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

Remi Jedwab† and Edward Kerby‡ and Alexander Moradi§

This Version: December 2013

Abstract: Little is known about the extent and forces of path dependence in developing countries. We examine whether locational fundamentals (i.e., the geographical endowments of various locations) or increasing returns (i.e., localized historical shocks) are the main determinants for the distribution of economic activity across space in these countries. The construction of the colonial railroad in Africa provides a natural experiment to study the emergence, persistence and optimality of an urban equilibrium. Using fine spatial data for Kenya over one century, we find that colonial railroads had a strong causal impact on the location and number of European settlers. Economic development in the European areas in turn determined the location of the main cities of the country at independence. Second, railroads fell into disrepair and most settlers left in the immediate post-independence period, yet these patterns of economic geography persisted. While colonial sunk investments (e.g., schools, hospitals and roads) partly contributed to path dependence, railroad cities mainly persisted because their early emergence served as a mechanism to coordinate contemporary investments in the subsequent period. This shows the stability of the equilibrium. Third, the railroad locations have worse geographical fundamentals than other locations that could have counterfactually received the cities, which suggests the equilibrium may not be optimal. Our findings are important as they inform regional economic policy in shaping future economic progress.

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

*We are grateful to François Bourguignon, Richard Hornbeck, Stephen Smith, Adam Storeygard, Anthony Yezer and seminar audiences at the African Economic History Network Meeting, Cambridge (RES), Cape Town, the Economic History Association Meeting, George Washington, Maryland, 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.

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 and economic geography 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. For example, 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. As explained by Davis & Weinstein (2002, 2008), the fact that Hiroshima and Nagasaki recovered their population and industries strongly argues in favour of the first theory. There is only one spatial equilibrium, and regional policy interventions, no matter how large they are, cannot undo it in favour of another equilibrium. Yet the literature has also shown how large (or in some cases, even small) localised shocks or events can have permanent effects (Bosker et al., 2007, 2008; Redding, Sturm & Wolf, 2011; Bleakley & Lin, 2012; Holmes & Lee, 2012; Michaels & Rauch, 2013). These studies demonstrate strong evidence in favour of the emergence of multiple spatial equilibria, which implies that regional policy interventions may impact regional development.

In parallel to the economic geography literature, the development literature has debated whether the underdevelopment of various parts of the world (e.g., sub-Saharan Africa) is better explained by physical geography or history, 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; Rappaport & Sachs, 2003; Beeson, DeJong & Troesken, 2001; Sachs, 2003; Bosker & Buringh, 2010; Maloney & Caicedo, 2012; Motamed, Florax & Matsers, 2013). Conversely, various articles have shown how historical shocks may shape the patterns of development, for a given physical geography (Acemoglu, Johnson & Robinson, 2001, 2002; Banerjee & Iyer, 2005; Nunn, 2008; Dell, 2010, 2012; Nunn & Puga, 2012; Maloney & Caicedo, 2012). Obviously, the effects of a specific physical geography will depend on the technologies used in the economy, and technological progress may itself be conditioned by the quality of institutions. Therefore, the spatial patterns of development result from the interaction of physical geography, technology and institutions.

How do history and path dependence interact with the process of economic development? To shed light on this question, imagine an unurbanised country. Locational fundamentals and increasing returns can explain why an urban equilibrium, i.e. the birth and growth of a network of cities, can emerge, and why it can persist. The *emergence* of the equilibrium can be due to locational fundamentals, such as the proximity to a coast or a river, being situated in a fertile plain, etc. The literature has shown how these natural advantages are the main determinants of the distribution of economic activity across space at the beginning of the development process. A historical shock can also put people together in selected locations. If there are increasing returns, people stay where they are, and the locations keep growing. The *persistence* of an urban equilibrium could be due to serial correlation, the fact that locational fundamentals have a continuous effect on the spatial distribution of economic activity, or path dependence, the fact that increasing returns solidify the initial advantages of the existing cities. There are various available tests of the path dependence hypothesis. For example, wars are sufficiently large negative shocks that we should expect them to have permanent effects. If destroyed cities recover their population, this suggests that locational fundamentals are the main determinants of spatial economic development. However, the fact that cities arising as a result of a natural or man-made advantage may persist after this advantage is eroded provides strong evidence for increasing returns and path dependence. As explained by Bleakley & Lin (2012), given that fixed costs are a source of increasing returns, *sunk investments* (e.g., schools, hospitals and roads) could account for path dependence. 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. The created cities solve this problem for the first period and each subsequent period. Lastly, as shown by Michaels & Rauch (2013), a historical shock, combined with increasing returns, may trap people in locations that were not ex-ante optimal, given the distribution of locational fundamentals. The persistence of an equilibrium does not ensure its *optimality*.

While more empirical evidence is needed on the mechanisms by which an urban equilibrium can emerge, persist and maximize welfare, there are few natural experiments that allow to investigate such mechanisms. Then, little is known about the extent and nature of path dependence in the specific context of developing countries. Studying this issue is all the more important that the growth of cities is essential for the development process in a country. Unfortunately, these countries are not only poor, but they also suffer from data shortage. This limits our understanding of the relationship between history, path dependence and development.

To address these difficulties, we use a natural experiment and a new data set on railroads, European settlement and city growth at a fine spatial level over one century (473 locations of about 16x16km, in 1901-2009) in Kenya. All of Kenyan railroad lines were built during the colonial era. We find that they had a strong causal impact on the location and number of European settlers. Economic development in the European areas in turn determined the location of the main cities of the country at independence. Railroads then fell into disrepair and most settlers left in the immediate post-independence period. Yet colonial railroads had long-term, permanent effects on urban and development patterns.[1] We use this natural experiment to

[1]As argued by Lucas (1988), cities are the main engines of growth. Urban growth is our primary measure of economic development, in line with the literature (Bairoch, 1988; Acemoglu, Johnson &

show how a historical shock can create an urban equilibrium, why this equilibrium may persist as a result of path dependence, even it may not be the most optimal equilibrium in the long run. The Kenyan setting is attractive for four reasons.

First, we provide evidence that the placement of the colonial railroad was exogenous to the population booms. Historical transport costs in Kenya were extremely high. Kenya lacked navigable waterways. Draft animals could not be used due to the Tsetse fly transmitting trypanosomiasis. There were only a few tracks, that were not developed into roads before 1945. Kenya exported goods which were head-loaded on short distances, or slaves who walked longer distances. Economic change was limited to the coast. The British coloniser sought to build infrastructure to penetrate to the hinterland, mostly to ensure military domination. The British built a railroad from Mombasa, on the coast, to Lake Victoria, in the west, in 1901. Although the line crossed the country from east to west, the ultimate objective was to link the coast to Uganda (across Lake Victoria), where the Nile originates. The railway was even called the "Uganda Railway", although it was not located in Uganda. The British were then able to travel from Cairo to the eastern coast of Africa, which was an asset at the time of the scramble for Africa. The line (and its branch lines) went through areas that were lowly-populated, for historical reasons. We show that the decrease in trade costs had a strong (and mainly unexpected) impact on the settlement of Europeans, who established cities from where they managed their farms and specialised in urban production activities. African population also increased along the lines, to work on European farms and in European cities.

Second, even if placement was not exogenous, Kenya's history gives us various identification strategies we exploit to confirm our effects are causal. We find no effects for a set of "placebo lines" that consist of various explorer routes from the coast to the west and branch lines that were planned but not built. Besides, as we have data at a fine spatial level, we can include many ethnic group fixed effects, district fixed effects, or even a quartic polynomial in longitude and latitude to identify our effects by comparing connected locations with only contiguous locations that were not connected. These strategies all give similar results. Lastly, even if the mainline was exogenous, the branch lines were potentially endogenous, as the colonizer 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, due to mismanagement and lack of maintenance in the rail sector, as well as massive investments in roads. While few locations of the hinterland were connected to the coast in the colonial period, almost all locations are now connected, and an entirely different spatial distribution could have emerged post-1963. Europeans were then asked by the first African president of the country to choose the Kenyan citizenship. Their refusal to give up their British citizenship caused their massive exodus from Kenya. Lastly, coffee was the main export of the colony, and the engine of growth of the European areas. Coffee production then collapsed in the 1980s, owing to low international prices. Locations along the railroad lines thus lost their initial advantage in terms of transportation, human capital and commercial 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,

Robinson, 2002; Bosker & Buringh, 2010; Dittmar, 2011; Jedwab & Moradi, 2013).

but is due to path dependence. Indeed, the long-run effects of colonial railroads are explained by colonial urbanisation. We have also collected large amounts of data on colonial and post-colonial infrastructure. We find that, while colonial sunk investments matter, railroad cities mostly persisted because their early emergence served as a mechanism to coordinate contemporary investments in the subsequent period. Finally, we hypothesize that cities would have eventually emerged in the African areas that were relatively more densely populated before the colonial era. We find that the railroad locations have worse locational fundamentals than these areas, which indicates that the colonial urban equilibrium may be suboptimal today.

We believe our paper makes a valuable contribution to the literature. Indeed, our natural experiment allows us to examine the causal emergence, persistence and optimality 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 either theoretically or empirically demonstrated how such investments may have permanent spatial effects. We find strong evidence for path dependence, in line with many papers of the literature, but in contradiction with Davis & Weinstein (2002, 2008) and Miguel & Roland (2011). Our data set also helps us disentangle the mechanisms of path dependence. As in Bleakley & Lin (2012), sunk investments are not the main force of path dependence, while the coordination problem of contemporary factors is. This shows the stability of the equilibrium. We then explain that the resulting equilibrium may be suboptimal, as it is associated with worse locational fundamentals. Nunn & Puga (2012) and Michaels & Rauch (2013) also provide interesting examples of spatial suboptimality.

Then, we focus our analysis on a developing country, for which the importance, extent and forces of path dependence could differ. First, developing countries are by definition poorer and more rural. Studying a poor country allows us to additionally analyse the emergence of the urban equilibrium, whereas most existing studies have tested if an existing equilibrium persists in the face of a shock. The paper is closely related to Bleakley & Lin (2012) who investigates the effects of portage over two centuries in the U.S. Second, the extent of path dependence may also differ. Locational fundamentals could play a larger role in more agrarian countries. Since these countries are more rural, any spatial shock could also have larger effects. These effects will be permanent if there are strong increasing returns. The literature has then shown how cities are centers of agglomeration effects and human capital accumulation in rich countries (Rosenthal & Strange, 2004; Glaeser & Gottlieb, 2009). Industrial agglomeration effects and human capital externalities may be more limited in poor countries, which could reduce the forces of increasing returns and path dependence. It is thus not obvious whether we should expect stronger path dependence in developing countries than in developed countries. Third, sunk investments may be less important in poor countries than in richer countries, as the materials used to build cities, schools, hospitals and roads are of poorer quality. However, there are fewer sunk investments per capita in poor countries, so that the locations that benefited from an early start may enjoy their initial advantage for a longer period. The coordination problem could also be less important, if the usual forces of agglomeration are more limited in such countries. However, the fact that these countries are poorer imply that many coordination problems have not been solved yet for various of their regions, which constrains local and overall capital accumulation. Spatial investments may solve these 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 mentioned as the main obstacle to trade expansion, and thus economic growth, in Africa (Rodrik, 1998; Buys, Deichmann & Wheeler, 2010). Transportation infrastructure can indeed facilitate the circulation of goods, people and ideas. Railroads have boosted exports in Kenya, in line with the literature on transportation and trade (Michaels, 2008; Duranton & Turner, 2012; Donaldson, 2013; Faber, 2013; Donaldson & Hornbeck, 2013). They have encouraged the movement of workers and firms, in line with the literature on transportation and population and employment growth (Baum-Snow, 2007; Atack et al., 2010; Banerjee, Duflo & Qian, 2012; Ghani, Goswami & Kerr, 2012). They have promoted the diffusion of innovations, here the establishment of cities and the adoption of a new crop, in line with the literature on information and communication technology and development (Jensen, 2007; Dittmar, 2011). Additionally, they can trigger an equilibrium in which cities emerge to facilitate the accumulation of factors, and thus have long-term effects on economic growth (Bleakley & Lin, 2012; Jedwab & Moradi, 2013). Few papers have studied the causal impact of transportation in Africa, with the exception of Storeygard (2012). Burgess et al. (2013) show how road building is driven by political considerations (ethnic favoritism) instead of economic considerations, which lowers the returns to such investments.[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 shows the persistence of this equilibrium, and studies its optimality. Section 6 concludes.

2. CONCEPTUAL FRAMEWORK

In this section, we offer a conceptual framework to theoretically analyze the relationship between history, path dependence and economic development. In particular, we reinterpret the path dependence literature to describe how an urban

[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).

equilibrium, i.e. the birth and growth of a network of cities, can emerge (what we call *urban emergence*), why it can persist (what we call *urban persistence*), and whether it is optimal (what we call *urban optimality*). We focus on the respective roles of locational fundamentals (i.e., the geographical endowments of various locations) and increasing returns (i.e., localized historical shocks) in shaping economic geography. We discuss how our natural experiment fits into this framework.

2.1 Locational Fundamentals vs. Increasing Returns

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

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

Increasing returns ($f'_P > 0$) can give rise to multiple spatial equilibria, as a dense location is likely to grow further. In this case, 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 have emerged could result from a bad geography (a low G) or a low economic return to a specific geography given the existing technologies in the economy (a low f'_G).

2.2 Urban Emergence

The emergence of an urban equilibrium, where no such equilibrium existed before, can be due to locational fundamentals or increasing returns. Locational fundamentals are "first nature" or "natural" advantages, such as the proximity to a coast or a river, being situated in a fertile plain or valley, etc. We know that the rise of complex civilizations was conditioned by their geographical environment (Bairoch, 1988; Diamond, 2011; Motamed, Florax & Matsers, 2013). Likewise, the literature has shown how the location of cities in Europe and the U.S. was strongly influenced by natural characteristics (Beeson, DeJong & Troesken, 2001; Bosker & Buringh, 2010; Bleakley & Lin, 2012). Increasing returns are then permitted by "second nature" or "man-made" advantages, such as having already built a city or being connected to transport networks (railroads, roads and airports). An urban equilibrium can emerge even without good locational fundamentals, if a historical shock puts people together in selected locations. If there are strong increasing returns (if f'_P is large enough), people stay where they are, and the locations keep growing. This paper argues that colonization provides such a shock, as colonial powers disproportionately "invested" in a few locations. In particular, the construction of the colonial railroad gave an early advantage to selected locations. Nunn & Puga (2012), Maloney & Caicedo (2012), Jedwab & Moradi (2013) and Bonfatti & Poelhekke (2013) also show how colonization caused a specific spatial equilibrium.

2.3 Urban Persistence

If an urban equilibrium persists over time, it could be due to serial correlation, the fact that locational fundamentals have a continuous effect on the spatial distribution of economic activity, or path dependence, the fact that increasing returns solidify the initial advantages of the existing cities. For example, many cities in Europe and the U.S. today are located along rivers or the coast, two natural advantages in terms of transport, although transport costs have considerably decreased over time. The persistence of these cities is still explained by serial correlation if these locational fundamentals are now valued as consumption amenities instead of production amenities (Rappaport & Sachs, 2003). People now live in these cities because of the quality of life they offer. If these natural advantages no longer matter (whether as production or consumption amenities), urban persistence must be due to increasing returns (Bosker & Buringh, 2010; Bleakley & Lin, 2012).

There are various available tests of the path dependence hypothesis. First, if serial correlation explains urban persistence, a historical shock should have no permanent effect on population. Various studies have examined the long-term effects of conflict on population patterns and economic development. For example, the war time bombings of Japan, Germany and Vietnam provide such a test. The fact that many destroyed cities recovered their population and industries indicates that serial correlation accounts for urban persistence (Davis & Weinstein, 2002, 2008; Miguel & Roland, 2011). However, Bosker et al. (2007, 2008) and Redding, Sturm & Wolf (2011) for Germany, and to a lesser extent Dell (2012) for Mexico, find a permanent effect of conflict on development. Second, if path dependence explains urban persistence, a geographical shock causing a change in locational fundamentals should have no permanent effect on population, as increasing returns will "protect" the affected locations. For example, 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 have become more important in the modern period, and the loss of a natural advantage may have lower effects now, as in Bosker & Buringh (2010) and Bleakley & Lin (2012). Lastly, whether we use exogenous changes in G or P to test the path dependence hypothesis, the persistence of an urban 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). Unfortunately, there are few natural experiments with a shock large enough to test the hypothesis.

There are then two main mechanisms of path dependence. As explained by Bleakley & Lin (2012), path dependence could be due to *sunk investments* or the *coordination problem* of contemporary factors. Firstly, given that fixed costs are a source of increasing returns, *sunk investments* (e.g., houses, 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. In other words, past population P_{t-1} could matter ($f_P' > 0$) because it proxies for past investments, or because it sends a signal to all contemporary factors that they should co-locate in this specific location. It will be important later on to disentangle the effects of both channels. Indeed, public policy recommendations will depend on which force of path dependence dominates.

2.4 Urban Optimality

If there are multiple spatial equilibria due to increasing returns, the existing urban equilibrium may be suboptimal. For example, if locational fundamentals or their value have changed over time, an urban 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. This may prevent any spatial redistribution, no matter how profitable it would be for everyone. Similarly, an urban equilibrium arising from a historical shock is suboptimal if the new urban distribution does not match the distribution of the locational 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.

This obviously assumes that relocation costs are not too high. In these conditions, we understand why a suboptimal (but efficient) equilibrium may persist. Nunn & Puga (2012) and Michaels & Rauch (2013) provide two interesting examples of spatial suboptimality. Nunn & Puga (2012) find that various African populations permanently moved to rugged areas to protect themselves against raids during the slave trade. As ruggedness usually 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, thus resulting in a suboptimal urban distribution.

There are two caveats to this analysis. First, a suboptimal equilibrium today may have been "dynamically" optimal. As we will discuss later, Kenya, and more generally Africa, 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 economy as a whole. Thus, even if the urban equilibrium created by colonization is suboptimal today, things would have been very different without this initial equilibrium. Second, technological change constantly modifies the value of locational fundamentals. An optimal distribution at one period may be suboptimal the next period, and so on. The optimal equilibrium may be continuously evolving, which may limit the potential benefits from any spatial reallocation.

2.5 Colonial Railroads in Kenya as a Natural Experiment

We argue that the construction of the colonial railroad in Kenya provides a natural experiment to study the emergence, persistence and optimality of an urban 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 was a historical shock that disproportionately favored some locations

($\Delta P > 0$). These locations became the main cities of the country at independence (high P). (ii) *Urban persistence*: The railroad locations lost their initial advantage in the immediate post-independence period, which should have led to a relative decline in their population ($\Delta P < 0$). Yet we find that these locations are relatively more urban and developed today (high P). (iii) *Urban optimality*: We compare the railroad locations with the locations that were densely populated by Africans before the colonial era, as these locations could have counterfactually received the main cities of the country. We find that the railroad locations are worse in terms of locational fundamentals (lower G). This points to urban suboptimality.

One question is whether our natural experiment provides a lower test of the path dependence hypothesis than the rest of the literature. Firstly, war bombings may not provide the ultimate test. While they killed people and destroyed buildings, many factors such as road networks or the reputation of a city 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, i.e. spatial interdependencies between cities. More importantly, we do not observe a pure market equilibrium in the post-war period as the governments of these countries massively reinvested in the destroyed cities to erase the trauma of the war, as acknowledged by Miguel & Roland (2011) for Vietnam. Urban persistence may be due to targeted public investments rather than serial correlation. Secondly, studies that focus 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 (destructions) to economic geography, investments in transport infrastructure (constructions) constitute positive spatial shocks. Yet, changes in transport technology may considerably erode the initial advantage of the locations using the old technology. For example, Bleakley & Lin (2012) show how portage sites have become major cities in the 19th century U.S. as water transport was the dominant form of transport before the railroad was built. Although portage declined post-1880, the portage locations are relatively more developed today. Our experiment is similar, in that the locations along the railroad had the main cities of Kenya at independence. These locations lost their initial advantage in the immediate post-independence period, yet we find that they remain relatively more developed fifty years later. We thus study how a man-made advantage has created an urban equilibrium, and how this equilibrium persisted when the advantage was eroded. Our test may not be as "clean" as Bleakley & Lin (2012), since the decline of railroads started only fifty years ago. Yet, the relative decline of the colonial railroad was a large, unexpected shock. Since no test is perfect, we think that we provide a valuable contribution to the literature, as more evidence is needed on the relationship between history, path dependence and development.

3. RAILROADS AND PATH DEPENDENCE: BACKGROUND AND DATA

We now discuss the historical background and the data we use in our analysis. The Online Data Appendix contains more details on how we construct the data. 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 construction on development, we construct a new data set of 473 locations, the third level administrative units (with a median area of 256 sq km, around 16x16km), 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 finished. Our analysis focuses on the rail network in 1930, as it did not change afterwards. 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 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 non-railroad 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 is likely to 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 transportation infrastructure in the region was on the agenda, to permit military domination and boost trade historically constrained by high transport costs. Uganda was then the most advanced and economically promising country of the protectorate. The British thus wanted to link

Uganda to the coast, and Kenya, which lies in-between, was simply considered as a transit territory. The main railroad line of Kenya, from Mombasa on the coast to Lake Victoria in the west of the country (see Figure 1), was even called the "Uganda Railway". Therefore, railroad construction to Uganda via Kenya provides us with an ideal natural experiment to identify the causal effects of transportation infrastructure on economic change in the short and long run, and path dependence.

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 trypanosomiasis transmitting Tsetse fly, and Kenya lacked navigable waterways. Headloading was the main, and very costly means of transport before the introduction of railways. Only high value goods were carried through the hinterland and slaves that could be walked. The main caravan route from Lake Victoria to the coast did not go through Kenya but through Tanzania in the South, since Zanzibar was the main port of Eastern Africa. The McKinnon-Sclater road, an ox cart track from the Ugandan border to Mombasa, was then started 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 strongly 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 Kenya-Uganda mainline (from Mombasa to Lake Victoria, see Figures 1 and 2) was constructed between 1896-1901. The railroad 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. The line was first built to Nairobi (see Figure 1), 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 way.[4]

Branch lines: The Ugandan railroad established the general urban pattern of Kenya,

[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 small, whereas Lake Nakuru became a National Park, which restricted the use of its water. Secondly, agriculture in Kenya is rain-fed and does not use irrigation. Thirdly, we will show later that the results are robust to controlling for the Euclidean distances to Nairobi and Nakuru.

as shown by Figures 1 and 2. The line produced its own nodes superseding the old caravan ones. Soja (1968) explains that the equal distribution of urban centres at key points along the main route reflects the weak influence of local economic factors in the initial urban growth. The interior nodes increased in size and importance as various branch railroad lines (and feeder roads to the mainline) were constructed (see Figure 1): Thika (1913), Mogadi (1915), Eldoret (1926), Kitale (1926), 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 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. Nevertheless, we will address these endogeneity concerns by comparing the main and branch lines. If the branch lines were endogenous, we would find stronger effects for the branch lines, which we do not.

Trade costs with the rail: Transport costs decreased massively 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 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 boomed without modern transport technology. Coffee could only be produced west of Nairobi. Given the export price in Mombasa, the distance from Nairobi to Mombasa, and trade costs between the two locations, coffee production was not profitable ex-ante. Railroads were thus essential to the economic colonisation of the hinterland.

Colonial roads. Roads were first complementary to the rail as they were feeders to it. The colonial government sought to "protect" the railroad by under-investing in roads and restricting the number of licensed vehicles. Road investments were also strongly affected the economic effects of the Great Depression and WW2 on Great Britain. The existing roads were then 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

[5]Head porterage rates were higher than in other African countries ca. 1910 (11s per ton mile), e.g. Ghana (5s per ton mile), Nigeria and Sierra Leone (2.5s per ton mile). 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).

that provide a good counterfactual for the railroad lines. We can then 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 traversed areas with better locational fundamentals. Indeed, 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 a result. Therefore, 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 (and not a potential upward bias when comparing the railroad locations with the other locations of the country). 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. Later on, we will show that no positive effects on population growth can be found for the placebo lines.

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 that was funded in 1899, there were only four localities with a population in excess of 2,000 in 1901, that were all coastal towns serving as slave ports. 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 given that conflicts were not infrequent between them. The creation of the Protectorate in 1895 established peace over the whole country. Trade costs remained high without modern transport technology, which constrained economic activity.

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 encouraged to generate economic development, justifying the railroad infrastructure and geopolitical imperative of "effective occupation". Moreover, through the introduction of an agricultural export industry, the railroads could be made to be profitable. Land was alienated and offered to European settlers, whose number increased to almost 60,000 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. In the early colonial period, almost 50% of the settlers were farmers. In the late colonial period, the share of farmers decreased to 25% while the share of civil and railway servants increased to 40%. The shares of merchants, professionals and missionaries were stable.

The exports of cash crops was the engine of the Kenya Colony's development. In the early colonial period, coffee, maize and sisal accounted for 40%, 20% and 10%

of exports respectively. Coffee was the premier export to Europe. Despite Kenya's proximity to Ethiopia, thought to be the origin of coffee, the crop was only introduced by missionaries in 1893. It took 30 years before coffee was widely grown by Europeans, making Kenya a large exporter as early as 1930. 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 also expanded. Then, it was not until the Swynnerton Plan of 1954 that the native African farmers were allowed to grow these crops (except maize, an "African" crop). The export of cash crops allowed for the imports of textiles, clothes, manufactured goods, oil, coal, alcohol, tobacco and vehicles. Most of the goods were transported by rail, as roads were not competitive until the late colonial period, 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 railway 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 increased to 42 at independence, as Europeans also established cities that served various functions. First, many of them lived in towns from where they administered 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. Third, part of the "surplus" generated by agricultural exports was also spent on locally produced manufactured goods and services, which further fueled the growth of these cities. A number of Asians (mostly Indians) that came to Kenya to build the Uganda railway also established themselves as merchants and traders in these cities. Lastly, many towns became seats of the colonial administration. While soldiers and policemen only accounted for 1% of European settlement in these areas, the number of bureaucrats increased to respond to the needs of the farmers and merchants.

The areas where Europeans settled to grow cash 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 through a system akin to "corvee labour". Their inbound migration to towns was limited by the "kipande" system which required all workers to obtain a registration certificate from their European employer. The population of Kenya increased fourfold during colonisation, and most of this growth occurred in these areas.

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 2). The number of localities with a population in excess of 2,000 increased to 247 in 2009. Many Africans also migrated to the existing estates and cities of the White Highlands. Europeans were then asked by the new president Kenyatta to choose the Kenyan citizenship. Most of them refused to give up their British citizenship, which caused their massive exodus to other countries (see Figure 1). 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 massive road investments in the immediate post-independence period (see Figure 1). Kenya remained specialised in the export of cash crops, but coffee production has continuously declined since the mid-1980s, as a result of low international prices, while tea, that is mainly grown in the African areas far away from the railway, has become the main export crop of the country, now accounting for 25% of total exports.

The historical analysis above suggests that railroads have permitted the economic exploitation of the hinterland. Initially, except for a few cities along the coast, there was no urban equilibrium. As Europeans settled along the railroad lines, they established farms and cities that thrived until independence, thus contributing to the emergence of an urban equilibrium. African in-migration to these rural and urban areas was then controlled by the Europeans. The econometric analysis in the following section will show that there were indeed causal effects of the railroads on European settlement, commercial agriculture, urbanisation, and African in-migration during the colonial period. We will then use the fact that railroads collapsed, most Kenyan Europeans left, and coffee production declined in the post-independence period, to test whether colonial railroads had long-term effects on economic development, as would imply the persistence of the urban equilibrium.

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 we implement to obtain causal effects. We also provide evidence on the various mechanisms by which railroads spurred population growth, in particular the commercialisation of agriculture.

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):

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

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 (all the lines were built by 1930): 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 almost 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 and mainly inhabited by pastoralists (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 apples with apples. We will show that our results hold when using the full sample. We also drop the three main nodes of the railroad: Mombasa, Nairobi and Kisumu. The effect then comes from being connected in between those nodes.

We express all population numbers in standard (z-)scores. Standardising the population variables has two advantages. First, the non-standardised values are much higher for some variables (e.g., total population vs. European population), so the coefficients β cannot be readily compared across outcomes. Standardised values eases the interpretation of the coefficients. Second, total population has been growing over time, so the coefficients β will mechanically increase for later periods, unless we standardise the variables. Logs could perform the same role. However, there are many locations with an European or urban population equal to 0. Using logs would drop these observations. Therefore, we will use standard scores in the main analysis, and show that results hold when using logs. Moreover, we add the location's area (sq km) on the right hand side as one of the controls. This let us conveniently interpret effects as densities. Infrastructure investments are typically endogenous: places with highest return will be connected. We argued in section 3.2 that placement of railroads was not endogenous to economic potential and population. We now describe the tests we perform 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 do a good job at controlling for 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 to include them in the regressions.

We add physical geography variables around 1901 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 their population was about 5,000 inhabitants) to control for historical urban patterns, and (ii) the area shares (%) of "major settled groups" and "pastoralist groups", and a "isolated groups" dummy to control for historical population patterns.

Firstly, 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 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 railroad, 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 eighteen locations of the 4,000 m high Mounts Kenya, Elgon and Kinangop. Treated locations are not that different when excluding these few outliers. Treated locations were then historically less populated than control locations (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, or an upward bias, if the railroad locations grew faster because they were initially lowly-populated.

Secondly, since we have data for 16x16 km locations, neighbouring locations may not differ in terms of unobservables. The location's median area is 256 sq km, about Boston's area. When comparing the 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 then 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 truly 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 ethnic 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 will control for heterogeneity at a fine spatial level, and identification will then come from variations in the treatment between very similar locations.[6]

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.

Fourthly, we can compare the connected locations with locations that would have been connected if the placebo lines had been "counterfactually" built. As discussed

[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. 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.

above, we should 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 more elevated, 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.

While we do not push one identification strategy in particular, and while none of these strategies is perfect, we will show that all of them give relatively similar results. This will comfort us in thinking 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 European and urban population growth, but these effect decrease as we move away from the railroad and are zero after 30 km and 10 km respectively (column (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 have then attracted even 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 capture 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 neighboring locations). We also find a strong effect of the railroad on Asian population, which disappears when adding the number of Europeans (columns (8)-(9)). The Asians only settled in the (European) cities, where they became merchants and traders. The coefficient of correlation between the Asian and urban population was even 0.90.

Table 3 investigates further the effects of the railroad on European settlement, since the other outcomes were 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 Kenya Colony. Column (1) of Table 3 replicates the main results on European settlement from column (1) of Table 1. In columns (2)-(3), we find that these railroads effects were 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 most of the settlers were farmers. The railroad also had positive effects on the number of European non-farmers (not shown). In column (5), we show that most of European settlers in 1933 lived in locations where there were already settlers 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 Europeans (columns (7), (9) and (11)). Sisal was not produced in the non-arid areas. We do not find any effect of the railroad on tea cultivation (not shown), as the areas suitable for tea were far from the railroad. Besides, tea did not become a major export before the 1940s.[7] The commercialization of agriculture, whether as a result of reduced trade costs or because railroads encouraged the diffusion of crop production techniques, was thus one the mechanisms by which the railroad contributed to economic development in these areas. One of the other mechanisms could have been the establishment of the "imperial peace" over these areas. However, the whole territory was pacified as soon as the Protectorate was created in 1895. 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 various identification strategies. Column (1) replicates our main results from Table 2 (see columns (1), (2) and (4)). For the sake of simplicity, we only 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.

Spatial Discontinuities. The fact that locations are small imply that neighbouring locations may not differ in terms of unobservables. Including 21 ethnic group fixed effects (35 district fixed effects) allow us to control for spatial heterogeneity across ethnic groups (districts). The railroad effects are then identified by comparing treated locations with neighboring 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

[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 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.

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 locations 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 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.

Main robustness checks. The results suggest that the railroad had strong causal ef-

[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 when they are built. Thus, even if their placement was endogenous, we use the exogenous timing of their construction to identify causal effects.

[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.

fects 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, and only focus on the non-arid areas west of Nairobi (column (1)), (ii) use the full sample of locations including the ones in the sparsely populated, arid parts of Kenya (column (2)), (iii) also control for the Euclidean distances to each railroad node (Mombasa, Nairobi and Kisumu) 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" at independence (column (4)): the identification then comes from the comparison of treated and untreated locations *within* and *outside* these areas),[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 square (column (8)), or a dummy equal to one if the location was 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. We also verify that the results hold when constructing the railroad dummies using the Euclidean distance to a railroad station in the 1930s (not shown). This is not surprising since the coefficient of correlation between the Euclidean distances to a railroad station and a railroad line is 0.99. Secondly, we test that the results are robust to using other population thresholds to define a locality as a city. We use 2,000 inhabitants because it is the lowest threshold for which we have consistent data post-1962. We obtain similar results if we use 5,000 instead, a threshold that is often used in the literature (Bairoch, 1988; Acemoglu, Johnson & Robinson, 2002): 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 not as significant as when standardised urban population is the dependent variable, 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). Thirdly, for the European and urban population, we can also follow a difference-in-difference (DiD) strategy where we compare connected and non-connected locations over time. The European population of the 403 non-arid locations was nil in 1901. Three locations had then an urban population of about 5,000. The DiD regressions allow us to include location and year fixed effects. We drop the controls since they are time-invariant. The results are even stronger now. For example, when the depen-

[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.

dent 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-sectional analysis (column (2) of Table 1). We cannot follow this strategy for total population, since there was no census in 1901. Fourthly, Conley standard errors do not account for spatial autocorrelation among the group of railroad locations, which are connected by the mere fact that they belong to a same network. We thus cluster standard errors using nine groups: one group for locations belonging to the main line, seven groups for locations belonging to each one of the seven branch lines, and one group for untreated locations. The results hold when doing so (not shown). Lastly, we cannot estimate the railroad effects for the arid locations only (see Figures 1 and 2), given the small number of observations (N = 67). Moreover, a few of the arid locations along the railroad lines were game parks in which settlement was restricted.

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 European settlers and created cities. We now examine other outcomes at independence. This is important, because it clarifies what other (colonial) factors the railroads brought, which could then cause path dependence. We use the same model as model (2) to examine whether railroad locations had better (non-railroad) 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. Since the Kenya Colony was extremely poor in the 19th century, no location had any of these types of infrastructure then. 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 then 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 the log because we expect a non-linear relationship between the number of people and the availability of public goods. Interestingly, the fact that the effects remain high and significant in Panel B but not in Panel C indicates 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 (mostly for Europeans, since African in-migration was controlled) were created as a result. Another interpretation of these results could be that the coloniser has invested in other public goods at the same time as it was building the railroad. In other words, a location could be connected to the railroad and benefit from public investments in educational infrastructure the same year. In that case, the railroad effects on population could be explained by these other colonial investments. However, there were almost no other investments in 1901. There were only 500 settlers then, and they all lived in Mombasa or Nairobi, which are not included in our analysis. A second interpretation of these results could be that railroads directly influenced the placement of other public goods between 1901 and 1962. Population then increased as a result of the presence of these factors. Even if we believe that the coloniser would not have invested in European public goods where there were no Europeans, we cannot rule out this possibility. As such, the railroad dummies could capture the effects of a package of colonial investments.[11]

General equilibrium effects. While settlers migrated from Great Britain, African population growth was due to in-migration from the African areas (see Figure 1). 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, which forced many to leave subsistence farming to work in the European areas. Africans worked as indentured farmers for the Europeans, while African wages were fixed at low levels in the cities. Europeans were thus able to capture most of the surplus 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 economic development. Railroad locations are more developed today, although they have lost their initial advantage in terms of transportation, European settlement and commercial agriculture. The persistence of the first urban equilibrium provides strong evidence for path dependence. We use our novel data set to study the channels of path dependence. We also discuss urban optimality.

[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, hence the integration of both transportation networks. However, the roads do not explain why the railroad locations are more developed, as discussed in subsection 4.4 (and shown in column (6) of Table 5).

5.1 Relative Decline of Railroads and Exodus of Europeans

The rail network reached its maximum length in 1930, with 2,084 km of track, and has stagnated since. As shown in Figure 3, freight volumes were more or less increasing until independence, with the noticeable exceptions of the World War I, Great Depression and World War II periods. Railroad traffic has then continuously declined since independence in 1963 (more exactly 1956 or 1973 depending on the measure of railroad traffic used). Traffic volumes are now much lower than what they used to be, although railroad 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) describes 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 in Kenya 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, in spite of its potential. Second, the first governments of Kenya massively invested in the road network (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 length of paved and improved roads increased from about 4,000 km in 1964 to 13,000 km in 2002. Rent-seeking favoured large construction projects prone to embezzlement, such as the construction of new roads. Then, road building in Kenya was driven by political considerations instead of economic considerations, as the various presidents of the country disproportionately targeted their ethnic homelands in terms of road investments. Maintaining the old (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/or were replaced by roads at nearby sites. However, they clearly lost their comparative advantage in terms of transportation. For the restricted sample of 403 non-arid locations, 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/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/improved road. However, even these 69 locations have access to an earthen road. Kenya not being as tropical as most countries of Western and Central Africa, most earthen roads are suitable for motor traffic throughout the year, and not just during the dry season. Therefore, while few locations of the hinterland were connected to the coast in the early colonial period

[12] This instability includes the Mau Mau Rebellion in the 1950s, the difficult transition to independence and democracy in 1963, the subsequent exodus of the European and Asian populations, the abolition of democracy in 1969, the death of Kenyatta and his replacement by his (not less dictatorial) vice-president Arap Moi in 1979, and the economic crisis in the 1980s and 1990s.

(the railroad age), almost all locations are now connected, which has considerably eroded the initial advantage of the railroad locations. Yet we will show that these locations are still relatively more developed today, although an entirely different spatial distribution could have emerged after independence 50 years ago.

The number of Europeans grew exponentially from 500 in 1901 to almost 60,000 in 1962, on the eve of independence. As described in section 3.3, Europeans were then asked by the new president Kenyatta to choose the Kenyan citizenship. Most of them refused to give up their British citizenship, which was not well-received by the native population and 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 1). The European settlers who accepted to acquire Kenyan citizenship then settled in Nairobi and Mombasa to specialise in service activities. The African farmers that replaced them remained specialised in coffee and tea, while maize stopped being exported. Yet coffee production has dramatically declined since the mid-1980s, as a result of low international prices and exhausted coffee soils in the old coffee-producing regions. Coffee only account for 10% of total exports now, while tea has become the main export of the country, now accounting of 25% of total exports. However, tea is not produced in the same areas as coffee. 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 European settlers represented the business elite during the colonial period, these areas must have experienced a transitory (if not permanent) 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? In particular, has the level effect on population density in 1962 narrowed over time? 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 first urban equilibrium. Besides, we do not have total population data for the most recent census year.

Cross-sectional analysis. We run the same regression for 403 non-arid locations l as model (1), except the dependent variable is now defined for the year 2009:

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

with $zUPop_{l,2009}$ being the standard score of urban population in 2009. We standardise the urban variable to account for demographic growth post-1962. Kenya's population has increased threefold in 50 years, and the average size of its cities has also dramatically risen. Standardised values will ease the comparison of the coefficients across periods. 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. Interestingly, the point estimates are not very different from what we found for the year 1962 (column (2) of Table 2). This shows the remarkable stability of the urban spatial distribution between 1962 and 2009, and suggests that the level effect in 1962 may have remained the same. 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. Likewise, controlling for roads in 1962 did not strongly modify the relationship between colonial roads and urban density in 1962, as discussed in section 4.4. Being close to a paved road has a positive effect on urban population, yet this effect is twice lower than for the rail. In column (3), we verify that these long-term effects are indeed causal by comparing the railroad locations with the placebo locations, as we did before. The results are unchanged. 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 centroid, or comparing the branch and main lines (not shown but available upon request). The 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 cities that thrived during colonisation. The long-run effects are then 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 primarily because they were more urbanized in the past.[13]

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 $t = [1962, 1969, 1979, 1989, 1999, 2009]$:

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

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 the various controls listed in Table 1. For the year 1979, the previous period is 1962. There is convergence (resp. divergence) in the period $[t\text{-}1;t]$ if κ_t is lower (resp. 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

[13] 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%.

there is a positive and significant effect (0.08**). The population effects are also small and not significant, with the exception of the year 1979 (vs. 1962). We first verify that the post-independence effects are not smaller than for the colonial period (1962 vs. 1901) for the mere reason that we are studying periods of 10-20 years instead of 60 years. For example, we run the same regression as model (3) for the year 2009, except we control for urban population in 1962 instead of 1999. This is basically what we did when studying the cross-section in 2009. Column (5) of Table 6 showed that the effect was also small, and not significant (0.08). Secondly, the effects are still positive until 1979. There are various reasons why we think it could be the case. Migration restrictions to the European cities and estates were lifted at independence. Many Africans migrated to these areas as a result. The fact that there was not significant impact for urban population between 1962 and 1969 could be due to the disruptive effects of the exodus of Europeans. Then, the various ethnic groups around the Rift Valley kept colonising the land around the railroad lines after independence. After 1979, roads had been built or upgraded throughout the country, railroad traffic was clearly declining, the Europeans had definitely left, and land expansion was decelerating as coffee production was falling. Yet the railroad locations are still relatively more developed in 2009, thirty years later.

5.3 Channels of Path Dependence

The railroad effects on urban population were unchanged between 1962 and 2009, although the railroad locations lost their initial advantage in transportation, European settlement and commercial agriculture. What could explain path dependence? As already explained in section 2.3, colonial *sunk investments* could induce people to stay at these locations. If schools and hospitals are expensive to build, people are less mobile and initial advantages have long-run effects. The long-term effects of historical factors will depend on how fast sunk capital depreciates. The loss of comparative advantage dates back from 40 to 50 years ago, 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 *coordination problem* as it is not obvious which locations should have the contemporary factors. In this case, it makes sense to locate factors in locations that are already developed, for example the railroad locations. The location of contemporary factors (including people) today then depends on past population density, without it being explained by historical factors. We study how the railroad effects on urban population today vary as we control for the various channels of path dependence, i.e. the historical factors and contemporary factors.

Historical factors. In section 4.5, we showed that connected locations had better non-railroad infrastructure per capita at independence. In particular, the railroad locations had better access to health, educational, institutional, communication, transportation and industrial infrastructure. If these historical factors had an independent long-term 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 type of infrastructure in the location, as we want to control as much as possible for sunk investments.[14] 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). However, the Kenyan context was very specific, in that 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*, but not as much *per European 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). The specific persistence of the railroad cities is thus not explained by the fact that the constraints on African migration were lifted at independence. To conclude, 25% of urban persistence post-1962 can be explained by sunk investments. These results are in line with the Table III of Bleakley & Lin (2012), where the inclusion of historical factors appears to reduce the long-term effects of portage in the U.S. by similar magnitudes. While historical factors matter to explain urban patterns today, they are not the main channel of path dependence. It is then interesting to note that the coefficients of the rail dummies are negative in column (6). In other words, once we control for the positive long-term effects of urban population and factors in 1962, the railroad locations are "losing" relative to the rest of the country. This is a clear illustration of path dependence. Without increasing returns (in people and other factors - the P in our conceptual framework), the railroad locations would have declined when losing their initial advantage.

Contemporary factors. We first verify that railroad locations have higher densities of contemporary 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.[15] 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 transportation 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 old railroad lines were paved over time. The point estimates are strongly reduced when controlling for population today (Panel B). In other words, 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 6, we find that controlling for contemporary factors does not modify the relationship between urban population in 1962 and urban population in 2009 (the coefficient remains almost unchanged, at 0.54***). These results point to the following story: railroad locations have higher densities today, because people co-locate where there are more people in the previous period, and other contempo-

[14]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.

[15]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.

rary factors "follow" people (instead of people following factors). There were then more people in the previous period because of the initial population effect during the colonial period and repeated co-location decisions afterwards.

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 as a result of ethnic favoritism. 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 already control for some of these effects. However, the fixed effects imperfectly capture the fact that there may be heterogeneous effects across ethnic groups. We run the same regression as in column (7) 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 played a role in protecting 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. For example, Jedwab & Moradi (2013) explain that 80% of African railroad lines were built during colonization to match the needs of the colonial powers. Twenty-five African countries in particular have received significant investments in the rail, and these countries were more urbanized as a result. Gwilliam (2011) then describes how railroad traffic has collapsed in these countries, with the exception of South Africa. Yet Jedwab & Moradi (2013) show that the locations along the colonial railroad lines are still relatively more urban and economically developed today.

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

5.4 Economic Development Today

We have shown that railroad locations have higher densities of contemporary factors today. However, these positive effects are somewhat reduced once we control for population densities. In per capita terms, railroad locations are not necessarily better endowed in factors than non-railroad locations of similar densities, whether we consider health or educational infrastructure. However, these locations could still be wealthier per capita than the other locations, 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 economic development as the dependent variable. This allows us to compare the railroad and non-railroad locations. We can then simultaneously control for contemporary urban population. This then allows us to only compare 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), for example as a result of tea cultivation. Therefore, the European cities and African cities have two different origins and geography, that are somewhat related but at different periods. 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), 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 (2)). 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, in line with 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, in line with their paper. 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 exactly 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)).

The evidence suggests that the railroad locations and cities are indeed wealthier today. This cannot be explained by the observable factors in our analysis, and has to be explained by unobservable factors. There must be unobservable contemporary factors that were repeatedly co-located along the railroads, such as entrepreneurial, social or institutional capital. These results suggest that colonial railroads 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 An Optimal Urban Equilibrium?

As explained in the conceptual framework, an urban equilibrium arising from a historical shock is suboptimal if the resulting urban distribution does not match the distribution of the locational fundamentals. Without the shock, the urban equilibrium would have eventually emerged in the locations with the best fundamentals. The historical shock has permitted an earlier emergence of the equilibrium. However, the locations that received the cities thanks to the shock may be characterized by worse locational fundamentals. For example, they could be situated in rugged areas, as in Nunn & Puga (2012), or they could be farther from a navigable waterway, as in Michaels & Rauch (2013). A social planner would then be able to increase overall welfare by spatially reallocating people to the fundamentally best locations. In the long run, without increasing returns, the efficient urban equilibrium would naturally evolve into the optimal equilibrium, as people would relocate to the best locations. With increasing returns, the suboptimal urban equilibrium will persist after the shock. Increasing returns are due to high relocation costs, whether as a result of sunk investments (the fact that non-labor factors are not costlessly mobile) or the coordination problem of contemporary factors (the fact that factors need to simultaneously relocate for this to work). The suboptimal urban equilibrium is thus likely to persist. Our natural experiment fits into this framework, even if the construction and decline of the colonial railroad (along with the mass inflow and outflow of Europeans, and the rise and fall of coffee production) is a spatial shock that lasted several decades. While sunk investments account for some of this persistence, the coordination problem remains the main constraint to any relocation.

The question is which urban equilibrium would have emerged in Kenya without the historical shock, i.e. without the construction of the colonial railroad? Such an analysis is necessarily speculative. There were very few cities before colonization. Had the railroad not been built, when and where would have locational fundamentals caused the emergence of another equilibrium? While we cannot answer the "when" question, it is likely that cities would have emerged in the areas that were historically densely populated by Africans before the colonial era. These populations must have been living in these areas, because of the better locational fundamentals they provided. Without colonization, economic development would have limited to these areas, and population would have eventually resulted in urban growth. We thus decide to compare the railroad locations with the historically settled locations. We only consider the railroad locations less than 10 km from a railroad, as there was only an urban effect until there. The historically settled locations are then the locations in which the area share of the "major settled groups" is more than 50%. Indeed, cities would not have emerged in the pastoralist areas. We regress various outcomes on the 0-10 km rail dummy to test if the railroad and historically settled areas significantly differ. We do not include province fixed effects or controls, as the exercise is simply a comparative one. Table 10 displays the results. In Panel A, we compare all the railroad and historically settled locations. In Panel B, we restrict the comparison to the locations with a city in 2009. In other words, are the cities that emerged as a result of the historical shock different from the cities that emerged later on? In Panel C, we also control for the urban population in 2009, so as to compare cities of similar sizes. In columns (1) and (2) of Panel A, we first verify that the railroad locations are indeed more urban and more economically developed, when using as a dependent variable the standard score of urban population in 2009 and the poverty headcount ratio (%) in 1999. In columns (3)-(11), the dependent variables are the geographical controls we used in our econometric analysis (see Table 1). These controls proxy for locational fundamentals. Results are similar across the panels. The railroad locations/cities are less likely to be situated along the coast (column (3)), have lower rainfall (column (5)), have higher altitude (column (6)), and their soils are less suitable for the cultivation of tea (column (11)), the main export of Kenya today. They are also less arid, but the coefficient is extremely small (-0.1%) and barely significant. Without increasing returns, the bad geography that characterizes these locations would have had a negative impact, as it raises (non-railroad) transport costs and reduces economic potential.

We test the robustness of these results. Firstly, the railroad led to the emergence of an urban equilibrium in the hinterland, since this was the region most constrained by high trade costs before the rail. If historically settled areas were closer to the coast, our results could be driven by the fact that we are comparing the coast and the hinterland. We thus verify that we obtain rather similar results when dropping the coastal province, so as to focus the analysis on the hinterland only, or even including province fixed effects, so as to compare railroad and historically settled locations within the same province (not shown, but available upon request). Secondly, if the colonial railroad created cities in fundamentally worse locations, it may not be an issue if it does not also prevent the emergence of cities in the fundamentally best locations. The equilibrium is only suboptimal if it leads to urban eviction effects. If we find that cities also emerged in the historically settled locations (vs. the other control locations), it means that the urban effects of the colonial railroad are welfare-improving, as even more cities are then created. We regress the standard score of the urban population in 2009 on the "historically settled location" dummy, while dropping the locations less than 10 km from the rail. This allows us to compare the historically settled locations and the other control locations. The effect is nil and not significant (not shown). The historically dense areas do not have more cities than other areas, which points to urban eviction effects, and therefore urban suboptimality. Lastly, we check that the results hold when using the railroad locations until 20 km, or using 75% as the threshold for the area share of the "major settled groups" when defining the historically settled locations (not shown).

6. DISCUSSION: PATH DEPENDENCE AND DEVELOPMENT

[To Be Completed]

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."

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, 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.

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 Joshua D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature*, 47(4): 983–1028.

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, 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.

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.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on The Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*. Vol. 4 of *Handbook of Regional and Urban Economics*, , ed. J. V. Henderson and J. F. Thisse, Chapter 49, 2119–2171. Elsevier.

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

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

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

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 Control Locations**		**(2) Locations 20-40 km**		**(3) Placebo Locations**	
Dependent Variable:	Coeff.	SE	Coeff.	SE	Coeff.	SE
"Coastal location" dummy	-0.03	(0.03)	0.01	(0.02)	0.00	(0.03)
Distance to the coast (km)	-8.01	(6.50)	-8.00	(7.55)	-7.85	(7.91)
Average annual rainfall (mm)	-58.78*	(31.08)	-135.89***	(36.53)	-27.61	(37.52)
Altitude: mean (m)	19.36	(39.54)	-30.02	(46.57)	110.77***	(41.43)
Altitude: standard deviation (m)	-36.22***	(11.15)	-30.13**	(13.45)	5.83	(12.03)
Share of arid soils (%)	-0.08**	(3.47)	-0.01	(0.01)	-0.10	(0.10)
Share soils suitable for agriculture (%)	-0.39	(3.65)	-3.25	(4.00)	-5.99	(5.15)
Share soils suitable for coffee (%)	-0.68	(2.75)	0.13	(3.59)	-5.54*	(3.30)
Share soils suitable for tea (%)	-11.87***	(2.99)	-15.37***	(4.01)	-10.12**	(4.07)
Area (sq km)	-84.64	(52.64)	-58.31	(60.10)	95.81**	(48.10)
"City in 1901" dummy	-0.01*	(0.01)	0.00	(0.00)	-0.02	(0.02)
"Provincial capital" dummy	0.00	(0.01)	-0.00	(0.02)	-0.01	(0.02)
Area share of "major settled groups" (%)	-10.72**	(4.23)	-15.75***	(5.35)	-21.43***	(5.38)
Area share of "pastoralists" (%))	-0.96	(3.53)	4.24	(4.03)	4.91	(4.36)
"Isolated groups" dummy	0.02	(0.05)	0.06	(0.06)	0.13**	(0.06)
Number of treated locations:	162		162		162	
Number of control locations:	241		89		81	

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 (colonial) railroad line, while simultaneously including province fixed effects (N = 8). Robust standard errors: * p<0.10, ** p<0.05, *** p<0.01. There are 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). 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*	*Placebo Treatment Effects: Drop if Rail 0-X km = 1*	*Control Group: Placebo Locations*	*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***	0.37***	0.23**	0.41***	0.35***	1.36***			0.41***
	(0.09)	(0.11)	(0.10)	(0.11)	(0.11)	(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)	(0.24)	(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***	0.28***	0.26***	0.16*	0.24**	0.80***			0.31***
	(0.10)	(0.07)	(0.09)	(0.10)	(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. 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

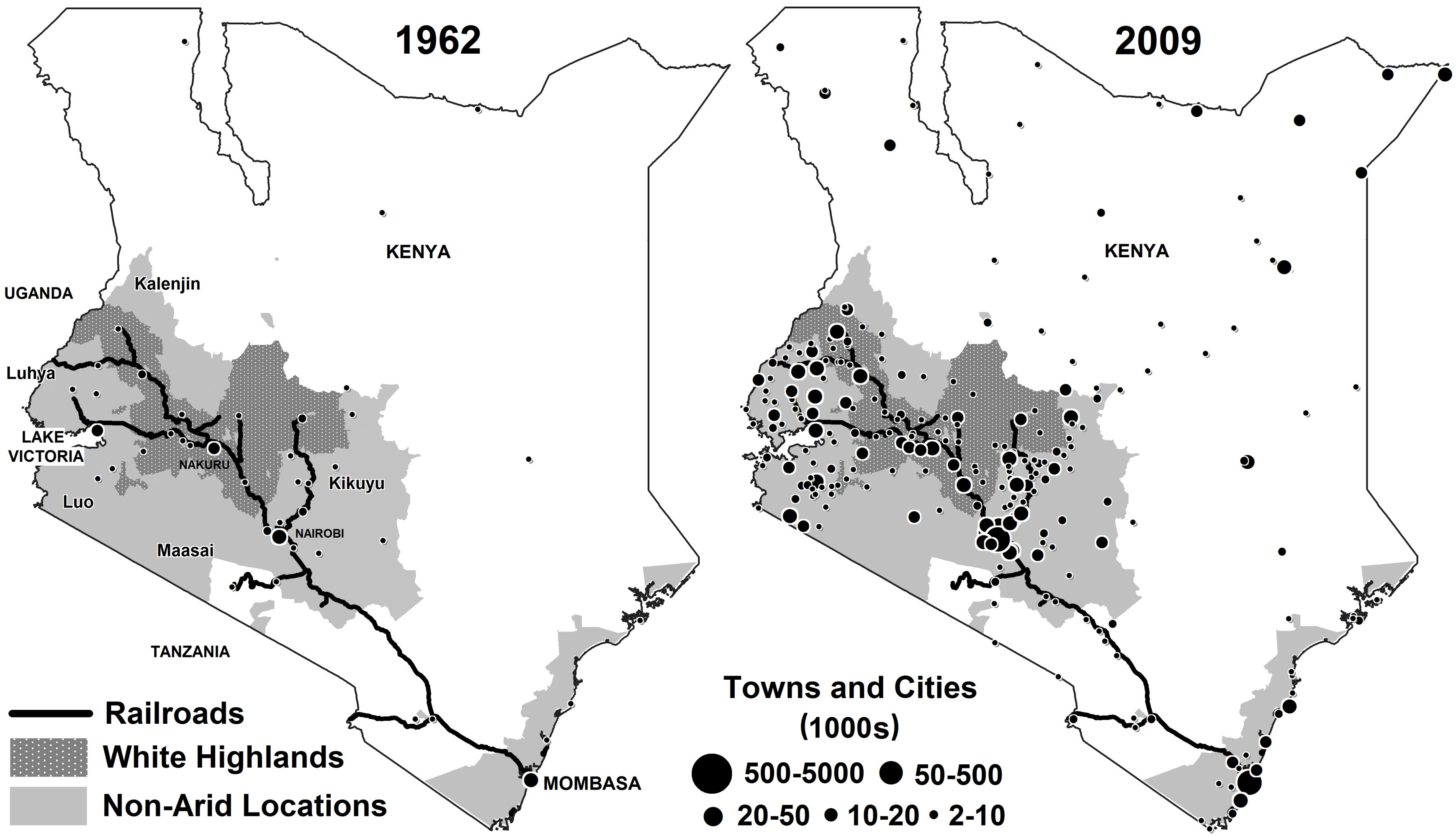
Notes: OLS regressions using data on 403 non-arid locations for the year 2000. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In columns (1)-(2), the dependent variable is the poverty headcount ratio in 1999, i.e. the percentage of the population of location l living below the 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.

TABLE 10: COLONIAL RAILROADS AND URBAN SUBOPTIMALITY TODAY

	Urban Pop. (Z)	Poverty Headcount Ratio (%)	Coastal Location Dummy	Distance to Coast (km)	Av. Annual Rainfall (mm)	Altitude Mean (m)	Altitude Std. Dev. (m)	Share Arid Soils (%)	Share Agri. Soils (%)	Share Coffee Soils (%)	Share Tea Soils (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: ***Railroad vs. historically settled locations (N=272)***											
Rail 0-10 km	0.63***	-13.5***	-0.1**	40.4	-112***	440***	-6	-0.1*	-0.5	-0.8	-14.6***
	(0.22)	(2.2)	(0.0)	(27.8)	(39)	(84)	(10)	(0.0)	(4.3)	(4.4)	(2.8)
Panel B: ***Railroad vs. historically settled locations, for urban locations only (N=101)***											
Rail 0-10 km	1.39***	-12.3***	-0.1**	16.0	-147**	410***	-7	0.0	-10.1	-3.6	-14.1***
	(0.44)	(3.3)	(0.1)	(45.0)	(64)	(125)	(18)	(0.0)	(6.7)	(6.5)	(5.2)
Panel C: ***Railroad vs. historically settled locations, for urban locations only (N=101), control for urban pop.***											
Rail 0-10 km		-11.0***	-0.1**	19.6	-104	373**	5	0.0	-8.1	-3.6	-13.5**
		(3.6)	(0.1)	(50.9)	(69)	(145)	(19)	(0.0)	(6.8)	(7.5)	(5.9)

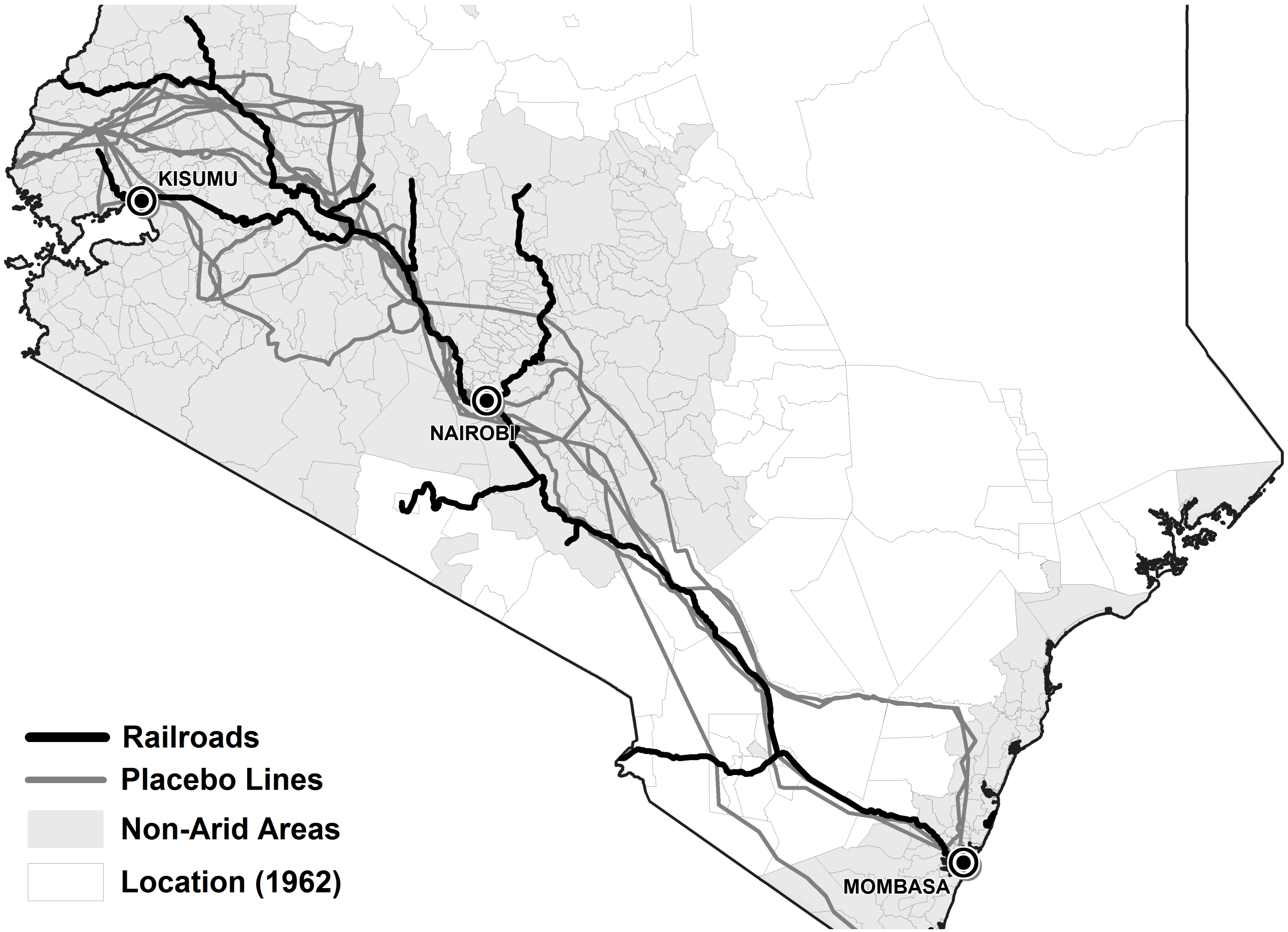
Notes: OLS regressions using data on 403 non-arid locations for the year 2009. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01. This table tests whether the locations just along the railroad and the historically settled locations before the colonial area are significantly different in terms of locational fundamentals. In Panel A, we simply compare the railroad and historically settled locations. In Panel B, we restrict the comparison to the locations with a city in 2009. In Panel C, we also control for the urban population in 2009. In columns (1) and (2), the dependent variable is the standard score of the urban population in 2009 and the poverty headcount ratio (%) in 1999 respectively. In columns (3)-(11), the dependent variables are the geographical controls we used in our analysis. These are described in Table 1, and in the Online Data Appendix. 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



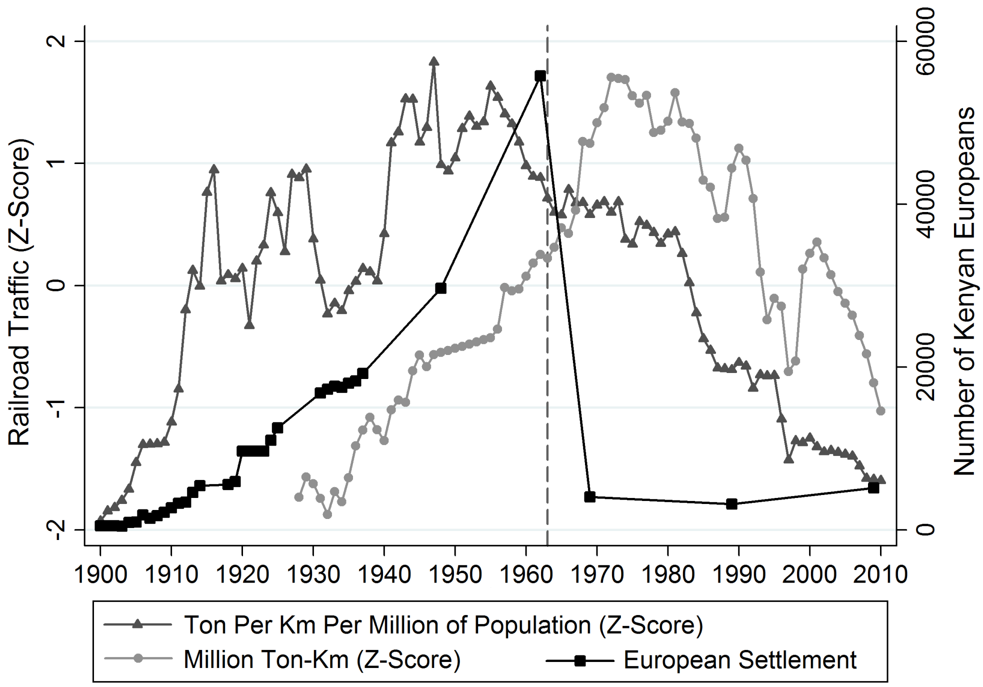
Notes: The map shows colonial railroads and cities in Kenya in 1962 and 2009. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria in the west was built in 1896-1903 to link Uganda to the coast. The non-arid areas 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



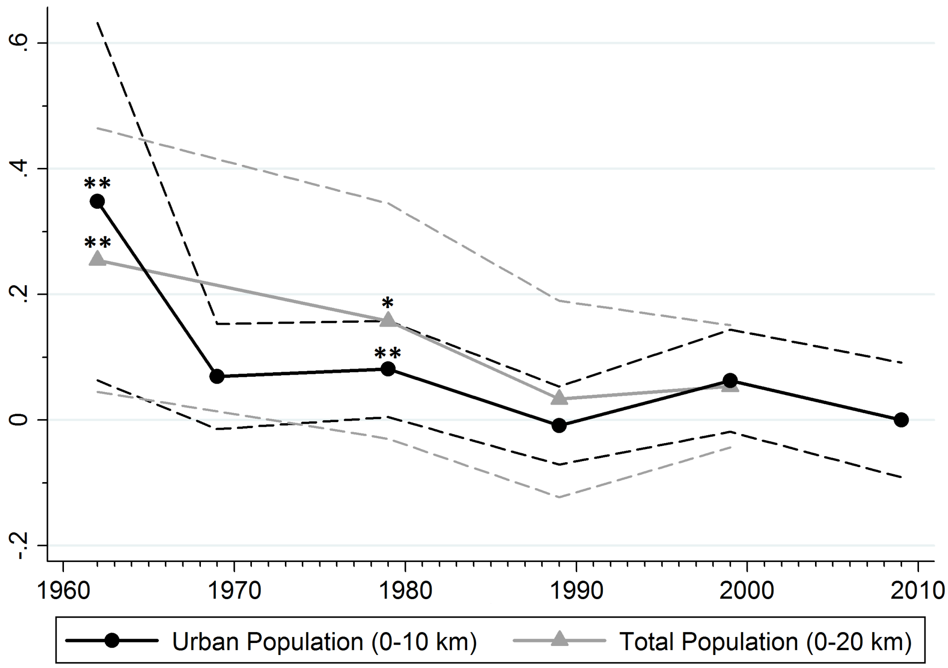
Notes: The map shows the railroad lines and the placebo lines in Kenya. Kenya's railroad lines were all built before the country gained independence in 1963. The main line from Mombasa on the coast to Lake Victoria (Kisumu) was built in 1896-1903 to link Uganda to the coast. The branch lines were built between 1913 and 1930. The map also shows seven placebo lines: the Thomson route (1883), the Pringle route (1883), the Fischer route (1885), the Lugard route (1889), 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 Mochakos (1926). Thomson, Pringle, Fischer, Lugard, McKinnon-Sclater, Austin-Pringle 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 standardized 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 kilometer 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 kilometers traveled. Data is interpolated for missing years. The vertical dashed line indicates the year 1963, when Kenyan gained independence. Most Kenyan Europeans left the country after 1963, as they had to acquire Kenyan citizenship to stay. 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 sublocations – the fourth level administrative units – in 1999 that is available on the website of the International Livestork Research Institute (http://www.ilri.org/GIS) to reconstruct the GIS map of the locations in 1962. Sublocations 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 *Survey of Kenya* (1959). A map of the branch lines is obtained from XXX. 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 XXX. 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 various lists of "urban localities" available in the paper and online reports of the *Population and Housing Censuses* 1962, 1969, 1979, 1989, 1999 and 2009. These reports consistently lists all localities above 2,000 and their population size. Defining as a city any locality with more than 2,000 inhabitants, we obtain a geospatialized sample of 247 different 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 then compare these estimates with what we find from 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 census. Total population for posterior census years – 1979, 1989 and 1999 – is obtained from various GIS databases available on the website of the International Livestork Research Institute (http://www.ilri.org/GIS). These GIS databases report the total population of each sublocation 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 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 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.

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 Livestork 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), improved (laterite), and earthen 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 *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 digitized 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 *Survey of Kenya* 1959. We digitized 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 centers and towns in 1962. We digitized the map using GIS and create an "industrial center" dummy if the location contained at least one of these industrial centers 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 health dispensaries) for each location. Data on health 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 facilities (primary schools and secondary schools) for each location.

Economic Development for Kenya:
We use geospatialized poverty maps available at the sublocation level - the spatial 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 Livestork Research Institute (http://www.ilri.org/GIS). We use two different 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.

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.

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.

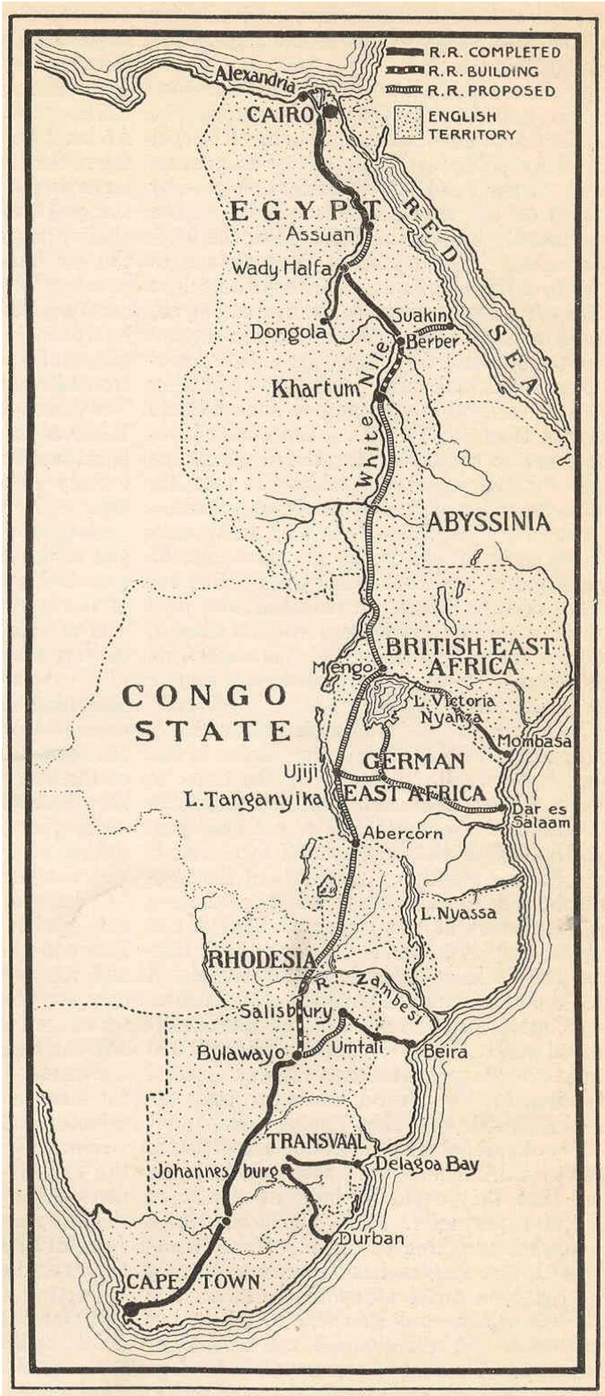
Soja, Edward W. 1968. *The Geography of Modernization in Kenya: A Spatial Analysis of Social, Economic, and Political Change.* Syracuse, NY: Syracuse University Press.

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

RAILROAD AND PLACEBO LINES [To be completed]

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	Pringle 1883	Fischer 1885	Lugard 1889	Jackson 1889	McKinnon Sclater 1897	Austin Pringle 1899	Macdonald 1899	Eldoret Sergoit 1926	Mochakos 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**	0.23	0.5	0.50*	0.44	0.05	-0.15	0.25	1.47	0.25	0.1	0.84
	-0.21	-0.14	-0.24	-0.24	-0.39	-0.26	-0.3	-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	-0.05	0.98	0.38	0.63	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)	(0.12)	(0.75)	(0.25)	(0.54)	(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*	0.16	0.58**	0.17	0.41	0.00	0.39	0.37	0.26	0.04	0.54	0.53*
	-0.14	-0.08	-0.23	-0.13	-0.29	-0.12	-0.28	(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	0.07	0.00	0.08	-0.11	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)	(0.09)	(0.08)	(0.11)	(0.10)	(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**	0.23	-0.02	0.06	-0.03	0.20	0.14	0.19	-0.27*	1.24	-0.19*	-0.03
	-0.11	-0.07	-0.1	-0.19	-0.06	-0.06	-0.1	(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	0.28	0.05	0.16	0.25	-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)	(0.18)	(0.22)	(0.15)	(0.32)	(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.