. In Panel B, we control for log(total pop.) in 1962. In Panel C, we also include log(European pop. + 1) in 1962. All regressions include province fixed effects and the same controls as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 7: COLONIAL RAILROADS AND URBAN POPULATION GROWTH, 1901-2009

Dependent Variable:	Urban Population in 2009 (Number of Inhabitants, Z-Score)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rail Dummy, 0-10 km	0.52***	0.40***	0.49***	0.11	0.08	-0.26*	-0.26*
	(0.16)	(0.15)	(0.18)	(0.12)	(0.10)	(0.15)	(0.14)
Rail Dummy, 10-20 km	0.11	0.03	0.18	-0.08	-0.04	-0.21	-0.19
	(0.13)	(0.13)	(0.13)	(0.12)	(0.09)	(0.13)	(0.13)
Rail Dummy, 20-30 km	-0.01	-0.07	0.10	-0.17*	-0.12	-0.20*	-0.16
	(0.09)	(0.10)	(0.12)	(0.10)	(0.09)	(0.11)	(0.10)
Rail Dummy, 30-40 km	-0.09	-0.12	0.13	-0.16*	-0.02	-0.04	-0.04
	(0.10)	(0.10)	(0.16)	(0.08)	(0.07)	(0.07)	(0.08)
Paved Road 2002, 0-10 km		0.24***	0.21	0.23***	0.17***	0.19***	0.15**
		(0.07)	(0.13)	(0.08)	(0.06)	(0.07)	(0.06)
Improved Road 2002, 0-10 km		-0.06	-0.16	-0.05	0.06	0.07	0.04
		(0.09)	(0.16)	(0.08)	(0.06)	(0.05)	(0.05)
European Pop. 1962 (Z)				0.37***	0.11	0.08*	0.07
				(0.14)	(0.07)	(0.04)	(0.04)
Urban Pop. 1962 (Z)					0.70***	0.53***	0.54***
					(0.11)	(0.09)	(0.08)
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y
Historical Factors 1962	N	N	N	N	N	Y	Y
Contemporary Factors 2000	N	N	N	N	N	N	Y
Number of Obs.	403	403	265	403	403	403	403
R-squared	0.22	0.23	0.26	0.34	0.60	0.73	0.74

Notes: OLS panel regressions using population data on 403 non-arid locations for the year 2009. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the standard score of the population of location l that resides in localities whose population is in excess of 2,000 inhabitants in 2009. In column (2), we control for roads in 2002: we add two dummy variables equal to one if the location is less than 10 km from a paved or improved road in 2002. In column (3), the control group consists of the locations less than 10 km from a placebo line, as in column (8) of Panel B in Table 3. In column (6), we control for the historical factors in 1962 listed in section 3.5 (we use both the numbers and dummies). In column (7), we control for contemporary factors in 2000. These are the same factors as in Table 7. All regressions include province fixed effects (N = 8), and the same controls as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 8: COLONIAL RAILROADS AND CONTEMPORARY FACTORS, 2000

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 (2000)***						
Rail Dummy, 0-10 km	0.10	0.34**	0.04	1.22***	0.10***	2.26**
Rail Dummy, 10-20 km	0.00	-0.07	0.07	0.82***	0.07**	0.65
Rail Dummy, 20-30 km	-0.04	-0.05	-0.03	0.62	0.03	0.58
Rail Dummy, 30-40 km	-0.11	-0.22	0.01	0.27	0.02	-1.37
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999)***						
Rail Dummy, 0-10 km	0.00	0.18	-0.07	0.76**	0.07***	0.35
Rail Dummy, 10-20 km	-0.08	-0.21*	-0.02	0.41	0.04	-1.03
Rail Dummy, 20-30 km	-0.08	-0.12	-0.07	0.42	0.02	-0.22
Rail Dummy, 30-40 km	-0.05	-0.13	0.07	0.55*	0.04	-0.25
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 (2000)***						
Rail Dummy, 0-10 km	0.02	22.05***	0.08**	7.25***	0.47***	-0.09
Rail Dummy, 10-20 km	0.01	8.25	0.04	4.54**	0.28***	-0.15**
Rail Dummy, 20-30 km	0.01	-0.13	0.01	0.31	0.25***	-0.07
Rail Dummy, 30-40 km	-0.02	-6.22	-0.08	-1.69	0.09	-0.05
Panel B: ***Railroads and contemporary factors, conditioned on contemporary population (1999)***						
Rail Dummy, 0-10 km	0.01	2.71	0.02	2.51	0.44***	-0.13
Rail Dummy, 10-20 km	0.01	-8.80	-0.02	0.37	0.25***	-0.18**
Rail Dummy, 20-30 km	0.01	-8.25	-0.01	-1.68	0.23***	-0.09
Rail Dummy, 30-40 km	-0.01	5.19	-0.05	1.10	0.11	-0.03
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

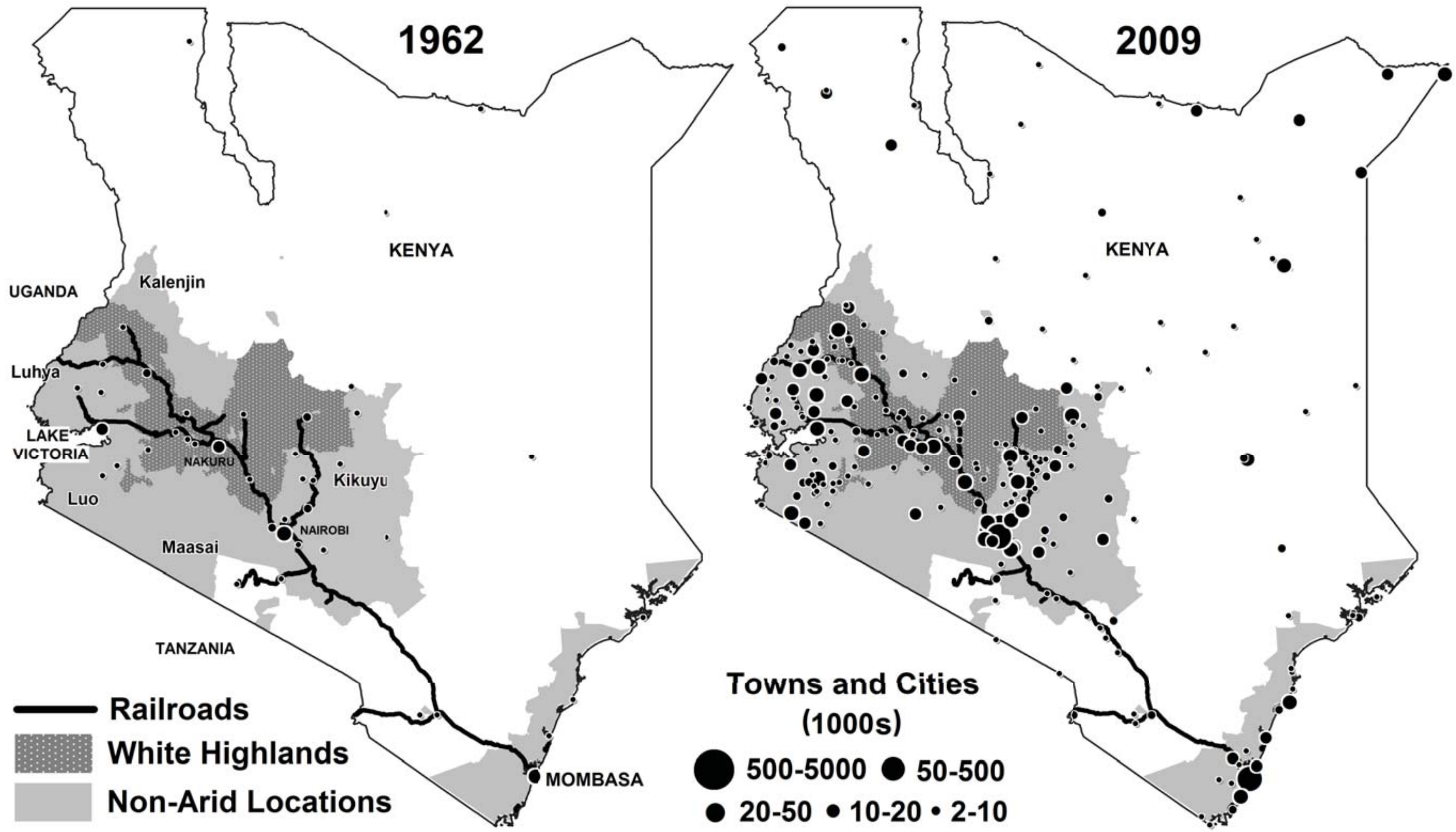
Notes: OLS regressions using infrastructure data on 403 non-arid locations for the year 2000. In the interest of space, robust SEs are not reported; * p<0.10, ** p<0.05, *** p<0.01. This table shows the effects of the rail dummies on twelve measures of contemporary factors in 2000. As these measures were close to 0 for the year 1901, these cross-sectional regressions should be interpreted as long-differenced estimations for the period 1901-2000. In Panel A, we regress each measure on the rail dummies. In Panel B, we control for contemporary population (log total population) in 1999. All regressions include province fixed effects and the same controls as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

TABLE 9: COLONIAL RAILROADS AND ECONOMIC DEVELOPMENT TODAY

Dependent Variable:	Poverty Headcount Ratio 1999 (%)		Poverty Gap 1999 (%)		Night Light Intensity 2000		ln(Lights/Area) 2000	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rail Dummy, 0-10 km	-7.81***	-6.15***	-3.10***	-2.51**	2.92***	1.00*	2.38***	1.15**
	(1.83)	(2.10)	(0.92)	(1.10)	(0.75)	(0.53)	(0.44)	(0.55)
Rail Dummy, 10-20 km	-5.95***	-3.69**	-2.63***	-1.62*	1.13***	0.16	1.57***	0.78*
	(1.49)	(1.66)	(0.77)	(0.88)	(0.37)	(0.33)	(0.42)	(0.46)
Rail Dummy, 20-30 km	-3.86***	-2.70*	-1.96***	-1.46*	0.52*	0.05	1.25***	0.36
	(1.41)	(1.52)	(0.74)	(0.80)	(0.29)	(0.27)	(0.41)	(0.44)
Rail Dummy, 30-40 km	-1.10	-1.08	-0.41	-0.50	0.24	0.12	1.00**	0.65
	(1.68)	(1.70)	(0.93)	(0.96)	(0.31)	(0.31)	(0.49)	(0.50)
Urban Pop. (Z-Score, 1999)		-0.88		-0.51		1.35**		0.39*
		(0.93)		(0.45)		(0.66)		(0.20)
Mean	51.4	51.4	18.9	18.9	1.36	1.36	0.90	0.90
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	Y
Euro. & Urban Pop. 1962	N	Y	N	Y	N	Y	N	Y
Historical Factors 1962	N	Y	N	Y	N	Y	N	Y
Contemporary Factors 2000	N	Y	N	Y	N	Y	N	Y
Number of Observations	403	403	403	403	403	403	269	269
R-squared	0.71	0.75	0.68	0.72	0.36	0.59	0.59	0.70

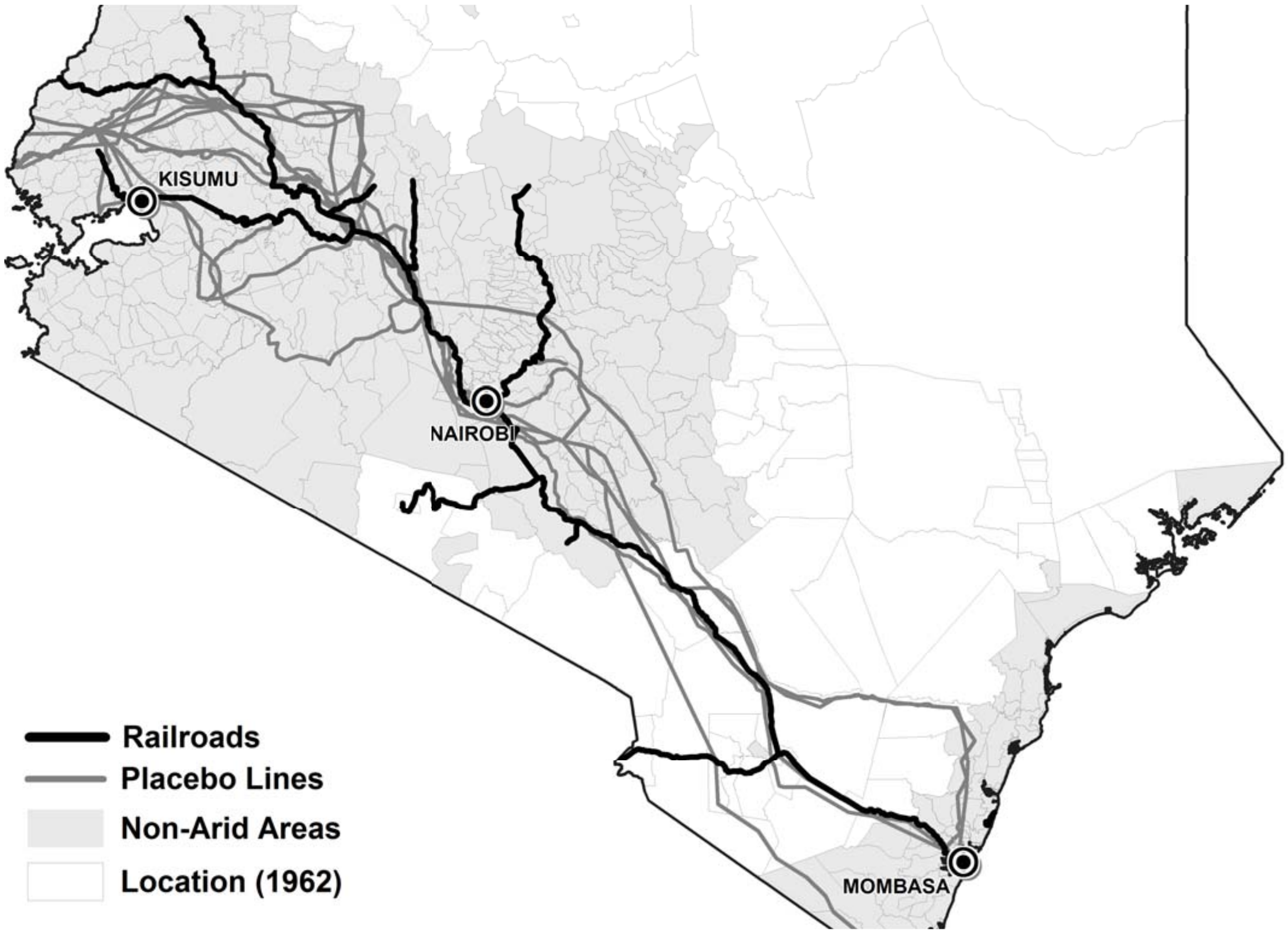
Notes: OLS regressions using data on 403 non-arid locations for the year 2000. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In columns (1)-(2), the dependent variable is the poverty headcount ratio in 1999, i.e. the percentage of the population of location l living below the national poverty line. In column (3)-(4), the dependent variable is the poverty gap in 1999, i.e. the mean shortfall from the poverty line of location l, expressed as a percentage of the poverty line. In column (5)-(6), the dependent variable is the mean night light density of location l in 2000. In column (7)-(8), the dependent variable is the log of the ratio of mean night light density to the area (thousand sq km) of location l in 2000. The measure is the same as the one used by Henderson, Storeygard & Weil (2012). In columns (2), (4), (6) and (8), the regression is similar to the regression of column (7) of Table 6, in which we include the standardised European and urban populations in 1962, the historical factors of 1962 and the contemporary factors of 2000. As we also control for standardised urban population in 1999, we are comparing locations with a same urban population. As in Henderson, Storeygard & Weil (2012), we use a relatively high threshold to be able to distinguish developed and less developed areas, so there are locations for which no light is observed. Since the variable is logged, these locations are dropped from the analysis. All regressions include province fixed effects, and the same controls as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.

Figure 1: Colonial Railroads, European Settlement and City Growth in Kenya, 1962-2009



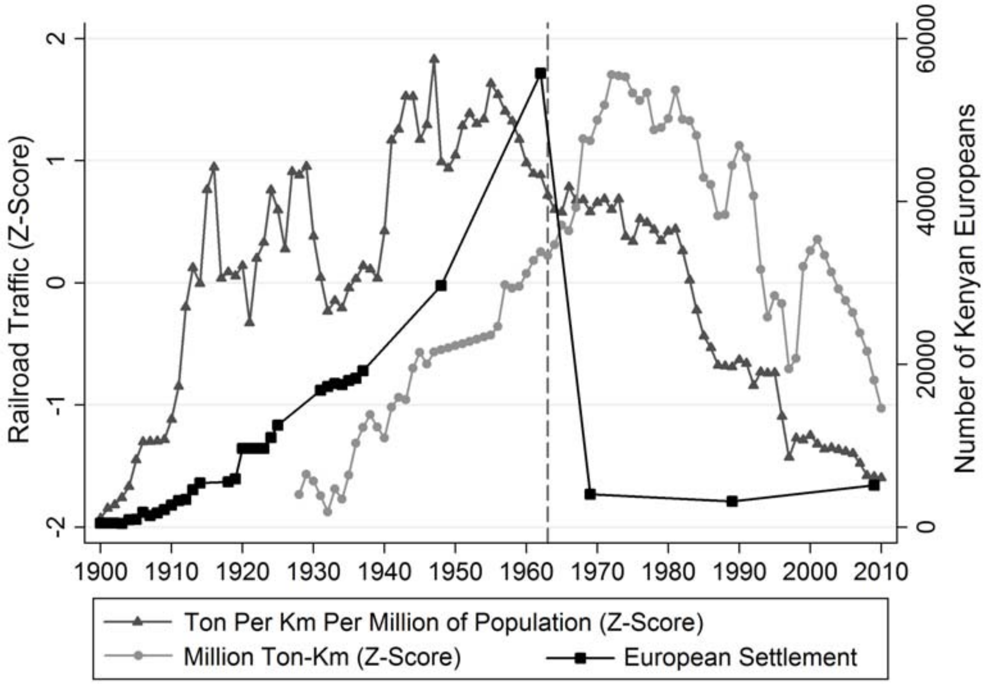
Notes: The map shows colonial railroads and cities in Kenya in 1962 and 2009. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria in the west was built in 1896-1903 to link Uganda to the coast. The non-arid areas 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 in excess to 2,000 inhabitants in 1962 (N = 42) and 2009 (N = 247). We do not have data for localities below that population threshold. There were only 5 cities in 1901, and all of them except Nairobi (Est. 1899) were on the coast. See Online Data Appendix for data sources.

Figure 2: Colonial Railroads and Placebo Lines in the Non-Arid Areas of Kenya



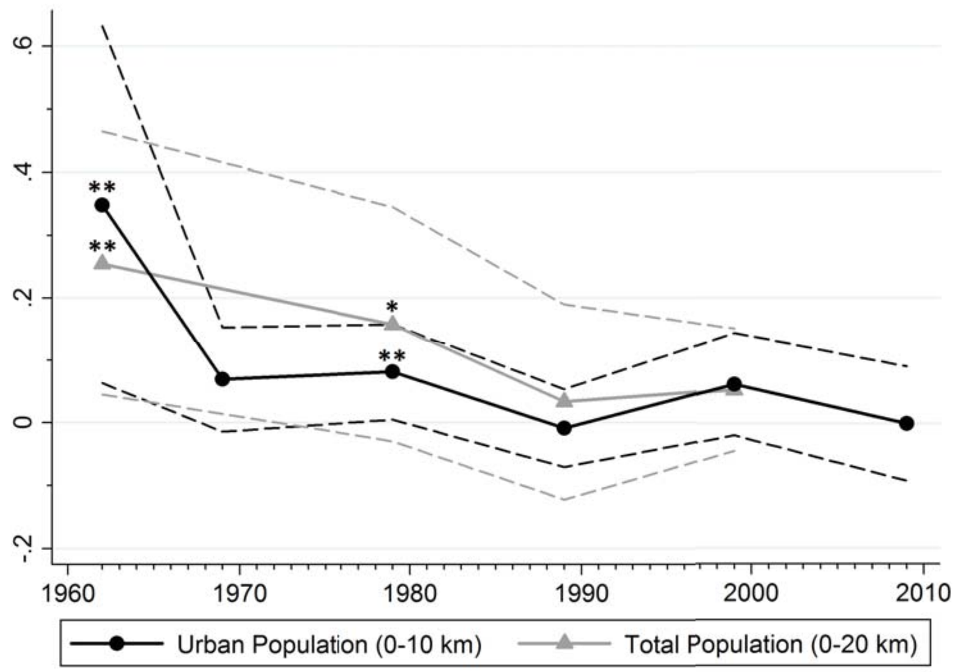
Notes: The map shows the railroad lines and the placebo lines in Kenya. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria (Kisumu) was built in 1896-1903 to link Uganda to the coast. The branch lines were built between 1913 and 1930. The map also shows several placebo lines: the Thomson route (1883), the Fischer route (1885), the Jackson route (1899), the Lugard route (1889), the Pringle route (1893), the McKinnon-Sclater route (1897), the Austin-Pringle route (1899), the Macdonald route (1899), Nakuru-Sergoit (1926), Kericho-Sotik (1926), Thika-Donyo Sabuk (1926) and Machakos (1926). Thomson, Fischer, Jackson, Lugard, Pringle, McKinnon, Sclater, Austin and MacDonald 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). See Online Data Appendix for data sources.

Figure 3: Railroad Traffic and European Settlement, 1900-2010



Notes: The graph shows the number of "Kenyan Europeans" and the standardised values of two measures of railroad traffic for selected years in 1900-2010. The first measure – *tons per km per million of population* – is the volume of goods transported by railway per capita, measured in metric tons per kilometre of rail network and per million of population. The second measure – *million ton km* – is the volume of goods transported by railway, measured in metric tons times kilometres traveled. Data is interpolated for missing years. The vertical dashed line indicates the year 1963, when Kenyan gained independence. 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 and 95% confidence intervals of repeated cross-sectional regressions (equation (4)) 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 national (African and European) census before this date. 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 ONLINE PUBLICATION: DATA APPENDIX

DATA SOURCES

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

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

Railway Data for Kenya:
We obtain the layout of railway lines in GIS from *Digital Chart of the World*. We use Hill (1950) and Ochieng and Maxon (1992) to recreate the history of railway construction. For each line, we know when it was planned, started, and finished. Placebo lines consist of explorer routes (from the coast to Lake Victoria) and branch lines that were proposed but not built. A map of these explorer routes is obtained from the *Government Survey of Kenya* (1959). A map of the branch lines is obtained from a report of the Colony and Protectorate of Kenya (1926). Most of those "placebo lines" became roads later. Using the GIS road network also available from *Digital Chart of the World*, we recreate those placebo lines in GIS. We calculate for each location the Euclidean distance (km) from the location centroid to each real or placebo line. Lastly, we create a set of location dummies equal to one if the location centroid is less than X km away from the line: 0-10, 10-20, 20-30 and 30-40 km. We create a location dummy equal to one if the location contains a rail station in 1938, whose list we obtained from one of the railway reports. We do not have data on the location of railroad stations at independence. Data on aggregate railroad traffic came from various annual reports of the Colony and Protectorate of Kenya, the East African Protectorate and the Kenya Railways Corporation, as well as World Bank (2013).

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

Population Data for Kenya:
Total population at the location level in 1962 is obtaining by entering the 1962 *Population and Housing Census*. Total population for posterior census years – 1979, 1989 and 1999 – is obtained from various GIS databases available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS). These GIS databases report the total population of each sub-location for each one of these years. Using GIS, we reaggregate the sublocations to match the location boundaries of the year 1962. Rather unfortunately, we

do not have total population data for the years 1969 and 2009. The annual total population of Kenya from 1900 to 2009 is then available on the website of Populstats, a database on the historical demography of all countries: http://www.populstat.info/.

Ethnic Data for Kenya:
Firstly, the 1962 *Population and Housing Census* reports the number of Europeans and Asians for each one of the 473 locations in 1962. The African population is then estimated by subtracting the European and Asian populations from the total population. Secondly, the 1962 census was the first exhaustive census in Kenya, whereas European censuses were conducted more or less every 5 years before independence. Rather unfortunately, the reports of these censuses do not display the distribution of the European settlers at a fine spatial level. Instead, we obtain the total number of European voters at the location level in 1919 and 1933. The *Voter Registries for the Election to the Legislative Council* of the *Official Gazette of the Colony and Protectorate of Kenya* of these years lists the name, occupation and address of all European voters for these years. The voter registries were then geospatialized using the exact address of the voters. The aggregate occupational distributions of the European settlers was then compared between the voter registries of 1919 and 1933 and the European *Population Censuses* of 1921, 1931 and 1948. Thirdly, the total number of Kenyan Europeans for other years from 1900 to 2009 is obtained from the reports of the European *Population Censuses* in the colonial period, and the reports of (exhaustive) *Population Censuses* post-independence. Fourthly, the 473 locations belong to 21 ethnic homelands when using the digitised ethnic map of Murdock (1959).

Commercial Agriculture Data for Kenya:
For each location in 1960, we know from the report European *Agricultural Census* how much land is devoted to the European cultivation of four crops (in thousand acres): coffee, maize, wheat and tea. We use these measures in 1960 as proxies for cultivation in 1962. We have no reliable spatial data on African cultivation in 1960-62. However, Africans were not allowed to grow these crops until 1954, with the exception of maize. We then have panel data on the cultivation of these crops at the district level (N = 22) every year from 1922 to 1932 (T = 11). Data on aggregate production and the composition of exports in the colonial period is obtained from the *Annual Reports of the Colony and Protectorate of Kenya* and from reports of the various European *Agricultural Censuses* that took place before independence. Data on aggregate production and the composition of exports in the post-independence period is obtained from FAO (2013).

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

Economic Geography Data for Kenya:
Firstly, we use GIS to create a "coastal location" dummy variable equal to one if the location borders the sea. Secondly, we use GIS to obtain the total area (sq km) of the location.

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

Other Transportation Networks Data for Kenya:
Roads in 1901 are described in Hill (1950) and Ochieng and Maxon (1992). The GIS maps of paved and improved roads in 1964 and 2002 are obtained from Burgess (2013). In their paper, they use Michelin paper maps to recreate the 1964 and 2002 road networks in GIS (which we use as proxies for 1963 and 2009 respectively), distinguishing paved (bitumenized) and improved (laterite) roads. We also use GIS to get the Euclidean distances of the location to a paved road or an improved road or both in 1964 and 2002.

Non-Transportation Infrastructure Data for Kenya:
We have data on health, educational, institutional, communication and industrial infrastructure at the location level in 1962. In particular, we know the number of hospitals, clinics, dispensaries, secondary schools, provincial police headquarters, divisional police headquarters, police stations, police posts, post offices and postal agencies for each location in 1962. Data on health infrastructure comes from a map of *Medical Facilities* in 1960 (which we use as a proxy for 1962) that was published by the *Government Survey of Kenya* 1959. We digitized the map using GIS and estimated the number of each type of facilities (hospitals, clinics and dispensaries) for each location. Data on education infrastructure comes from a map of Secondary Schools in 1964 (which we use as a proxy for 1962) that was published in Soja (1968). We digitised the map using GIS and estimated the number of secondary schools for each location. Data on institutional infrastructure comes from a map of *Police Organisation* in 1960 (which we use as a proxy for 1962) that was published by the *Government Survey of Kenya* 1959. We digitised the map using GIS and estimated the number of each type of police stations (provincial police headquarters, divisional police headquarters, police stations and police posts) for each location. Data on industrial infrastructure comes from a map published in Ogendo (1967). The map shows the location of the main and secondary industrial centres and towns in 1962. We digitised the map using GIS and create an "industrial centre" dummy if the location contained at least one of these industrial centres or towns. We then have data on health and educational infrastructure at the location level in 2000. In particular, we know the number of hospitals, health clinics, health dispensaries, primary schools and secondary schools for each location around 2000. Data on health infrastructure comes from a public GIS government database of *Health Facilities* in 2008 (which we use as a proxy for 2000). We use GIS to estimate the number of each type of facilities (hospitals, health clinics and dispensaries) for each location. Data on educational infrastructure comes from a public GIS government database of *Primary and Secondary* Schools in 2007 (which we use as a proxy for 2000). We use GIS to estimate the number of each type of schools for each location.

Economic Development for Kenya:
We use geospatialized poverty maps available at the sublocation level - the unit below the location – for the year 1999 to reconstruct average poverty rates at the location level for the same year. These GIS maps are available on the website of the International Livestock Research Institute (http://www.ilri.org/GIS). We use two poverty measures: the poverty headcount ratio, the percentage of the population of each location living below the national poverty line (%), and the poverty gap, the mean shortfall from the poverty line in each location, expressed as a percentage of the poverty line (%). The source of the satellite data on night lights is NOOA (2012). We follow the approach of Henderson, Storeygard and Weil

(2012) and estimate average light intensity for each location for the year 2000-2001.

REFERENCES

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

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

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

Henderson, J. 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.

Murdock, George. 1959. Ethnographic Atlas. Pittsburg, PA: University of Pittsburg Press.

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

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

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

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

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

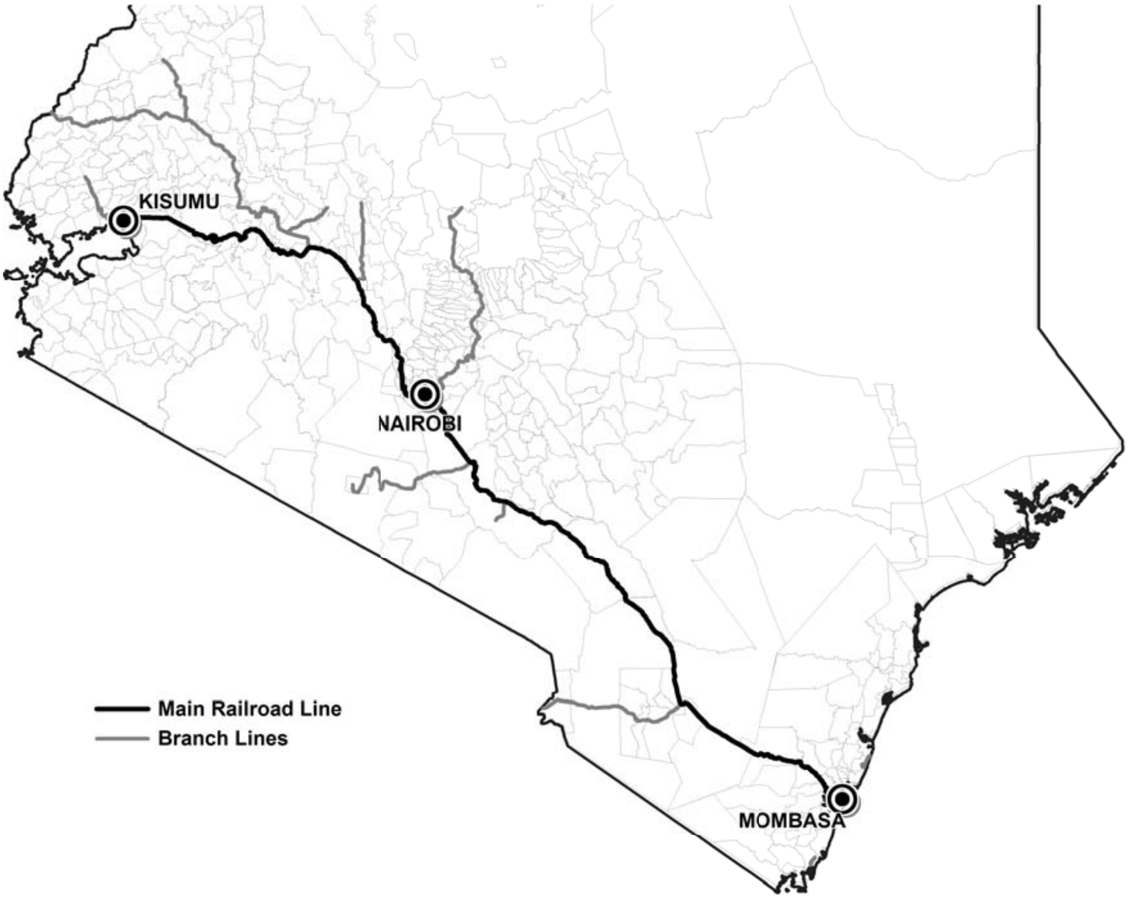
RAILROAD AND PLACEBO LINES

We now describe the various railroad and "placebo" lines that we use in our analysis. The following summaries draw on Hill (1950), Hazlewood (1964), Soja (1968) and King & van Zwanenberg (1975), as well as specific sources for each railroad or placebo line.

Railroad Lines

Main line from Mombasa to Kisumu. Acknowledging that the journey along the McKinnon-Sclater ox cart road was slow and difficult, the Imperial British East Africa Corporation (IBEAC), a royal charter company tasked with the administration of the Protectorate, sought to build a railroad from the coast to Lake Victoria (Gunston, 2004). Suffering financial losses, the IBEAC ceded control over Uganda and Kenya to the British Government in 1896 and the Kenya-Uganda mainline was constructed between 1896-1901. The "Uganda Railway" was initially named after its original destination being Uganda. It was built for three principal reasons. Firstly, for strategic and geopolitical reasons. The line shielded the region against competing European powers, by allowing the fast transportation of troops. Lake Victoria is also the source of the Nile River, and the British thought that by linking Uganda to the coast they could unify all their colonies in Northern, Eastern and Southern Africa (Appendix Figure 1 shows the map of the "Cape to Cairo Railway", a plan to unify British Africa from south to north by rail). Secondly, Uganda was seen to hold vast wealth with further trade potential. Linking Lake Victoria to the coast would open up Uganda by reducing trade costs. Thirdly, it had a deemed civilising mission, bringing Christianity and the abolishment of slavery. The construction was debated fiercely within the British parliament. Critics doubted the usefulness of the railroad "from nowhere to utterly nowhere". Since Kenya was just a transit country, the itinerary of the line was chosen in order to minimize its construction costs, i.e. the rail distance from Mombasa (the largest city of Kenya then, with 15,000 inhabitants) to Lake Victoria. Unlike the explorer and caravan routes, the railroad engineers built temporary bridges with steel girders to cross ravines and rivers. This allowed for the shortest distance rather than to follow the contours of the escarpment or human settlements. Figure A shows the itinerary of the main line. The line was first built to Nairobi, an uninhabited swamp that traces its urban origin as a railroads depot in 1899 but later became the railroads headquarters and the country's capital. Nairobi was chosen because it was a water hole (*Enkare Nyrobi* means "cold water" in Maasai) that could supply the rail construction workers with water. The construction of the railroad until Nairobi proved difficult, as the track had to be cleared through extremely dense jungle and lava rocks, cross large tsetse fly infested seasonal swamps, etc. The line then went to Kisumu on Lake Victoria, via Lake Nakuru, another source of water along the way. The track also had to cross rugged areas and dense forests between Nairobi and Kisumu, which considerably increased construction costs.

Figure A: Main and Branch Railroad Lines

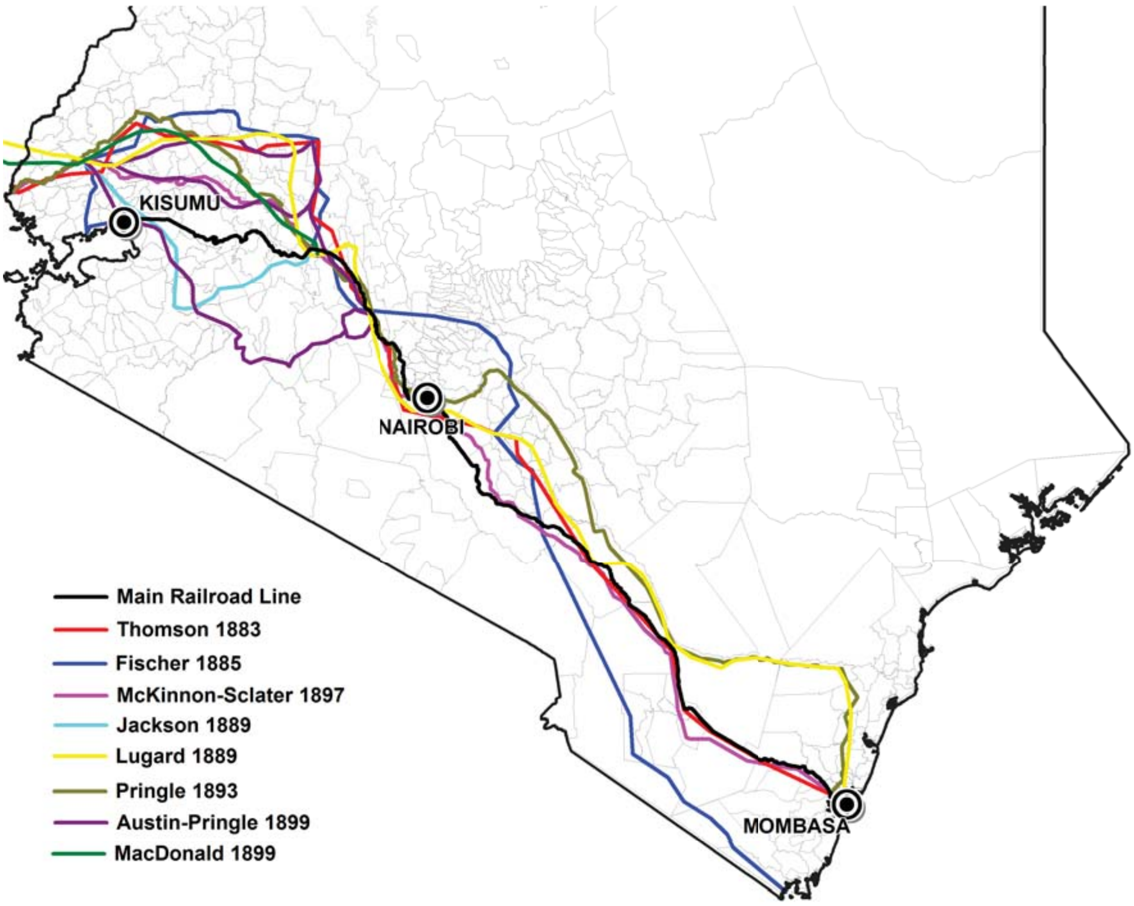


Branch lines. The Ugandan railroad established the general urban pattern of Kenya. The line produced its own nodes superseding the old caravan ones. Various branch railroad lines were constructed between 1913 and 1930. No railroad was built post-1930. Figure A also shows the location of the branch lines. While the placement of the mainline could be considered as exogenous to future population growth, the branch lines were potentially endogenous, as the colonial government sought to connect specific areas with high economic potential. The branch lines were actually not profitable, which could question the ability of the government to appraise the economic potential of various areas at the time. The following lines were the branch lines: (i) Nairobi-Thika (1913): The branch line was built to tap the fertile lands towards Mt Kenya; (ii) Konza-Magadi (1915): The branch line was built to serve the trona mines at Lake Magadi. The Madagi Soda Company was founded in 1911 to convert the trona to soda ash. According to the reports of the Blue Books, soda-ash and carbonates of soda accounted for about 9% of total exports in the colonial period. The construction of the line was funded by both the company and the protectorate; (iii) Voi-Taveta

(1918): The Northern areas of Tanzania, that was then included in German East Africa, were captured by the Allies in 1916 during World War I. Tanzania became a British Mandate in 1918, which meant that the British could now easily travel to the Kilimanjaro. They built a branch line from Voi to Taveta, at the border with Tanzania; (iv) Nakuru-Eldoret (1926): The branch line was built as the European farmers needed access to these markets, especially those beyond Eldoret; (v) Eldoret-Kitale (1926): The new branch line to Eldoret showed a lot of potential for carrying more traffic, and was planned as the main trunk line to Uganda, and, ultimately, the Congo; (vi) Rongai-Solai (1926): The branch line was built to access the Rift Valley and the agricultural lands which were dominated by coffee plantations; (vii) Eldoret-Tororo (1928): The branch line was to connect the Kenyan Protectorate with the Ugandan Protectorate by railroad. Totoro was then a village on the Ugandan side of the border. This branch line, and not Eldoret-Kitale, thus became the main trunk line to Uganda; (viii) Gilgil-Thomson Falls (1929): The branch line was built to Thomson Falls, the access point for the timber milling industry; (ix) Thika-Nanyuki (1930): The branch line was built as an extension to Nanyuki of the branch line from Nairobi to Thika. The objective was to connect the agricultural lands of Fort Hall, Nyeri and the base of Mt Kenya at Nanyuki with the railroad system; and (x) Kisumu-Butere (1930): The branch line was built to extend the mainline to Butere, in order to connect the settler areas in Yala and Butere (maize and cattle) with the railroad system.

Placebo Lines

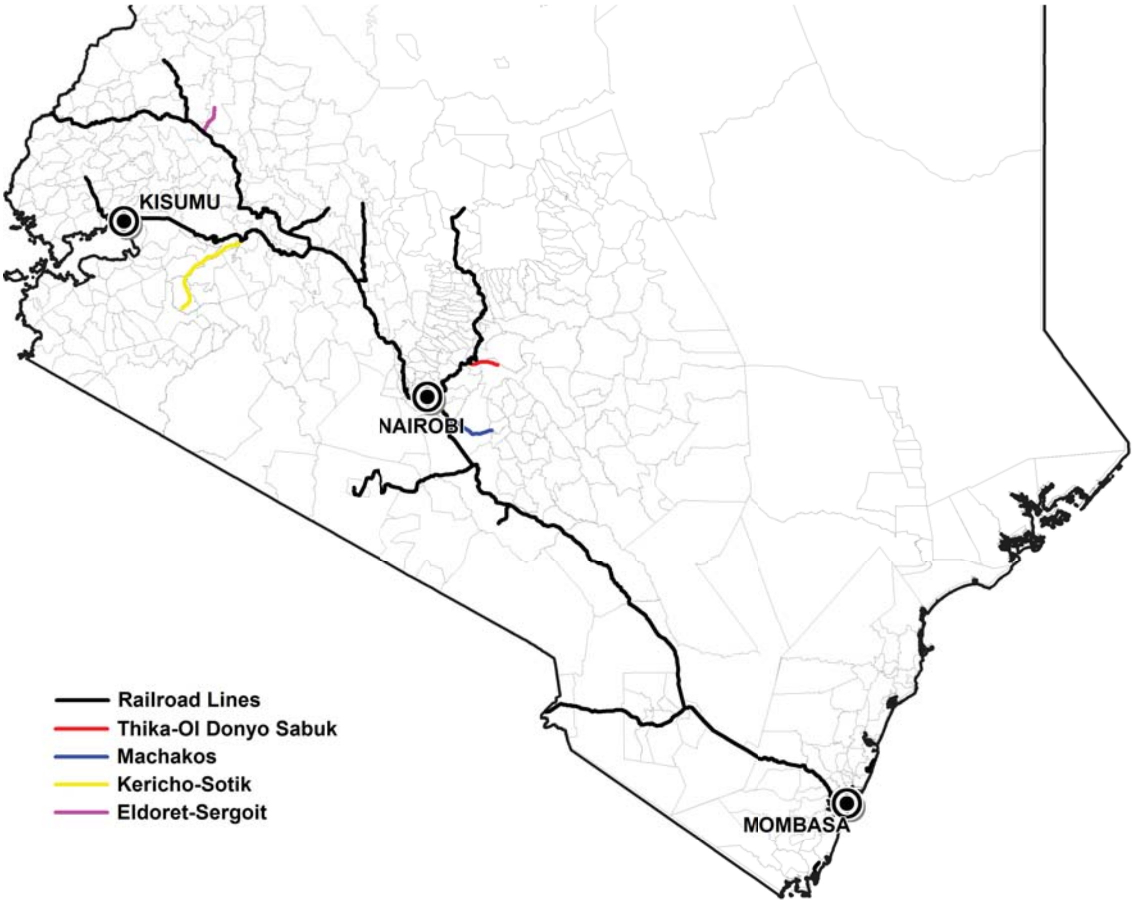
Figure B: Mainline and Explorer Routes (from Mombasa to Lake Victoria)



Explorer routes. We now describe the various explorer routes we use in our analysis. The explorer routes are taken from the map published in the *Government Survey of Kenya* (1959). We then digitised the routes using GIS. Figure B shows the mainline and each explorer route. These explorer routes from the coast to Lake Victoria provide a good counterfactual for the mainline. First, various segments of these routes could have been alternatively selected to become a segment of the mainline. Second, the explorer routes traversed areas with better locational fundamentals. They often went through more densely populated areas as a result. Therefore, as we show in the main text, the economic potential of these "placebo lines" was better than for the mainline. The following explorer routes are the "placebo lines" we consider: (i) Thomson 1883: Thomson was leading a Royal Geographical Society expedition explicitly assigned to find a trade route from the coast to the Northern end of Lake Victoria. His travel report includes a map of his route (Thomson, 1884). We ignore his excursions to Mt Kenya and Mt Elgon. Thomson joined an Arab caravan in Taveta to Njemps indicating that he was following a frequented trade route; (ii) Fischer 1885: Fischer's expedition was intended to rescue lost explorer W. Junker who was believed to be somewhere on the western shore of Lake Victoria. He organised the expedition in Zanzibar, and started ashore from Vanga, a small village on the coast. Fischer wanted to see the King of Buganda (in Uganda) asking for help, hence he planned a route to Kampala (Seegers, 2008). We exclude his journey after reaching Kisumu, which was the railroad terminus; (iii) Jackson 1889: Jackson led an IBEAC expedition intended to open up the regions between Mombasa and Lake Victoria, to mark out or establish trading stations, and to make treaties with chiefs (Hill, 1950). Jackson was instructed not to enter Uganda, for fear of becoming involved in the religious turmoil in Uganda. Jackson followed the caravan route. He reached Machakos, and camped on the north shores of Lake Naivasha a few months later. In Nakuru, the expedition deviated from the caravan route and took course to Sotik. The later expedition arrived at Mumias. (iv) Lugard 1889: IBEAC engaged Captain Lugard to open up a trade route into the hinterland (Hill, 1950). Lugard first reached Machakos, at that time the only established Company station in the hinterland (and the administrative capital of the Colony). He then built a trading station at Dagoretti, close to Nairobi. When he received instruction to proceed to Uganda, he left Dagoretti traveling by way of Naivasha to Nakuru. From there he tried to find a way to cross the Mau range, which he rightly thought was the shortest route to Lake Victoria and Uganda. As Lugard could not find a way through the dense forest, he instead followed the well-known caravan route via Baringo. He thereby missed a route, later known as Guaso Ngishu route, which railroad surveyors found excellently suited for the railway. After Lake Baringo he crossed the Kamasia range and the Uasin Gishu Plateau. He thereby abandoned the route followed by Thomson, Fischer and Jackson, who had all circled around the north of the Nandi Hills and reached Kavirondo by the valley of the Nzoia. Lugard noted that the way from the summit of the Elgeyo escarpment to the verge of the Kavirondo country was completely devoid of people. He later arrived at Mumias, where he found a garrison of 10 men and goods that had been left behind by Jackson. His journey had been more rapid than that of any previous caravan. Lugard stressed that, at Mumias, large quantities of flour were brought in for sale. He took the cheap rate at indication of the economic potential of the territory. Lugard then reached Kampala, the capital of Uganda; (v) Pringle 1893: He led an expedition from Mombasa to Uganda. He first followed the usual caravan routes to Machakos. He then went from Machakos to Naivasha, via Nairobi. At Naivasha, Pringle hesitated between three main routes to Uganda (Pringle, 1893). He chose the one that went through more accessible territory; (vi) McKinnon Sclater 1897: The construction of the McKinnon-Sclater road, an ox cart track from Mombasa to the Ugandan border, was started in 1890 by the IBEAC. However, the road did not reach Uganda before 1896, the year when rail construction began. The track was "of the simplest kind, unmetalled, and in fact, the roughest track along which a bullock-cart would go"(Smith, 1899). The road did not strongly reduce trade costs, as the journey was slow and difficult; (vii) Austin Pringle 1899: Captains Pringle and Austin, both of the Royal Engineers, did extensive survey work in the west of Kenya. They explored three routes; and (vii) Macdonald 1899: IBEAC made him the "Chief Engineer" of the railway survey. In 1899, MacDonald received news that the fighting in Uganda was over and that it was again safe for a caravan to cross the Nile. He thus led an expedition from Nakuru to Uganda via Mumias (MacDonald 1899a, 1899b).

Branch lines that were planned but not built. We now describe the various branch lines

Figure C: Branch Railroad Lines and Branch Lines Planned but not Built



that were proposed in 1926 but not built. Figure C shows the railroad lines and the proposed branch lines. We argue that the economic potential of the branch lines that were built and the branch lines that were planned but not built was very similar. Some of the proposed branch lines were dismissed because they did not have enough promise to be profitable. However all lines that were actually built were unprofitable, including in 1929 before the economic crisis (Kenya and Uganda Railways and Harbours (1931)). This could question the ability of the government to appraise the economic potential of various areas and profitability of specific lines at the time. These lines should thus provide a good counterfactual for the railroad lines. The following lines were the branch lines that were planned but not built (see Colony and Protectorate of Kenya (1926) for a description of each line): (i) Eldoret-Sergoit 1926: The branch line was proposed to access the agricultural lands of the Nzoia River with crops of maize, wheat and the potential for cattle ranching and coffee cultivation. The report of the proposed line found that the line showed great potential for European settlement. The report concluded that the line "would ultimately pay and lead to a more intensive development of an area which is undoubtedly attractive". Yet it never materialised; (ii) Machakos 1926: Machakos was established in 1887, and was the first capital of the Colony. The town lies 34 km southeast of the mainline. The construction of the railroad and the relocation of the capital to Nairobi resulted in the isolation of this European settler enclave, hence the need for a branch line. The report however notes that the proposed line would serve no extensive area of importance, as the area around Machakos had not much potential would Machakos had not been a major town then; (iii) Kericho-Sotik 1926: The branch line was motivated as a viable line for its agricultural output (coffee, tea and cereals) and its potential for the processing of large quantities of bamboo from the forest belt for paper pulp. Another motivation was that the line would provide access to various African areas, which would boost trade between Europeans and the natives. This potentially rich area suffered from isolation with "the distance from the railway (about 65 miles) being too great for crops of a bulky nature to be attempted." Although these areas showed high potential, the line was never built, because expected construction costs were also high. The line necessitated "a considerable amount of heavy earthworks, and probably viaducts." The report concludes "The most optimistic expectations of increased output from the Kericho and Sotik area as it stands today, would not show enough traffic to warrant the construction of a railway on irs own account"; and (iv) Thika-Ol Donyo Sabuk 1926: Ol Donyo Sabuk lay within close proximity to the existing railway siding at Thika. The objective of the line was to open up more land for the cultivation of coffee. However, the report said that the branch line was unlikely to have any effect, as it was too close to the existing branch line from Nairobi to Thika.

REFERENCES

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

Gunston, H. (2004). "The Planning and Construction of the Uganda Railway." *Transactions of the Newcomen Society* 74: 45-71.

Hill, M. F. (1950). *Permanent way*. Nairobi: East African Railways and Harbours.

Kenya and Uganda Railways and Harbours. (1931). *General Manager's Bulletin No 1*. Nairobi: Office of the General Manager of Railways and Harbours.

MacDonald, J. R. L. (1899a). Journeys to the North of Uganda. I. *The Geographical Journal* 14(2): 129-148.

MacDonald, J. R. L. (1899b). Notes on the Ethnology of Tribes Met with During Progress of the Juba Expedition of 1897-99. *The Journal of the Anthropological Institute of Great Britain and Ireland* 29(3/4): 226-247.

Pringle, J. W. (1893). With the Railway Survey to Victoria Nyanza. *The Geographical Journal* 2(2), 112-139.

Seegers, L. (2008). "The Fishes Collected by G. A. Fischer in East Africa in 1883 and 1885/86." *Zoosystematics and Evolution*, 84(2), 149-195.

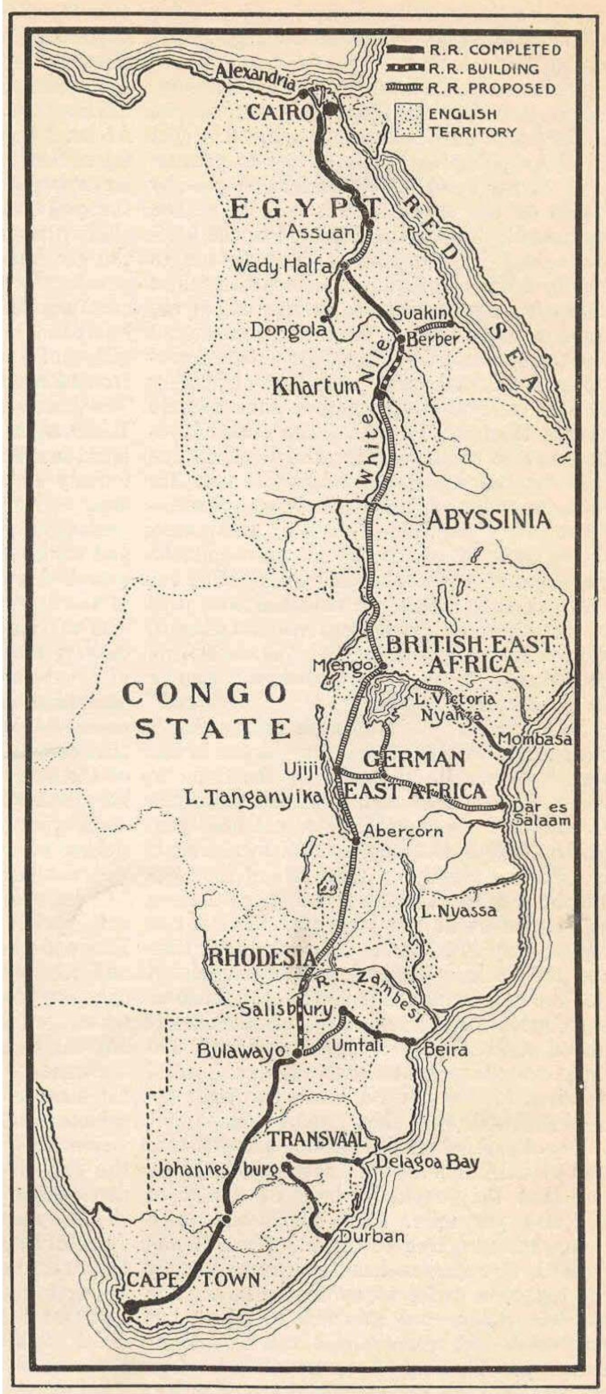
Smith, R. E. (1899). "Road-Making and Surveying in British East Africa." *The Geographical Journal* 14(3): 269-289.

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

Thomson, J. (1884). *Through the Masai Country to Victoria Nyanza*. Proceedings of the Royal Geographical Society and Monthly Record of Geography 6(12): 690-712.

FOR ONLINE PUBLICATION: APPENDIX FIGURES AND TABLES

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



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

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

Rail/Placebo	**Rail**	**Placebo**	**Col. (3)-(10): Explorer Routes**								**Col. (11)-(14): Planned But Never Built**			
Placebo Line:		All	Thomson 1883	Fischer 1885	Jackson 1889	Lugard 1889	Pringle 1893	McKinnon Sclater 1897	Austin Pringle 1899	Macdonald 1899	Eldoret Sergoit 1926	Machakos 1926	Kericho Sotik 1926	Thika ODS 1926
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Panel A:	*Dependent Variable*: **Number of European Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km	0.60***	0.31**	0.57**	0.5	0.44	0.50*	0.23	0.05	-0.15	0.25	1.47	0.25	0.1	0.84
	(0.21)	(0.14)	(0.24)	(0.39)	(0.30)	(0.26)	(0.24)	(0.12)	(0.15)	(0.33)	(1.08)	(0.23)	(0.25)	(0.68)
Panel B:	*Dependent Variable*: **Number of European Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km		0.17	-0.03	0.98	0.63	0.38	-0.05	0.10	0.02	0.65	3.27	0.83	0.11	2.47
Drop if Rail, 0-10 km = 1		(0.11)	(0.04)	(0.75)	(0.54)	(0.25)	(0.12)	(0.15)	(0.16)	(0.62)	(2.69)	(0.75)	(0.21)	(1.86)
Panel C:	*Dependent Variable*: **Number of Urban Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km	0.35**	0.21**	0.43*	0.58**	0.41	0.17	0.16	0.00	0.39	0.37	0.26	0.04	0.54	0.53*
	(0.14)	(0.08)	(0.23)	(0.29)	(0.28)	(0.12)	(0.13)	(0.12)	(0.35)	(0.23)	(0.29)	(0.29)	(0.45)	(0.29)
Panel D:	*Dependent Variable*: **Number of Urban Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km		0.02	0.08	0.00	-0.11	0.08	0.07	0.04	-0.14	-0.16	0.03	0.42	-0.09	0.15
Drop if Rail, 0-10 km = 1		(0.06)	(0.05)	(0.08)	(0.10)	(0.11)	(0.09)	(0.08)	(0.09)	(0.10)	(0.17)	(0.31)	(0.19)	(0.17)
Panel E:	*Dependent Variable*: **Number of Total Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km	0.25**	0.03	0.23**	-0.02	-0.03	0.06	0.23	0.20	0.14	0.19	-0.27*	1.24	-0.19*	-0.03
	(0.11)	(0.07)	(0.10)	(0.06)	(0.10)	(0.06)	(0.19)	(0.18)	(0.17)	(0.13)	(0.15)	(1.01)	(0.11)	(0.11)
Panel F:	*Dependent Variable*: **Number of Total Inhabitants in 1962 (Z-Score)**													
Rail/Placebo, 0-10 km		0.13	0.13	0.05	0.25	0.16	0.28	-0.20	0.12	0.47	-0.09	0.28	0.7	0.19
Drop if Rail, 0-10 km = 1		(0.10)	(0.13)	(0.22)	(0.32)	(0.15)	(0.18)	(0.18)	(0.35)	(0.33)	(0.20)	(0.28)	(0.56)	(0.25)
Province FE, Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

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. This table tests that there are no spurious effects for twelve placebo lines in 1962. In Panels A and B, the dependent variable is the standard score of the European population in 1962. In Panels C and D, the dependent variable is the standard score of the urban population in 1962. In Panels E and F, the dependent variable is the standard score of the total population in 1962. Column (1) replicates the main results of Table 2, using the 0-10 km rail dummy for the sake of simplicity. In columns (2)-(14), we test that there are no effects when using the placebo lines instead. In Panel (B), (D) and (F), we drop the locations less than 10 km from a railroad line, in order to compare the placebo locations with the other control locations (N = 323), while suppressing the effects from the railroad lines. In column (2), we consider all placebo lines. In columns (3)-(10), we consider eight explorer routes as placebo lines. In columns (11)-(14), we consider four branch lines that were planned but not built. Thika-ODS stands for Thika-Ol Donyo Sabuk. All regressions include province fixed effects (N = 8), and the same controls as in Table 2. The nodes are dropped from the analysis. See Online Data Appendix for data sources.