CENTRE FOR ECONOMIC HISTORY
THE AUSTRALIAN NATIONAL UNIVERSITY
DISCUSSION PAPER SERIES

THE PERMANENT EFFECTS OF TRANSPORTATION REVOLUTIONS IN POOR COUNTRIES: EVIDENCE FROM AFRICA

REMI JEDWAB
GEORGE WASHINGTON UNIVERSITY

ALEXANDER MORADI
UNIVERSITY OF SUSSEX

DISCUSSION PAPER NO. 2015-01

JANUARY 2015

THE AUSTRALIAN NATIONAL UNIVERSITY
ACTON ACT 0200 AUSTRALIA
T 61 2 6125 3590
F 61 2 6125 5124
E enquiries.eco@anu.edu.au
http://rse.anu.edu.au/CEH

The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa*

Remi Jedwab† and Alexander Moradi‡

January 2015

Abstract: We exploit the construction and eventual demise of the colonial railroads in Africa to study the impact of transportation investments in poor countries. Using Ghana and Sub-Saharan Africa as a whole, we assembled new data on railroads and cities spanning over one century to show that: (i) Railroads had large effects on the spatial distribution and aggregate level of economic activity during the colonial period, as they constituted a transportation revolution in a context where no modern transportation technology previously existed. (ii) These effects have persisted to date, although railroads collapsed and road networks expanded considerably in the post-independence period. The analysis contributes to our understanding of the heterogeneous impact of transportation investments. It shows that initial investments may have a large effect in poor countries with basic infrastructure. As the countries develop, increasing returns may then solidify their spatial distribution, and subsequent investments may have a smaller effect on local economic development.

Keywords: Transportation; Railroads; Development; Cities; Path Dependence; Roads

JEL classification: O1; O3; O18; R4; R1; N97

*We thank the editor, three anonymous referees, Hoyt Bleakley, Robin Burgess, Denis Cogneau, Melissa Dell, Taryn Dinkelman, Dave Donaldson, Marcel Fafchamps, James Fenske, Edward Kerby, Richard Hornbeck, Elise Huillery, Stelios Michalopoulos, Guy Michaels, Edward Miguel, Nathan Nunn, Henry Overman, Jean-Philippe Platteau, James Robinson, Elizabeth Sadoulet, Adam Storeygard, Leonard Wantchekon and seminar audiences at Barcelona, Berkeley, Bocconi (CEPR/AMID), Cambridge (RES), Carlos Tercero, CIDE-Princeton (EPEN), Dartmouth (NEUDC), DIAL, Geneva (AEHW), Georgetown, George Mason, GMU-GWU Workshop, George Washington, Harvard, Houston, IEA (Jordan), IFPRI, 3ie-IFPRI, Kent, London School of Economics (STICERD, SERC), Oxford (CSAE), Namur, Paris School of Economics, RAND, Royal Holloway (RES), San Diego (ASSA), SFSU (PACDEV), Stellenbosch (WEHC), Sussex, Tufts, Virginia and World Bank, for helpful comments. We thank Amy Guisinger, Michelle Khilji and Jon Rothbaum for excellent research assistance. The financial support of PSE, AMID-LSE, GWU-IIEP and ESRC First Grant (RES-061-25-0456) is gratefully acknowledged.

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

‡Alexander Moradi, Department of Economics, University of Sussex, Jubilee Building, Falmer, BN1 9SN, 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, UK, and Research Associate, Department of Economics, Stellenbosch University, South Africa.

1. INTRODUCTION

Developing countries have extensively invested in railroads and roads over the past century. Transportation investments continues to remain the single largest item in the budgets of developing countries and development aid agencies alike (World Bank, 2008).[1] Transportation infrastructure investments are indeed considered key to promoting economic growth. We may however wonder if the returns are higher in poor countries with basic infrastructure and high trade costs, and whether there are contexts in which these investments may have little effect. Despite a vast public policy literature on transportation infrastructure and the considerable amounts of money at stake, there is a distinct lack of quantitative evidence on the heterogeneity of the impact of these investments, depending on the context in which they take place.

This paper exploits the construction and later demise of the colonial railroads in Sub-Saharan Africa. One century ago, trade costs were very high south of the Sahara (Chaves, Engerman & Robinson, 2012), due to a lack of reliable transportation (i.e., lack of trails, waterways, and animals immune to the Tsetse fly). Africa only exported high value goods that were headloaded for short distances or slaves who walked longer distances (Nunn, 2008). Economic change was limited to a few areas. Colonial powers built railroads, to ensure military domination and boost trade (Chaléard, Chanson-Jabeur & Béranger, 2006). These investments were substantial, costing the lives of many construction workers and one third of the colonial budgets.[2] Although profitable, post-independence governments ceased investing in railroads. Due to the lack of maintenance, many lines fell into disarray in the 1970s (Gwilliam, 2011). At the same time, road networks were considerably expanded. In this paper, we show that colonial railroads had large, permanent effects in Africa.

We focus on Ghana, for which we create a new dataset on railroads and cities at a precise spatial level over one century: 2,091 cells of 0.1x0.1 degrees (11x11km) in 1891-2000. We first show how the construction of the colonial railroad had a large impact on city growth. The British colonizer built two lines to link the coast to mining areas and the hinterland. These lines went through dense, lowly-populated tropical forests, but the decrease in internal

[1]Due to recent commodity booms, several African countries have initiated various giant transportation projects, such as the Tanzania-Gabon railway ($33 billion), the Mombasa-Kampala-Kigali railway ($14 billion), the Trans-Kalahari railway ($9 billion) and the Abidjan-Lagos motorway ($8 billion).

[2]Our data indicates that 90% of total rail mileage in Sub-Saharan Africa was built prior to independence.

trade costs encouraged the local cultivation of cocoa, of which Ghana became the world's largest exporter. Rural populations increased along the lines as cocoa cultivation required more labor in cocoa-producing villages. Urban populations increased because villages used the towns as trading stations. We use various identification strategies in an attempt to measure causal effects. Our results also apply to 39 African countries, for which we have similar data for 194,000 cells with the same degree of precision in 1890-2010.

Next, we show the persistence of this spatial equilibrium following the demise of the colonial railroad in the post-independence period. Although the railroad locations have lost their initial relative advantage in terms of transportation, we find that they remain relatively more developed today. The effects can be attributed to colonial urbanization. We show that the spatial equilibrium became stable as soon as the rail was built, and that subsequent transportation technologies (i.e., roads) did not decentralize economic activity away from the railroad locations. We argue that while colonial sunk investments matter in explaining why the railroad locations are still more attractive today, railroad cities also persisted because their emergence helped coordinate investments across space for each subsequent period.

The literature has argued over the importance of geography versus history for the spatial distribution of economic activity. If geographical factors are the main drivers, any localized historical shock will only temporarily affect the distribution (Davis & Weinstein, 2002, 2008; Miguel & Roland, 2011). Thus regional policy interventions may be ineffective. However, the literature has also found support for the presence of local increasing returns and that localized historical shocks can have permanent effects (Bosker et al., 2007, 2008; Redding, Sturm & Wolf, 2011; Ahlfeldt et al., 2012; Bleakley & Lin, 2012; Michaels & Rauch, 2013; Jedwab, Kerby & Moradi, 2014).[3] In this case, regional policies may impact regional development. In this paper, we use two symmetrical shocks: the construction and demise of colonial railroads. If geographical factors are time-invariant, having two symmetrical shocks may help identify the role of local increasing returns.[4] Krugman (1991) contrasts the role of "history" and "expectations' in explaining local increasing returns'. On the one

[3]In a companion paper (Jedwab, Kerby & Moradi, 2014), we use the construction and demise of colonial railroads as well as the settlement and exodus of skilled Europeans and Indians in Kenya to study urban path dependence in the face of shocks to both physical capital and human capital.

[4]There are benefits to studying the obsolescence of a temporary man-made advantage instead of a natural advantage. First, a man-made advantage provides a valid natural experiment if its placement is exogenous and its obsolescence is unexpected. Second, it is hard to prove that a natural advantage is entirely obsolete.

hand, cities are costly to build. *Sunk investments* in infrastructure (history) may therefore account for local increasing returns. On the other hand, if there are returns-to-scale, factors must be co-located in the same locations. There is a spatial *coordination problem* as it is not obvious which locations should have the factors. Existing cities solved this problem for each subsequent period when people expect them to keep thriving. These effects may be different in poor countries, because there are initially fewer cities and no solidified spatial equilibrium as in richer countries. We will argue that the results in our context of mostly rural and poor areas are consistent with the various forces of path dependence in space.

This paper is also related to the literature on transportation infrastructure (see Redding & Turner (2014) for a survey). First, we study the impact in a poor country. Rail building constituted a transportation revolution, as no modern transportation technology existed before. Contrary to Chandra & Thompson (2000), we find that railroads created a change in the level of economic activity rather than just reorganized existing economic activity. Indeed, there is not much economic activity to reorganize in agrarian countries. Railroads thus produced agricultural change, trade integration and structural change, as in 19th century America (Atack & Margo, 2009; Atack et al., 2010; Fajgelbaum & Redding, 2014).[5] Second, these century-old effects have persisted to date. The literature has studied the medium-term effects (0-40 years) of transportation investments, whereas we also study their long-term effects (0-110 years). While transportation infrastructure may facilitate the movement of goods and people, we highlight a new channel by which it has a durable impact. The effects also persisted because the railroads created cities that served as a mechanism to coordinate spatial investments for each subsequent period. Third, we argue that there could be decreasing marginal returns to transportation investments, both *across* and *within* countries. Such investments may have large effects in poor countries, because trade costs are high and there are fewer cities. As countries develop, increasing returns solidify their spatial distribution and subsequent investments may have lower effects on local economic development. In Africa, the spatial equilibrium became stable after the railroads were built, and subsequent investments had smaller effects. Lastly, with the exception of Storeygard (2012), the impact of transportation infrastructure in Africa has been understudied.

[5] Other papers focus on the following dimensions: land values (Bogart, 2009; Donaldson & Hornbeck, 2013), trade (Donaldson, 2013; Duranton, Morrow & Turner, 2013), labor (Michaels, 2008; Duranton & Turner, 2012; Ghani, Goswami & Kerr, 2012) or income (Banerjee, Duflo & Qian, 2012; Storeygard, 2012; Faber, 2014).

Finally, our work is related to the literature on the determinants of city growth in developing countries. Transportation infrastructure may promote urbanization, when reduced trade costs boost trade and raise incomes, as in Atack et al. (2010) and Fajgelbaum & Redding (2014). During the colonial period, the decrease in trade costs has spurred urbanization outside the few large cities. Transportation infrastructure could thus decentralize economic activity, as in Storeygard (2012), Baum-Snow et al. (2012) and Ghani, Goswami & Kerr (2012). However, Faber (2014) finds that transportation investments reinforce the large cities if there are core-periphery effects. We find that urban systems were stable in post-colonial Africa, despite increased road investments in areas unrelated to the railroad lines. Second, we observe the emergence of a complex urban system consisting of villages, agro-towns and trading cities, in line with the central place theory (Christaller, 1933). Rural economic conditions feed into the growth of secondary towns servicing the farming sector, consistent with Jedwab (2013) and Henderson, Storeygard & Deichmann (2013).

The article proceeds as follows. Sections 2 and 3 present the background and the assembled data. Section 4 explains the methodology. Sections 5 and 6 display and discuss the results.

2. HISTORICAL BACKGROUND

In this section, we describe the historical context of our natural experiment, from the construction of the railroad one century ago to its demise in the post-independence period.

2.1 The Railroad Age in Ghana, 1901-1931

Once the British had consolidated their control over Ghana in 1896, they then sought to build transport infrastructure to permit military domination and boost trade previously constrained by high trade costs (Gould, 1960; Luntinen, 1996). Ghana lacked waterways, and draft animals were not used due to the Tsetse fly making head-loading the main method of transportation. There were then few trails because of the thick forest. Railroads were the latest and best mass transportation technology, but the British had to choose between a western, central, or eastern route.

The first line followed the western route (see (W) in Figure 1). Strong interest groups of British capitalists lobbied to connect the gold fields of Tarkwa and Obuasi to the coast, as the mines needed heavy machinery and large quantities of firewood or coal. The colonial administration gave in to the pressure, turning down alternative lines for which surveys attested a higher potential for agricultural exports. The Governorship of Maxwell (1895-1897) was instrumental in the decision-making process. He previously worked in colonial Malaya, where railroads served the tin mines, and he supported the same model of "mining first" for Ghana. Additionally, there were military considerations. The British had fought four wars before they annexed the Ashante Kingdom. The railroad was meant to allow the quick dispatch of troops to the Ashante town of Kumasi. The line started from Sekondi (1898) and reached Tarkwa, Obuasi and Kumasi in 1901, 1902 and 1903 respectively. The line went through virgin forest. Initially, mining accounted for two thirds of the line's traffic.

Colonial governors then long favored a central route (see (C) in Figure 1), but a series of events led to the governorship of Rodger (1904-1910) who decided that the capital Accra should be the terminus of this second line to Kumasi, via the eastern route (see (E) in Figure 1). By 1905, several additional motivations were cited for its construction: the export of cash crops, the exploitation of the goldfields (at Kibi), and the development of tourism (around Abetifi). Construction started in 1908, but completion was delayed by wartime shortages. By 1918, only Tafo had been reached. Kumasi was connected in 1923. A potential concern for the later empirical analysis is whether the placement of this line was exogenous, since cocoa was already grown in small quantities in the area. As the growing industry could be a cause of rail construction, it will be important to show that: (i) production would have remained low without the railroad, (ii) both lines had similar effects, (iii) no positive effects are found for various "placebo" lines, and (iv) results are robust to instrumentation.

Five alternative routes were proposed before the first line was built (see Figure 1). We can address concerns regarding endogeneity by using these lines as a placebo check of our identification strategy. Random events explain why the construction of these routes did not go ahead. First, the Cape Coast-Kumasi line (1873) was proposed to link the then capital Cape Coast to Kumasi to send troops to fight the Ashante. The project was dropped because the war came to an abrupt halt (1874). Second, Governor Griffith wanted a central line from Saltpond to Kumasi (1893) in order to tap the palm oil areas and link the coast to Kumasi. When he retired in 1895, he was replaced by Governor Maxwell who favored the mining lobbies and built the Western Line. For the third and fourth proposed routes, Maxwell thought that a second line was needed, and two projects had two different termini, Apam-

Kumasi and Accra-Kumasi. A conference was to be held in London to discuss the proposals, but Maxwell died before reaching London and neither route was built. Finally, Governor Hodgson favored Accra, but he thought that the line should be built to Kpong rather than Kumasi. He retired in 1904 before work begun, and was replaced by Governor Rodger who built the Eastern Line instead. In addition, we studied lines that were not built in time to affect cocoa production in 1927, the year for which we have GIS data on cocoa cultivation. Cocoa is a perennial crop, and harvest begins after 5 years (Ruf 1995*b*). Hence, to observe any impact on production in 1927, farmers must have planted cocoa trees before 1922. The extension of the Eastern Line from Tafo to Kumasi in 1923, and the Huni Valley-Kade line that was built parallel to the coast in 1927 (to connect a diamond field), thus provide another set of counterfactuals.

The rail reduced trade costs. While the freight rate per ton mile was 5 shillings (s) for head-loading, 3.2s for canoes, 2.5s for lorries, and 1s for steam launch, it was 0.4-0.6s for railroads. This comparison underestimates the magnitude of trade costs as: (i) it only accounts for head-loaders that walked along the few trails, (ii) the few waterways did not serve the areas suitable for cocoa, and (iii) the roads were of poor quality until the late 1920s when the "Tarmet Program" made some roads suitable for motor traffic year-round.[6] We verify that pre-rail trade costs were prohibitively high for cocoa cultivation. Using port prices for cocoa, historical estimates of production costs, and the least cost route to any port in 1901, we estimate the profitability of cocoa cultivation for each location. We find that without railroads production would have been limited to a narrow coastal strip, while with railroads cultivation was profitable in the hinterland (see Web Appendix Figure 1). As Luntinen (1996) commented, "The very existence of the transport network encouraged the production of surplus for the market. It was cocoa that made the Gold Coast the richest colony in Africa. The farmers seized the opportunity as soon as the rail reached them." This is evident in the fact that after 1910, 80% of the cocoa produced was transported by rail.

2.2 The Post-Railroad Age in Ghana

Rail usage continued to grow from 1500 ton miles in 1931, averaging 2,500 ton miles in 1944-1974 (Luntinen, 1996). However, rail traffic collapsed after 1974. In 1984 and

[6]Roads were first complementary to the rail as they were feeders to it. There were only two lorries in 1914 and roads were seasonal or of low quality until 1931. We focus on the railroad age because it gives us a natural experiment to study the impact of modern transportation technology versus no technology.

2000, rail only transported 500 and 900 ton miles, respectively; although, total GDP (and presumably also total traffic) has increased sevenfold since 1931. Similarly, while railroads accounted for more than 70% of cocoa transport until 1970, this share decreased to 5% in 2000. What caused the obsolescence of rail after independence (1957)? Political and economic instability had a detrimental impact on past public investments and management.[7] In 1974, the Ghana Railway Corporation (GRC) employed 15,000 workers, twice as many as in 1957 even though the traffic had not increased. Payroll absorbed 70% of expenditure and the GRC was often in deficit. Service quality was poor, which reduced traffic and freight revenues, thus delaying the maintenance and accelerating the decline of the network.

Furthermore, an agronomic feature of cocoa is that it is produced by "consuming" the forest (Ruf, 1995*b*). Cocoa farmers identify a patch of virgin forest and then replace forest trees with cocoa trees. Pod production peaks after about 25 years, and declines thereafter. When trees are too old, young adult males of producing households move to a new forest, while other household members convert the old cocoa plot into farmland for food production. Jedwab (2013) describes how production in the provinces along the Eastern and Western Lines peaked in the late 1930s and 1960s respectively, and decreased afterwards. The railroad locations have thus lost their initial relative advantage in terms of commercial agriculture.

Lastly, the post-independence governments significantly invested in the road network. Roads were three times cheaper to build, yet maintenance costs were lower for railroads. Ghana's road network increased from 6,000 km in 1931 to 40,000 km today. The railroads were replaced by roads at nearby sites over time, so that railroad cities did not lose their "absolute" access to transportation. Yet they lost their initial "relative" advantage in terms of transportation. Among the main 554 grid cells we use in our analysis, about 50 cells were crossed by a railroad. Among the 500 non-railroad cells, about 30, 250 and 450 of them were crossed by a road in 1901, 1931 and 2000, respectively. Therefore, while few cells were connected to a transportation network (the rail) in 1918, most cells are now connected (via a road). The population could have spread along the roads over time, away from the old railroad lines. Therefore, a different spatial distribution of economic activity could have potentially arisen as a result. However, we find it did not.

[7] This instability includes the overthrow of president Nkrumah and the succession of military coups after 1966, the economic downturn in 1966-1969, and the major economic crisis from 1974-1983.

2.3 Patterns of Economic Transformation in Ghana, 1901-2000

Cocoa has been the main engine of Ghana's development (Hill 1963; Austin 2008). While cocoa was introduced by missionaries in 1859, production remained close to zero in 1901. By 1931, only thirty years later, total production reached 250,000 tons and Ghana was the world's largest exporter of cocoa.[8] Figure 2 shows cells that were suitable or highly suitable for cultivation. Cocoa originally spread out around Aburi, where the British distributed cocoa seedlings. As Ghanaians realized how profitable cocoa was, more and more people specialized in it.[9] While Ghana was then hardly urbanized at the turn of the 20th century, it is now one of the most urbanized countries in Africa. When arbitrarily defining any locality with more than 1,000 inhabitants as a town, Ghana's urbanization rate increased from about 20% in 1901 to 40% in 1931 and to 70% in 2000. Figure 3 shows the spatial distribution of these towns over the last century. Before 1901, towns were chieftaincy towns or coastal trading centers (Dickson, 1968). In the 20th century, most of urban growth took place in the forest zone, with the development of modern transportation, cocoa production and mining.

3. DATA DESCRIPTION

We construct a new data set of 2,091 grid cells of 0.1x0.1 degrees (11x11 km) for the following census years: 1891, 1901, 1931, 1948, 1960 (three years after independence), 1970, 1984 and 2000.[10] We obtain the layout of the railroads (in GIS) from the *Digital Chart of the World*. We use various documents to recreate the history of each line and station, including lines that were planned but not built. Our analysis focuses on the rail network in 1918. We also construct a GIS database on rivers, tracks and roads in 1850-2000.

The data on cocoa land suitability was derived from a survey of soil quality for cocoa. A cell is defined as *suitable* if it contains cocoa soils. It is *highly suitable* or *very highly suitable* if more than 50% of its area consists of forest ochrosols, or first class or second class forest ochrosols respectively. We use an administrative map of cocoa production and GIS to calculate the amount produced (tons) for each cell in 1927. There was no production in 1901. We also have data on cocoa tonnages brought to each railroad station in 1918. We then use census gazetteers to construct a GIS database of localities above 1,000 inhabitants. While we have

[8]The real value of total exports per capita increased fourfold from 1901-1931, which can explained by the export of cocoa. While production was nil in 1901, it amounted to 80% of exports by 1931.

[9]Production boomed in Ashanti, around Kumasi, and not in the South-West, closer to Sekondi, because of very poor cocoa soils in the South-West.

[10]See Web Appendix for data sources and construction of variables.

exhaustive urban data for all census years, we only have georeferenced population data for Southern Ghana in 1901 and the whole country in 1931, 1970 and 2000. All cells have the same area, so population levels are equivalent to population densities. Lastly, we have data on local public goods in 1901, 1931 and 2000.

We also construct a new data set of 194,000 grid cells of 0.1x0.1 degrees (11x11km) for 39 Sub-Saharan African countries for the following years: 1890, 1900, 1960, 1970, 1980, 1990, 2000 and 2010. As we did for Ghana, we reconstruct the history of each actual or planned railroad line, and we also create a GIS database on rivers and roads in 2000. We then use census gazetteers and other sources to reconstruct a GIS database of localities above 10,000 inhabitants. We select 10,000 and not 1,000 because consistent historical urban data does not exist below that threshold for most countries. Figure 4 shows the spatial distribution of the colonial railroads and the cities for the 39 countries in 1960 and 2010. There were 73, 436 and 2,911 cities in 1900 (not shown), 1960 and 2011 respectively.[11]

4. ECONOMETRIC MODEL

In this section, we describe the main specifications and identification strategies. Our analysis consists of two steps. First, we will show that the construction of the colonial railroad gave rise to a specific spatial equilibrium. Second, we will test the persistence of this spatial equilibrium, despite the demise of the colonial railroad post-independence.

4.1 Econometric Specifications

First, to study if rail construction gave rise to a specific spatial equilibrium during the railroad age in Ghana (1901-1931), we will run the following models for cells, c, in 1901-1931:

$$\tilde{C}_{c,31} = \alpha + Rail_{c,18}\beta + \lambda\tilde{R}_{c,01} + \rho\tilde{U}_{c,01} + X_{c,01}\gamma + u_c \quad (1)$$

$$\tilde{P}_{c,31} = \alpha' + Rail_{c,18}\beta' + \lambda'\tilde{R}_{c,01} + \rho'\tilde{U}_{c,01} + X_{c,01}\gamma' + u'_c \quad (2)$$

where our dependent variables are the standard scores, or z-scores, of cocoa production in tons ($\tilde{C}_{c,31}$) and rural/urban population ($\tilde{P}_{c,31}$) of cell c in 1931 (e.g., $\tilde{P}_{c,t} = (P_{c,t} - E(P_{c,t})) \setminus \sigma_{P_{c,t}}$). $Rail_{c,18}$ are cell dummies capturing rail connectivity of being: 0-10, 10-20, 20-30 or

[11]For the period between 1960-2010, data for 33 countries was obtained from *Africapolis I: West Africa* and *Africapolis II: Central & Eastern Africa*. *Africapolis* is an attempt by a team of geographers to recreate a consistent database for all African cities using as many sources as possible: Population censuses (reports and gazetteers), administrative counts, demographic surveys and electoral counts. We used the same types of sources to create the data for the 6 remaining countries. For the years 1890 and 1900, we use colonial census reports and colonial handbooks. The sources used for each country are listed in Web Appendix Table 12.

30-40 km away from a line. We control for initial population conditions, by including the z-scores of the rural and urban populations in 1901 ($\tilde{R}_{c,01}; \tilde{U}_{c,01}$) (cocoa production was nil). Population has been growing over time, so the coefficients will mechanically increase for later periods, unless we standardize the variables. Standardizing the variables also helps us compare the effects across the outcomes. Our hypothesis is that railroads have positive effects ($\beta > 0$; $\beta' > 0$) on cocoa production and population growth. We will next include $\tilde{C}_{c,31}$ in model (2) to see if rail connectivity has an effect on population growth through more cocoa production along the lines. To investigate if that spatial equilibrium is persistent after the demise of the colonial railroad, we will run the following model for cells, c, in 1901-2000:

$$\tilde{P}_{c,100} = \alpha'' + Rail_{c,18}\beta'' + \lambda''\tilde{R}_{c,01} + \rho''\tilde{U}_{c,01} + X_{c,01}\gamma'' + u''_c \quad (3)$$

where our dependent variables are the z-scores of the population variables for cell c in 2000 ($\tilde{P}_{c,100}$). We control for the initial population conditions in 1901. We hypothesize that the long-term effects (β'') (i.e., after the demise of the railroads) do not differ from the short-term effects (β') (i.e, when the railroads were built). Our analysis is performed on the sample of cells suitable for cocoa, i.e. the cells of the forested South (see Figure 3). If we use the full sample of 2,091 cells, we run the risk of comparing the southern and northern parts of Ghana, whose geography and history are different (Austin, 2007). Additionally, if unobservable factors explain why the South is more developed than the North, excluding the northern cells will provide more conservative estimates. We also restrict our sample to those cells for which we have data on rural population in 1901. This leaves us with 554 cells.

To show that railroads also gave rise to a specific spatial equilibrium in colonial Africa (1900-1960), and that the spatial equilibrium also persisted post-independence, we will run the following models for 194,000 cells, c, in 39 countries, s, in 1900-1960 and 1900-2010:

$$\tilde{U}_{c,s,60} = \alpha_{ssa} + Rail_{c,s,60}\beta_{ssa} + \rho_{ssa}\tilde{U}_{c,s,00} + \pi_s + X_{c,s,00}\gamma_{ssa} + \nu_{c,s} \quad (4)$$

$$\tilde{U}_{c,s,110} = \alpha'_{ssa} + Rail_{c,s,60}\beta'_{ssa} + \rho'_{ssa}\tilde{U}_{c,s,00} + \pi'_s + X_{c,s,00}\gamma'_{ssa} + \nu'_{c,s} \quad (5)$$

where our dependent variables are the z-score of the urban population of cell, c, in country, s, in 1960 and 2010 ($\tilde{U}_{c,s,60}; \tilde{U}_{c,s,110}$), and we control for the z-score of the urban population in 1900 ($\tilde{U}_{c,s,00}$). $Rail_{c,s,60}$ are the four rail dummies (defined in 1960). We include 39 country fixed effects (π_s). We also drop 77 cells that contain the capital city, largest city and second largest city of each country, as they may have grown for political reasons.

4.2 Identification Strategies

We include various controls to account for contaminating factors. For Ghana ($X_{c,01}$), we add geography variables such as the proportions of suitable, highly suitable, and very highly suitable cocoa soils; the mean and standard deviation of altitude; and average annual rainfall from 1900-1960. We control for economic geography by adding dummies for bordering another country or the sea and Euclidean distances to the coast, a port, a navigable river, Accra, Kumasi and Aburi (the town of origin of cocoa). We also add a dummy equal to one if the cell has a mine in 1901 and add the value of mineral production in 1931.[12] Table 1 shows the mean of each variable for various groups of cells for Ghana. Columns (1) and (2) report the means for the cells within and beyond 10 km from a railroad line, respectively. To test whether the railroad cells (column (1)) differ from the other cells (column (2)), we regress each variable on a dummy equal to one if the cell is within 10 km from a line and test if the difference is significant. The 0-10 km railroad cells have higher rural densities and are closer to main cities, which suggests an upward bias. However, these cells also have a worse cocoa soil quality, which suggests a downward bias. Thus, the directionality of any bias is not obvious. We implement a battery of strategies to attempt to measure causal effects.[13]

Spatial Discontinuities. The cells have the same area as Manhattan. Therefore, neighboring cells should have similar characteristics. In columns (1) and (3) of Table 1, we compare for Ghana the cells 0-10 km with the cells 10-20 km from a line, and find no differences in observables. We can also include nine ethnic groups (sixty-two districts) fixed effects to compare the neighboring cells within a same group (district), as in Michalopoulos & Papaioannou (2012). For the all Africa sample, we can include 755 ethnic group and 2,304 district fixed effects. Alternatively, we can include a fourth-order polynomial of the longitude and latitude of the cell centroid, in order to flexibly control for unobservable spatial factors. The effects are identified from spatial discontinuities, as in Dell (2010).[14]

Mining and Military Lines. In Ghana, while the Eastern Line was potentially endogenous,

[12]For the regressions for Africa, we include the following controls for physical and economic geography ($X_{c,s,00}$): The shares (%) of class 1, class 2, class 3, undetermined, sparsely vegetated and submerged soils in the cell, mean and standard deviation of altitude (m), average annual rainfall (mm) in 1900-1960, two dummies equal to one if the cell is within 10 km from the coast or a navigable river, Euclidean distances (km) to the coast, a navigable river, the largest city, the second largest city or the capital city, respectively.

[13]We find no effects before the lines are built for Ghana. When estimating model (2) with urban population in 1901 as the dependent variable (1891 is the previous year), the coefficients (p-values) of the 0-10, 10-20, 20-30 and 30-40 km rail dummies are: -0.12 (0.16), -0.05 (0.14), -0.09 (0.15) and 0.02 (0.15), respectively.

[14]Web Appendix Figures 2 and 5 map the ethnic groups and districts for Ghana and Africa, respectively.

the Western Line was built for mining and military domination, making its placement less endogenous to cocoa production. Endogeneity is less of a concern if we find similar effects for both lines. Likewise, for each line in Africa, we know if it was built for military domination (because a colonial power wanted to put a claim on a territory or control the native population by being able to dispatch troops), or to connect a mine to the coast. We will test whether the railroad effects are the same when only considering the military and mining lines to create the rail dummies (Web Appendix Figure 6 shows these lines).

Placebo Lines. We can test that there are no spurious effects for cells along the placebo lines, as in Donaldson (2013). We can also use the placebo cells as a control group for the railroad cells, assuming they had the same economic potential. Column (4) of Table 1 reports the means of the controls for the cells within 10 km from a placebo line in Ghana. The placebo cells differ from the railroad cells, but it is not obvious in which direction the estimates could be biased for this control group. We can also compare the railroad lines with each placebo line, as some of them could prove to be a better counterfactual. Likewise, we will use for the 39 African countries various placebo lines that were proposed by the colonizers in 1916 and 1922 but never built (Web Appendix Figure 6 shows these placebo lines).

Instrumentation (IV). We can use as an instrument for the rail in Ghana the distance from the straight lines Sekondi-Tarkwa-Obuasi-Kumasi and Accra-Kumasi (Web Appendix Figure 3 displays the lines). This strategy echoes the work of Michaels (2008) who also uses straight lines. Likewise, for the all Africa sample, we use GIS to create an Euclidean Minimum Spanning Tree (EMST) network based on the initial urban network in 1900 (including the capital, largest and second largest cities and the other cities of each country). We can then use as an instrument for the rail the distance from the EMST straight lines.[15]

5. ESTIMATION RESULTS

We now present the main results of our analysis. We describe the emergence of the spatial equilibrium in 1901-1931, and then show its persistence in 1931-2000.

[15]Web Appendix Figure 7 shows the EMST and the nodes used to compute it for Africa. The EMST is the network that the colonial powers would have built if they had collaborated to optimally connect the initial cities while minimizing construction costs (using the Euclidean distance between them).

5.1 Emergence of the Spatial Equilibrium in Ghana, 1901-1931

5.1.1 Main Results

Table 2 reports the results for models (1)-(2) for the period 1901-1931. Rail connectivity has a strong effect on cocoa production (column (1)), but the effect decreases as we move away from the line. There is then a strong effect on rural population growth up to 30 km (column (2)) and urban population growth up to 10 km (column (4)). In columns (3) and (5), we include the z-scores of the amounts of cocoa produced and cocoa brought to railroad stations in the cell (and a dummy equal to one if the cell contains a train station) to see if rail connectivity drives population growth through more cocoa production along the lines. The population effects are indeed picked up by the cocoa variables. The rail effect on rural growth can be explained by the fact that more production requires more rural labor (column (3)). Cocoa is often produced on the farms surrounding villages (Ruf, 1995*b*).[16] The rail effect on urban growth can then explained by the fact that the more cocoa being transported requires larger rail and trading stations (column (5)). Cocoa farmers also established small producing towns (Hill, 1963). In column (6), the dependent variable is the z-score of a dummy equal to one if there is a town in the cell in 1931 (we already control for urban population in 1901). The rail led to the creation of new towns. Railroads thus induced a cocoa boom, which contributed to rural and urban growth.[17]

Table 3 displays the results when we implement the identification strategies described above. Column (1) replicates our main results from Table 2 (columns (1) and (4)). For the sake of simplicity, we focus on the 0-40 km dummy for cocoa production (Panel A) and the 0-10 km dummy for urban population (Panel B), as there are no effects beyond.[18] The results are robust to: (i) adding ethnic group or district fixed effects (columns (2)-(3)), (ii) including a fourth-order polynomial in longitude and latitude (column (4)), (iii) creating the rail dummies using the Western Line only (column (5)), (iv) restricting the control group to all placebo cells (column (7)). We verify in column (6) that there are no spurious effects for

[16] The relationship is not from railroads to population and then to production. Settlement was limited in the forest due to thick vegetation and high humidity (Ruf, 1995*b*). However, farmers overcame these constraints to grow cocoa. Additionally, cocoa production only required non-urban inputs: land, axes, machetes and labor.

[17] 50% of urban males worked in agriculture according to the 1931 census. Wealthy farmers settled in towns as they offered better living conditions (Hill, 1963). These towns also served as trading stations for exports and imports, as trade accounted for 20% of urban male employment. Also, cocoa generated an income surplus that was spent on urban goods and services (Jedwab, 2013). Manufacturing and services accounted for 30% of urban male employment, whereas consumption goods amounted to two thirds of imports.

[18] Web Appendix Table 1 shows that the effects are also strong for rural and total population growth, when using the same identification strategies as for cocoa production and urban population growth.

the placebo lines,[19] and (v) instrument the rail dummy by a dummy equal to one if the cell is within 40 km from the IV straight lines while dropping the IV nodes (column (8)).[20]

Additionally, these results hold if we (see Web Appendix Table 3): (i) drop the controls, (ii) use the full sample, (iii) drop the railroad nodes, (iv) drop the nodes and their neighboring cells to account for spatial spillovers from the nodes, (v) use the distance to rail stations to create the rail dummies, (vi) use the distance to the rail (km) instead of the rail dummies, since it is a variable of interest that has been often used in the literature, (vii) use the change in the z-score between 1901 and 1931 as the dependent variable,[21] (viii) run a panel regression including cell and year fixed effects,[22] (ix) follow Black & Henderson (1999) in normalizing the dependent variable (for P_t, $P_t^* = P_t/E(P_t)$), instead of standardizing it, (x) use the same specification as in their paper (the dependent variable is $P_t^* - P_{t-1}^*$), (xi) regress the log of the dependent variable on the rail dummies, (xii) use a log-log specification (population density falls by 53% with a doubling of the distance to the rail), and (xiii) use Conley standard errors (100 km) to account for spatial autocorrelation.

5.1.2 Other Measures of Development

For each cell, we know the number of schools, hospitals and churches, and whether the cell is crossed by a road, in both 1901 and 1931. We use model (2) to examine whether railroad cells had better infrastructure by 1931; although, no difference is observed in 1901 (not shown). Results are shown in Panel A of Table 4 (for the sake of space, the coefficients of the 20-30 km and 30-40 km dummies are not reported). We find positive effects on the number of schools (column (1)), and the probabilities of having a hospital (column (2)) or being crossed by a class 1 road (column (4), for 10-20 km as roads were feeders to the rail). We do not find any effect for churches (column (3)). These positive effects decrease when we include the z-scores of the urban and rural populations in 1931 (Panel B). This suggests that railroads increased population density, and public goods were then created.[23]

[19]We create a dummy equal to one if the cell is within X km from a placebo line (where X = 40 and 10 km). We find no spurious effects for each of the seven placebo lines (see Web Appendix Table 2).

[20]The IV F-statistics are high: 164 and 20 respectively. The IV minimizes measurement errors. We also verify that production was not better measured along the lines. Total production was 218,200 tons in the map, against 210,600 tons that were registered at the ports. We use exhaustive census data for population estimates.

[21]We run the following model for cells, c, in 1901-1931: $\tilde{P}_{c,31} - \tilde{P}_{c,01} = \alpha' + Rail_{c,18}\beta' + X_{c,01}\gamma' + u_c'$.

[22]We run the following model for cells, c, and years $t = [1901, 1931]$: $\tilde{P}_{c,t} = Rail_{c,t}\beta + \theta_c + \delta_t + X_{c,01}\gamma_t + u_{c,t}$.

[23]Another interpretation could be that the colonizer invested in public goods at the same time as it was building the rail. In that case, the rail effects on population could be explained by the other investments. However, we do not find that public goods in 1901 were located along the lines (not shown). A second interpretation could be that railroads influenced the placement of the other public goods between 1901 and

Anthropometric data indicates that living standards increased along the lines. Using individual data on Africans recruited by the British Army before independence (Moradi, 2009), we run the following regression on height (H) for 5,725 soldiers, s, born in cell, c, in year, t, (1867-1937): $H_{s,c,t} = a + \zeta Rail_{s,c} + \tau 1(t \geq 1918) + \beta_h Rail_{s,c} * 1(t \geq 1918) + X_{s,c,t}\phi + \upsilon_{s,c,t}$. We compare soldiers born in railroad to non-railroad cells (10 km is the cut-off), after and before 1918. We include the same controls as in the previous model and add individual controls for age, farming, literacy and ethnicity. While railroad cells were not different from other cells ex ante (ζ = -0.04), the height of the soldiers was $\beta_h = 0.66^{**}$ cm higher for those born along the lines after 1918 (see Web Appendix Table 4).

5.1.3 Investigation of General Equilibrium Effects

We analyze whether the railroads created new economic activity, or simply reorganized economic activity across space. Existing pre-colonial cities may have declined as a result of rail building and the emergence of new cities. The railroad effects would then be over-estimated. First, there have been few urban collapses. Only six cells with a city in 1901 did not have a city in 1931. Next, we test whether population increases in railroad cells were accompanied by population decreases in adjacent cells. Column (4) of Table 2 shows that the adjacent 10-40 km cells did not lose urban population relative to the control cells. Likewise, we find no negative effects for any group of adjacent cells (e.g., 10-20 km) when compared to adjacent cells farther away (e.g, 20+ km), as shown until 80 km in Table 5 columns (1)-(8). We also use model (2) on the full sample (N = 2,091) to test if cells that were relatively more developed before the rail lost urban population relative to other cells. The new urban residents also came from the non-forested areas, hence the need to also study them. We create a "Pre-Rail" dummy equal to one if the cell already had a city or had access to transportation before 1901, which we interact with two dummies equal to one if the cell belongs to the "Forest" (N = 554) and the "Non-Forest" (N = 1,537), respectively. We drop the 0-10 km railroad cells in order to compare the pre-rail cells with the other cells. In columns (9)-(12) of Table 5, the pre-rail dummy is equal to one if the cell already had a city in 1901, or if the cell was within 10 km from the coast, a trade route in 1850, or a road in 1901, respectively. We find no negative effect for these other competing cells.

1931, which could cause population to increase. While we cannot rule out this possibility, controlling for public goods in 1901 and 1931 does not alter the relationship between railroads and population (not shown).

Our results suggest that rural-to-urban migration accounts for city growth. Since the new urban residents in the railroad cells were not coming from the same cells (as their rural population also increased), they should have come from villages in the rest of the country. Hill (1963), Ruf (1995*a*) and Austin (2007, 2008) document how the cocoa farmers that populated the towns of the forest were previously farmers that lived in other parts of the country. The rail gave access to a new factor of production – forested land – that increased productivity, as it was used to grow cocoa for export. Data on production costs in 1931 indicate that cocoa farmers are 90% wealthier than subsistence farmers (not shown). The employment share in the cocoa sector was almost one third then. More generally, the number of cities increased by 150 towns and the urbanization rate of the forest by 10 percentage points between 1901 and 1931 (see Figure 4.a). We find the same patterns for the whole country (see Figure 4.b). This structural break in the urban pattern is related to the structural breaks observed in the cocoa exports, total exports and GDP series (see Figure 4.c).

How much of the aggregate changes in urbanization can be attributed to the rail?[24] When estimating model (2) with the cell urbanization rate (%) in 1931 as the dependent variable while controlling for the cell urbanization rate in 1901, we find a positive railroad effect until 10 km (Table 2 column (7)). Since the fast urban growth along the lines was not due to urban reorganization, the aggregate urban population of the forest increased. We thus multiply the number of cells for which the 0-10 km rail dummy is equal to one by the 0-10 km railroad effect when using the non-standardized urban population as a dependent variable. The product gives the "net" increase in the urban population caused by the rail in 1901-1931. As these *urban* residents would have remained *rural* without the railroad, we can estimate by how much lower the total urban population and the urbanization rate of the forest would have been in 1931 without the introduction of the rail. In particular, we estimate that the rail was potentially responsible for 75% of the change in the urbanization rate of the forest and 36% of the national change in the urbanization rate.[25]

[24]The social savings amount to 27% of GDP. This is higher than the estimates for the U.S. in the 19th century, namely 5-10% (Fogel, 1964; Fishlow, 2000), or India in the 20th century, namely 10% (Donaldson, 2013).

[25]This effect is large considering that the 0-10 km railroad cells represent only 3% of Ghana's habitable area. The effect is also a lower-bound estimate since it only includes the local effects of the lines. The cocoa farmers received 65% of the export price then. The difference was captured by the British at the ports, and the cocoa rents spurred the growth of the administrative cities (Dickson, 1968). The effects also spread across the country as the migrant farmers sent remittances to their families in other parts of the country (Hill, 1963).

5.2 Persistence of the Spatial Equilibrium in Ghana, 1931-2000

Column (1) of Table 6 confirms that the long-term effect for the urban populations in 2000 (see model (3)) is similar to the short-term effect for the urban population in 1931 (see model (2), Table 2 columns (4)). Panel C of Table 3 then shows that the long-term effect (0-10 km) is robust to using the same identification strategies as before. Using the same methodology as in the previous subsection, we find that the long-term effects may have potentially accounted for 42% of the change in the urbanization rate of the country in 1901-2000.

The long-term effects are also strong for the other population variables (e.g., rural population, see Web Appendix Table 5).[26] Likewise, railroad cells also have better infrastructure today. For example, we find that railroad cells are more likely to be crossed by a paved (bitumenized) or improved (laterite) road, and their residents also live closer to a secondary school, a clinic and a hospital (see Web Appendix Table 5).

We further investigate the dynamics of urban growth between 1891 and 2000. We run the following model separately for each year where $t = [1901, 1931, 1948, 1960, 1970, 1984, 2000]$: $\tilde{U}_{c,t} = \alpha_t + Rail_{c,18}\beta_t + \rho_t \tilde{U}_{c,t-1} + \lambda_t \tilde{R}_{c,01} + X_{c,01}\gamma_t + u_{c,t}$. The coefficient β_t indicates the rail effects for each period $[t\text{-}1;\ t] = [1891\text{-}1901,\ 1901\text{-}1931\ \ldots\ 1984\text{-}2000]$.[27] Figure 5 displays only the effect of the 0-10 km rail dummy, because the coefficients of the other rail dummies are zero. There was no effect in 1891-1901, a strong effect in 1901-1931, and no additional effect of the railroad post-1931.[28] Cells that had an initial advantage during the railroad age remain relatively more developed today. We also estimate ρ_t, the coefficient of autocorrelation of the urban population. A coefficient of 1 shows that the ranking of the cells in terms of urban population is the same in year t as in year t-1. Here, the coefficient jumps immediately after 1931, from about 0.50 to 0.90, and converges towards 1. This shows how stable the spatial equilibrium became over time. The stability of the equilibrium can also be observed from the evolution of the city size distributions (Black & Henderson,

[26] Results also hold when performing various robustness checks (see Web Appendix Table 6). Another concern is that the mean city size increased over time. The urban population is standardized, so we account for its mean and standard deviation each year. Then, 1,000 may be too low as a cut-off in 2000. Results also hold when using higher cut-offs (see Web Appendix Table 7): (i) 1,666, keeping the same ratio of the minimum to mean city size for 1931 and 2000 (following the methodology of Black & Henderson (1999)), (ii) 10,000, as the population increased tenfold since 1901, and (iii) 2,000, 5,000, 15,000 and 20,000.

[27] The rail dummies are defined using the lines in 1918, since they are more exogenous.

[28] Results hold when running panel regressions with cell and year fixed effects, and cell-specific trends (see Web Appendix Figure 4). We cannot add a lag of the dependent variable in the panel regressions (Nickell, 1981), so the change in the z-score is the dependent variable: $\tilde{U}_{c,t} - \tilde{U}_{c,t-1} = \alpha_t + Rail_{c,18}\beta_t + \lambda_t \tilde{R}_{c,01} + X_{c,01}\gamma_t + u_{c,t}$.

1999). Changes in the distribution can be described as a stationary, first order homogeneous Markov process from 1960 onwards (see Web Appendix Table 8). In particular, the relative size distribution of cities in 2000 is not different from the steady-state size distribution.

5.3 External Validity: Results for Sub-Saharan Africa

We find similar results for the rest of the continent. Column (1) of Panel D in Table 3 uses model (4) to show that there is a positive effect of the 0-10 km rail dummy on the urban population in 1960.[29] Columns (2)-(8) then show that the 0-10 km effect is robust to using the same identification strategies as for Ghana.[30] We also do not find evidence for urban reorganization, since there is no negative effect for the adjacent cells and other competing cells (see columns (9)-(16) of Web Appendix Table 13). Additionally, the total number of cities and urbanization rate of the 39 countries increased by 250 cities and 8 percentage points in 1900-1960, whereas they remained stable in 1890-1900 (see Web Appendix Figure 8). These effects are large. Indeed, Europe's urbanization rate also increased by 8 percentage points during the Industrial Revolution (1700-1870), when using the same city threshold of 10,000 (Malanima & Volckart, 2007). This structural break of urban patterns is related to the structural break observed in GDP per capita post-1900. Using the same methodology as for Ghana, the 0-10 km railroad effects potentially accounts for 55% of the aggregate change in urbanization (excluding the largest cities). These effects are large when considering that the 0-10 km railroad cells only represent 3% of Africa's habitable area. Lastly, we run the main regression for each country individually (the urban population is standardized using only the observations of that country). The mean effect is 0.35, but there are countries where the effect was large, such as Ghana (0.65), and countries where it was nil, such as Namibia (0.00). We then find that the aggregate change in urbanization and the rail effect across countries in 1900-1960 are positively correlated (see Figure 4.d).

The long-term effect on the urban population in 2010 is not different from the short-term effect in 1960 (0.34*** for 0-10 km in column (1) of Table 7, versus 0.37*** in column (1) of Panel D in Table 3).[31] We then investigate the dynamics of urban growth between 1890 and

[29] The coefficients of the other rail dummies are zero (not shown but available upon request).

[30] In column (8), we instrument the rail dummy by a dummy equal to one if the cell is within 40 km from the EMST straight lines while dropping the EMST nodes. The IV F-statistic is 92. Results then hold when only considering the military or mining lines, or the placebo lines in 1916 or 1922 (see Web Appendix Table 13), when performing the robustness checks (see Web Appendix Table 14), or when using higher cut-offs (see Web Appendix Table 15). Many cities were also created along the lines (column (8) of Web Appendix Table 13).

[31] We test that the long-term effect is causal using the main identification strategies (Web Appendix Table

2010, and run the following model separately for each year where $t = [1900, 1960, 1970, 1980, 1990, 2000, 2010]$: $\tilde{U}_{c,s,t} = \alpha_{ssa,t} + Rail_{c,s,60}\beta_{ssa,t} + \rho_{ssa,t}\tilde{U}_{c,s,t-1} + \pi_{s,t} + X_{c,s,00}\gamma_{ssa,t} + \nu_{c,s,t}$. Figure 5 displays the rail effect (0-10 km) for each period. There was no rail effect in 1890-1900, a strong effect in 1900-1960, and no additional effect post-1960.[32] The coefficient of urban autocorrelation, $\rho_{ssa,t}$, jumps after 1960, from about 0.50 to 0.90, and converges towards 1 (it decreased to 0.83 in the 2000s, as economic growth led to urban growth in areas that were not previously developed). Lastly, Africa's city size distribution in 2010 is also not different from its steady state distribution (see Web Appendix Table 17).

6. DISCUSSION

In this section, we investigate the potential channels of path dependence. We focus our analysis on Ghana, as we have better data than for the rest of Africa. We then conclude.

6.1 Discussion on the Channels of Path Dependence

Column (1) of Table 6 shows the baseline long-term effects (0.87 for 0-10 km) for Ghana. They remain unchanged when adding two dummy variables equal to one if the cell is within 10 km from a paved road or an improved road in 2000 (column (2)). Thus, railroad cells are not more developed today simply because they are the only cells with good-quality roads. While few cells were connected to the rail network in 1918, most cells are now connected to the road network. The population could have spread along the roads over time, away from the railroads, but it hardly did. The effects of the railroads (0.83) are four times larger for the roads (0.23). However, the road effects are not necessarily causal. When using the same identification strategies as before, when possible, the road effects still remain lower than the railroad effects (see Web Appendix Table 7).[33] The long-term rail effects are then greatly reduced when we control for the z-scores of urban and rural population in 1931 (column (3)). Thus, the rail effects in 2000 are explained by the rail effects in 1931.[34] Moreover, the road effects become small, which suggests that only the roads built before 1931 had long-

16). The long-term effects also hold when using higher cut-offs in 2010 (see Web Appendix Table 15).

[32] The results hold if we run panel regressions with cell and year fixed effects (see Web Appendix Figure 9).

[33] The road effects remain two to three times lower than the railroad effects, when adding ethnic group or district fixed effects, or a fourth-order polynomial of the longitude and latitude of the cell centroid. We also instrument the paved road dummy with dummies equal to one if the cell was along a trade route in 1850 and/or a class 1/class 2 road in 1931 (we do not instrument the improved road dummy, since the instruments do not predict this dummy). The "old" roads were indeed likely to be paved over time. The IV results should be taken with caution as we cannot be sure that the instruments satisfy the exclusion restriction.

[34] Rural growth should not explain path dependence. Here we control for rural population in 1901 and 1931, and thus their effects on urban population today. The effects are then robust to using other city thresholds.

term effects. In Web Appendix Table 10, we compare the short-term effects of new railroads and roads during the periods 1901-1931, 1931-1960 and 1960-2000. We find positive effects of the transportation investments in 1901-1931. After 1931, new cities were created along the roads, but they are to small to modify the equilibrium. This persistence in spatial patterns is consistent with the existence of local increasing returns, which we now examine.
In section 5.1.2, we showed that connected cells had better infrastructure (schools, hospitals and roads) by 1931. If these colonial *sunk investments* account for urban population today, including them in the model should capture some of the effects of population in 1931 on urban population in 2000. Including the historical factors in 1901 and 1931 reduces the coefficient of urban population in 1931 by 20% (from 0.64 to 0.53, column (4)). One potential concern here is that we may not correctly measure sunk investments, and thus underestimate their contribution to path dependence. Our analysis is biased if we omit other expensive public assets in existence in 1931. However, there were no universities, airports or dams at that time, so we may properly capture the historical factors. Thus, 20% of urban persistence could be potentially explained by colonial sunk investments.[35]

If there are increasing returns, factors of production must also be co-located in the same locations. There is a spatial *coordination problem* as it is not obvious which locations should have the factors. Initially, it is reasonable to co-locate factors in locations that are already developed (i.e., in the railroad cells). The location of factors today depends on past population density, without it being explained by sunk investments. First, we find that urban population in 1931 has a large effect on urban population in 2000, because it leads to higher urban densities in 1960, without it being explained by sunk investments in 1931 (Table 6 column (5)).[36] Second, in section 5.2, we explained that connected cells have better infrastructure in 2000 than non-connected cells, because they are more populated. In column (6) of Table 6, we show that controlling for various contemporary factors (see notes below the table) does not modify the relationship between urban population in 1960 and urban

[35] We find that the coefficient of urban population in 1931 is unchanged when we include (see Web Appendix Table 11): (i) the square of the numbers of each subtype of infrastructure, to account for non-linearities in the effects of infrastructure on population today, (ii) multiple interactions of these factors, to account for complementarities between the different subtypes, and (iii) cocoa production (1927) and the amount of cocoa brought to the rail station (1918), to account for income and thus private capital. The few stone houses that existed in 1931 were built by the wealthy cocoa farmers living in town (Hill, 1963). Otherwise, the towns consisted of houses built using thatch or wood, two materials that are not highly resistant. Therefore, the housing stock in 1931 is unlikely to explain why the railroad cells are more populated today.

[36] We find similar results when using the intermediary years of 1948, 1970 and 1984 (not shown).

population in 2000. This could lead us to believe that the population in the past drives the population today which then attracts other factors. The factors would simply "follow" people. Third, the stability of the spatial equilibrium after 1931 is consistent with the coordination hypothesis. Lastly, one implication of the coordination hypothesis is that there could be marginal decreasing returns to new transportation investments. If initial (rail) investments have created enough large cities, subsequent (road) investments may not create new cities large enough to modify the equilibrium, which is what our results seem to suggest.

Are railroad cities wealthier than non-railroad cities? We estimate model (3) except we use average night light intensity (Table 6 column (7)) and the employment share (%) of industry and services (column (8)) in 2000 as dependent variables. We control for the z-scores of the urban and rural population in 2000, to compare cities of similar sizes. We find a positive effect for railroad cities. While railroad cells are not necessarily better endowed per capita in observable factors than other cells today (see Panel B of Web Appendix Table 5), there could be unobservable factors that were repeatedly co-located along the lines. For example, the railroad cities initially specialized in the export of cocoa, and the import of foreign goods and the production of local goods to satisfy the needs of farmers. Seventy years later, they may still have a comparative advantage in the production of non-food goods and services.

The scarcity of data for Africa does not allow us to precisely examine the channels of path dependence for the rest of the continent. We use the information we have on the year of "connection" of each cell to test if the lines built relatively earlier have larger effects today. Rail building in colonial Africa can be separated into three episodes: 1890-1918, 1919-1945 and 1946-1960. 60% of the cells were connected during the scramble for Africa (1890-1918), and World Wars I and II reduced the number of new connections, due to budget restrictions in Europe (see Web Appendix Figure 10). Whether a line was built in 1890-1918, 1919-1945 or 1946-1960 should not make a major difference in terms of sunk investments. All the lines were built more than 50 years ago, and sunk capital could be equally depreciated for all these periods now. In column (2) of Table 7, we interact the rail dummy with three dummies for each episode of the rail building. The rail effect is higher the earlier the period (the 1946-60 dummy is omitted). In column (3), these differences are attenuated when controlling for the existence of roads in 2000. The pre-1918 effect remains higher than the post-1946 effect. The rail effects then disappear when controlling for urban population in 1960 (column (4)).

This finding is consistent with the coordination problem hypothesis: the cells connected earlier became large cities in 1960, as they solved a coordination problem earlier, and they remain large cities to date. The sudden increase in the coefficient of urban autocorrelation after 1960 is also consistent with the hypothesis. The road effects (0.12-0.20, column (3)) also remain smaller than the railroad effects (0.34, column (1)). Road investments then led to some economic decentralization post-1960, as the road effects remain positive and significant when controlling for urban population in 1960 (column (4)). While the roads were associated with the creation of cities (column (5)), these cities were small since the population effects are lower than the city creation effects (column (4) versus column (5)). In column (6), we use average night light intensity in 2010 as the dependent variable and find that railroad cities are wealthier than non-railroad cities of similar sizes.

To summarize, the persistence of the railroad effects in both Ghana and Africa is consistent with the existence of various forces giving rise to local increasing returns.

6.2 CONCLUSION

To study the impact of transportation investments in poor agrarian countries, this paper exploited the construction and later demise of colonial railroads in Ghana, and Africa as a whole. Railroads constituted a transportation revolution, and had large effects on the spatial distribution and aggregate level of economic activity during the colonial period. These effects have persisted to date; although the railroad systems have collapsed and road networks were considerably expanded in the post-independence period. Our results on the channels of path dependence are then consistent with the existence of local increasing returns.

Our findings suggest that the impact of transportation investments could be heterogeneous. Initial investments may have larger effects in poor countries with basic infrastructure and high trade costs induced agricultural adaption, trade integration and structural change. Then, as countries develop, increasing returns may solidify their spatial distribution and subsequent investments may have lower effects on local economic development. Whether the giant transportation infrastructure projects recently initiated in Africa will have the same local and aggregate effects as the transportation investments that took place over a century ago could thus potentially depend on the context in which they take place.

REFERENCES

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2012. "The Economics of Density: Evidence from the Berlin Wall." Centre for Economic Performance, LSE CEP Discussion Papers dp1154.

Atack, Jeremy, and Robert A. Margo. 2009. "Agricultural Improvements and Access to Rail Transportation: The American Midwest as a Test Case, 1850-1860." National Bureau of Economic Research, Inc NBER Working Papers 15520.

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.

Austin, Gareth. 2007. "Labour And Land In Ghana, 1874-1939: A Shifting Ratio And An Institutional Revolution." *Australian Economic History Review*, 47(1): 95–120.

Austin, Gareth. 2008. "Resources, Techniques, and Strategies South of the Sahara: Revising the Factor Endowments Perspective on African Economic Development, 1500-2000." *Economic History Review*, 61(3): 587–624.

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, Loren Brandt, Vernon J. Henderson, Matthew A. Turner, and Qinghua Zhang. 2012. "Roads, Railroads and Decentralization of Chinese Cities." Unpublished manuscript, Brown University, Department of Economics.

Black, Duncan, and Vernon Henderson. 1999. "Spatial Evolution of Population and Industry in the United States." *American Economic Review*, 89(2): 321–327.

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

Bogart, Dan. 2009. "Turnpike Trusts and Property Income: New Evidence on the Effects of Transport Improvements and Legislation in Eighteenth-Century England." *The Economic History Review*, 62(1): 128–152.

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.

Chaléard, Jean-Louis, Chantal Chanson-Jabeur, and Chantal Béranger. 2006. *Le chemin de fer en Afrique.* Paris: Karthala Editions.

Chandra, Amitabh, and Eric Thompson. 2000. "Does Public Infrastructure Affect Economic Activity?: Evidence from the Rural Interstate Highway System." *Regional Science and Urban Economics*, 30(4): 457–490.

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.

Christaller, Walter. 1933. *Central Places in Southern Germany.* Translation into English by Carlisle W. Baskin in 1966, Englewood Cliffs, NJ: Prentice-Hall.

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.

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

Dickson, Kwamina. 1968. *A Historical Geography of Ghana.* Cambridge: Cambridge University Press.

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.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner. 2013. "Roads and Trade: Evidence from the

US§." *The Review of Economic Studies*.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *The Review of Economic Studies*.

Fajgelbaum, Pablo, and Stephen J. Redding. 2014. "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914." National Bureau of Economic Research Working Paper 20217.

Fishlow, Albert. 2000. *Internal Transportation in the Nineteenth and Early Twentieth Centuries.* in S. L. Engerman and R. E. Gallman (eds.), *Cambridge Economic History of the United States, vol. 2. Cambridge*: Cambridge University Press, 543-642.

Fogel, Robert. 1964. *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: Johns Hopkins University Press.

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.

Gould, Peter R. 1960. *The Development of the Transportation Pattern in Ghana.* Chicago: Northwestern University Press.

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

Henderson, Vernon, Adam Storeygard, and Uwe Deichmann. 2013. "Has Climate Change Promoted Urbanization in Sub-Saharan Africa?" Unpublished manuscript, Department of Economics, Brown University.

Hill, Polly. 1963. *The Migrant Cocoa-Farmers of Southern Ghana: A Study in Rural Capitalism.* Cambridge: Cambridge University Press.

Jedwab, Remi. 2013. "Urbanization without Industrialization: Evidence from Consumption Cities in Africa." Unpublished manuscript, Department of Economics, George Washington University.

Jedwab, Remi, Edward Kerby, and Alexander Moradi. 2014. "History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya." CSAE Working Paper Series 2014-04.

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

Luntinen, Pertti. 1996. *Railway on the Gold Coast: A Meeting of Two Cultures : a Colonial History.* Helsinki: Suomalainen Tiedeakatemia.

Malanima, Paolo, and Oliver Volckart. 2007. *Urbanisation 1700-1780.* Chapter prepared for the Third RTN/CEPR Summer Symposium in London: An Economic History of Modern Europe.

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. 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. *Men under Arms in Colonial Africa: Gold Coast Regiment.* University of Sussex.

Nickell, Stephen J. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417–26.

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

Redding, Stephen J., and Matthew A. Turner. 2014. "Transportation Costs and the Spatial Organization of Economic Activity." National Bureau of Economic Research Working Paper 20235.

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

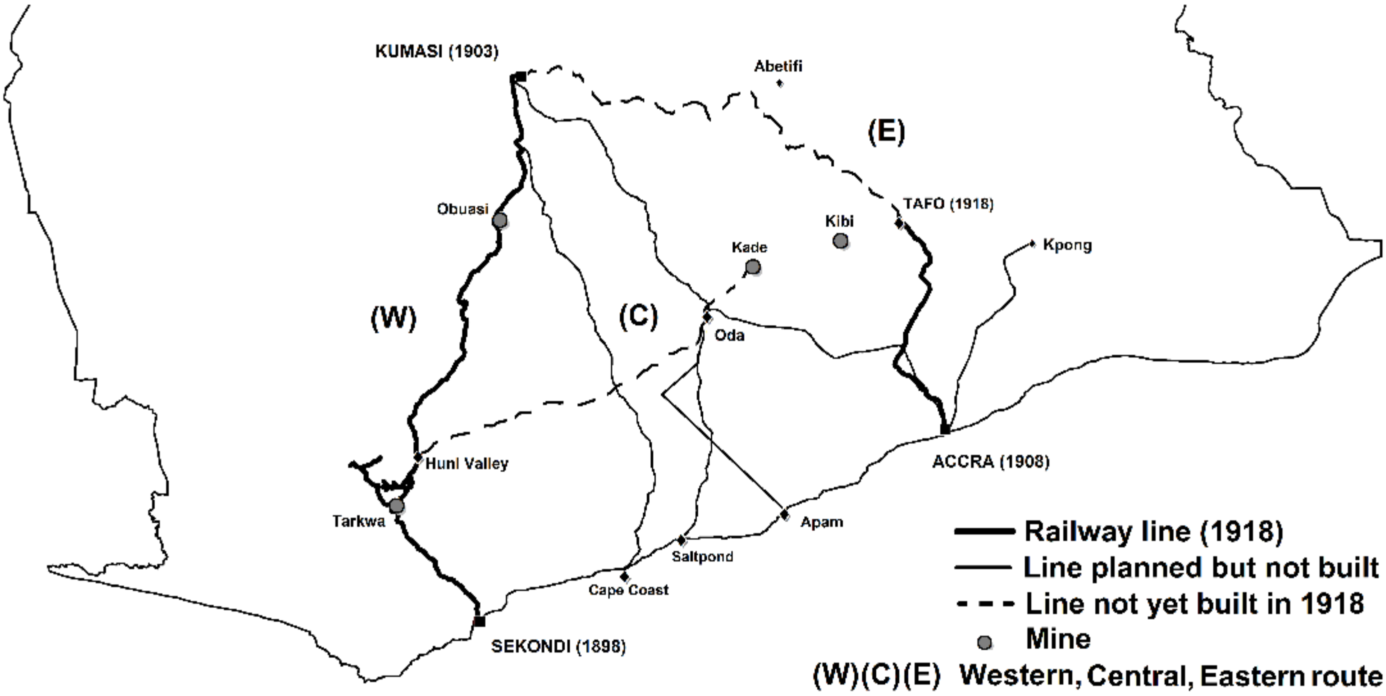
Ruf, François. 1995*a*. *Booms et crises du cacao: Les vertiges de l'or brun.* Paris: Ministère de la Coopéeration, CIRAD-SAR and Karthala.

Ruf, Francois. 1995*b*. "From "Forest Rent" to "Tree Capital": Basic Laws of Cocoa Supply." In *The Economics of Cocoa Supply*. , ed. Francois Ruf and P S Siswputranto. Cambridge: Woodhead Publishing Limited.

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.

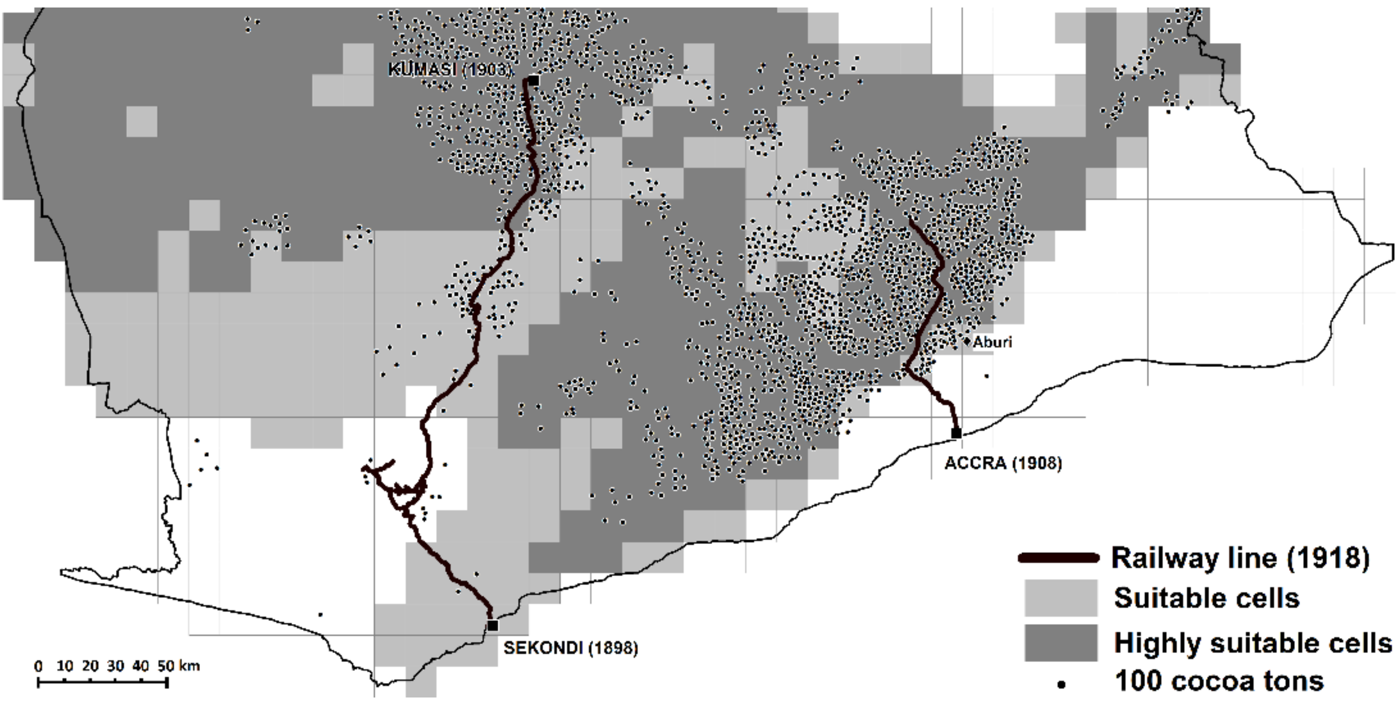
World Bank. 2008. *Safe, Clean, and Affordable... Transport for Development: The World Bank Group's Transport Business Strategy for 2008-2012.* Transport business strategy notes. Washington, DC: World Bank.

Figure 1: Colonial Railroads and Placebo Lines in Southern Ghana



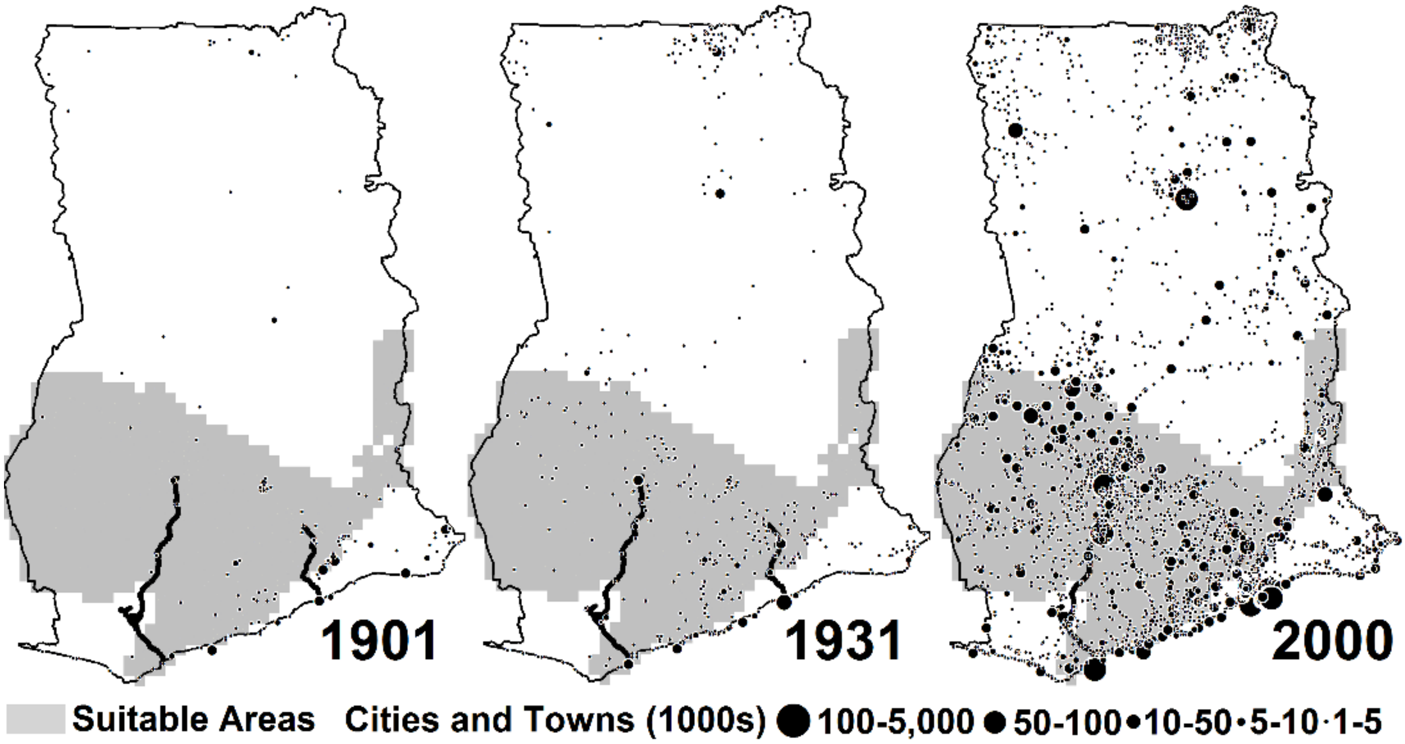
Notes: The map displays the railroad lines in 1918, the years line construction begun and finished for both the *Western Line* (each segment completion is listed below: Sekondi 1898, Tarkwa 1901, Obuasi 1902 and Kumasi 1903) and the *Eastern Line* (Accra 1908, Tafo 1918 and Kumasi 1923), and the seven placebo lines. The Western line was built to connect the mines of Tarkwa and Obuasi to the coast (Sekondi). The line was extended to Kumasi to allow the quick dispatch of troops in the Ashanti areas. The Eastern line was built to connect Accra and Kumasi. Several additional motivations were cited for its construction: the export of cash crops, the exploitation of goldfields at Kibi, and the development of tourism at Abetifi. There are five *lines that were planned but not built* (Cape Coast-Kumasi 1873, Saltpond-Kumasi 1893, Apam-Kumasi 1897, Accra-Kumasi 1897 and Accra-Kpong 1898) and two *lines that were not built early enough to affect cocoa production in 1927* (Tafo-Kumasi 1923 and Huni Valley-Kade 1927). (W), (E) and (C) show the western route, the eastern route, and the lines of the proposed central route respectively. See *Web Appendix* for data sources.

Figure 2: Colonial Railroads, Cocoa Suitability and Production in 1927



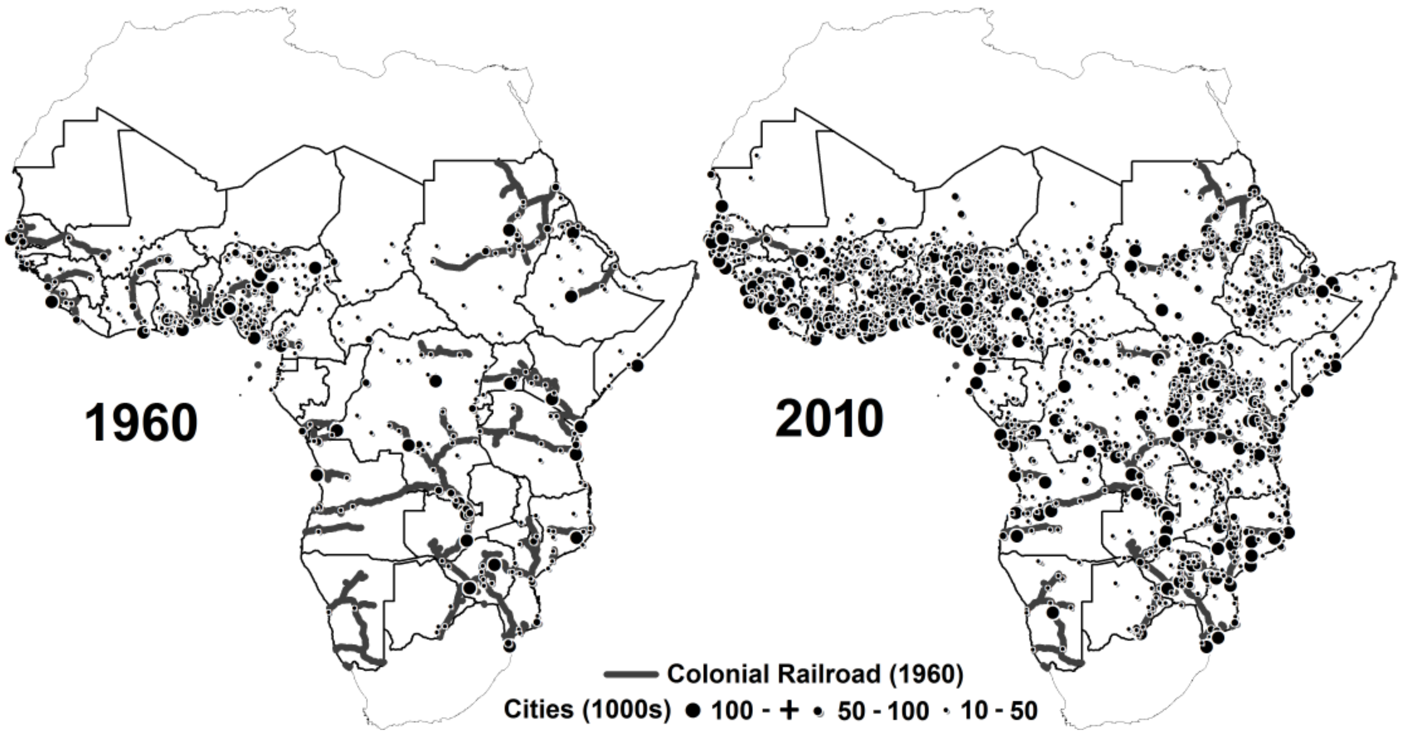
Notes: The map displays the railroad lines in 1918, suitable cells, highly suitable cells, and cocoa production in 1927. The map only shows Southern Ghana. A 0.1x0.1 degree (about 11x11 km) cell is defined as suitable if it contains cocoa soils, and highly suitable if more than 50% of its area consists of forest ochrosols, the best cocoa soils. Each dot represents 100 tons of cocoa production. See *Web Appendix* for data sources.

Figure 3: Colonial Railroads and City Growth, 1901-2000



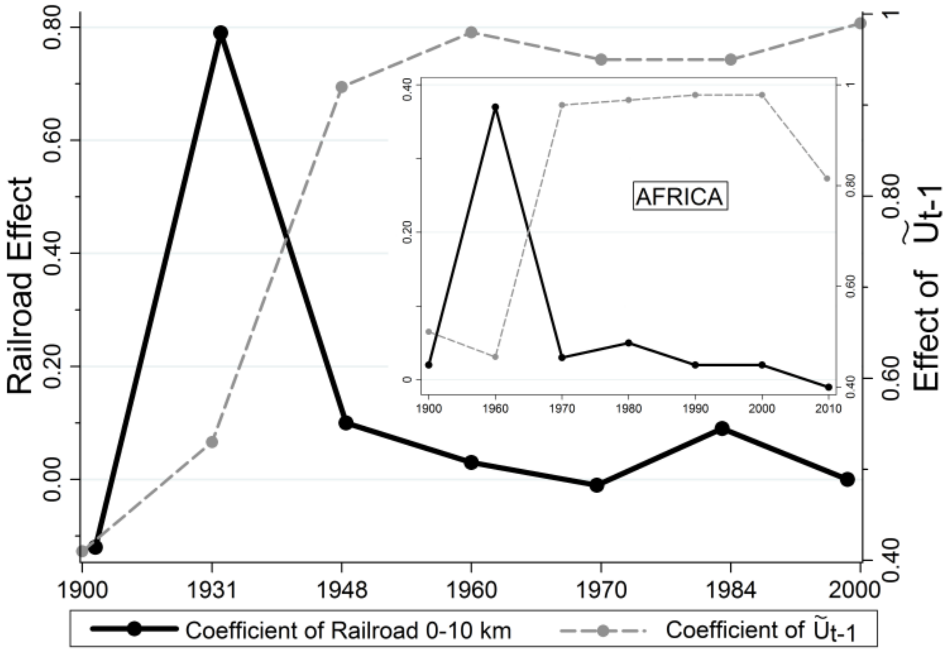
Notes: The maps display the areas suitable for cocoa cultivation (the areas shaded in grey), the railroad lines in 1918 and the cities and towns in 1901, 1931 and 2000. The suitable areas are the cells that contain cocoa soils. Cities and towns in our analysis are localities where the population is greater than 1,000 inhabitants in 1901 (N = 143), 1931 (N = 438) and 2000 (N = 2,991). See *Web Appendix* for data sources.

Figure 4: Colonial Railroads and City Growth for 39 African Countries, 1960-2000



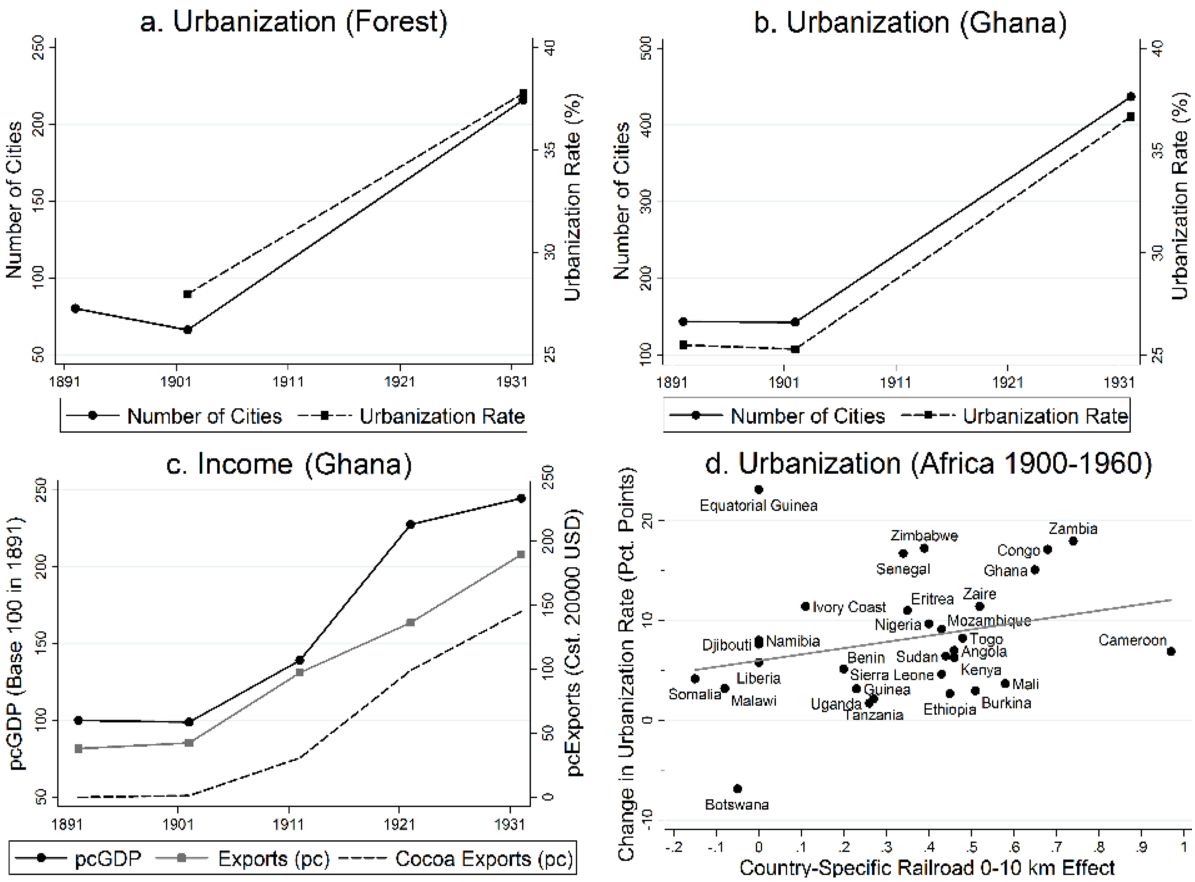
Notes: This map shows colonial railroads and cities for 39 selected Sub-Saharan African countries in 1960 and 2010. We exclude from the analysis four small island countries (Cape Verde, Comoros, Mauritius and São Tomé and Príncipe) and four Southern African countries for which urban data could not be obtained (Lesotho, Madagascar, South Africa and Swaziland). Colonial railroads are defined as lines that were built before the year 1960, when most African countries became independent. Cities are localities where population is greater than 10,000 inhabitants in 1960 (N = 436) and 2010 (N = 2,911). We do not have data for localities below the 10,000 population threshold, except for Ghana. See *Web Appendix* for data sources.

Figure 5: Effects of Colonial Railroads for Each Period, 1891-2000



Notes: The graph displays the coefficients of the 0-10 km rail dummy (1918 for Ghana, 1960 for Africa) for each period in Ghana in 1891-2000 and for 39 selected Sub-Saharan African countries in 1890-2010. The dependent variable is the z-score of the urban population where the years are [1891, 1931, 1960, 1970, 1984, 2000] for Ghana and [1890, 1900, 1960, 1970, 1980, 1990, 2000, 2010] for Africa. Cities are localities where population is greater than 1,000 and 10,000 inh. in Ghana and Africa, respectively. We only display the coefficients of the 0-10 km rail dummy, since the other effects are nil. We show the coefficients of the lag of the dependent variable (the z-score of the urban population in the previous year, $\tilde{U}_{t-1}$). See *Web Appendix* for data sources.

Figure 6: Aggregate Urbanization and Income Patterns for Ghana and Africa



Notes: Figures a. and b. show the evolutions of the number of cities (loc. ≥ 1,000 inh.) and the urbanization rate for the 554 cells of the "forest" and the 2,091 cells of Ghana as a whole in 1891-1931 respectively. Figure c. shows the evolutions of income (per capita GDP, base 100 in 1891) and total and cocoa exports (per capita, constant 2000 USD) in Ghana in 1891-1931. Figure d. shows the relationship between the change in the urbanization rate (pct. points) and the country-specific rail effect (0-10 km) for the 39 African countries in 1900-60 (the solid line is a linear fit for the data). See *Web Appendix* for data sources.

TABLE 1: SUMMARY STATISTICS (MEAN) FOR TREATED AND CONTROL CELLS IN 1901

Group of Cells:	0-10 km (1)	10+ km (2)	10-20 km (3)	Placebo (4)
Mine Dummy	0.04	0.00	0.00	0.01
Urban Population 1901 (Z-Score)	0.06	-0.01	0.16	0.43*
Rural Population 1901 (Z-Score)	0.73	-0.07***	0.28*	0.46*
Soils Suitable for Cocoa (%)	0.76	0.89**	0.83	0.89**
Soils Highly Suitable (%)	0.43	0.64***	0.46	0.72***
Soils Very Highly Suitable (%)	0.11	0.09	0.07	0.13
Altitude: Mean (m)	153	182**	175	167
Altitude: Std. Dev. (m)	36	40.4	44	37
Av. Annual Rainfall (mm)	1,447	1,487	1,481	1,418
Border Dummy	0.00	0.05***	0.00	0.00
Coastal Dummy	0.06	0.02	0.04	0.04
Distance to the Coast (km)	82	111***	83	79
Distance to Port 1901 (km)	97	144***	99	107
Distance to a River (km)	174	161	170	137***
Distance to Accra (km)	149	188***	152	116***
Distance to Kumasi (km)	107	119	110	102
Distance to Aburi (km)	149	185***	150	115***
Mineral Production 1931 (Z-Score)	0.45	-0.04***	0.17	0.18
Number of Cells	49	505	55	129

Notes: This table shows the mean of each variable for various groups of cells. We report the means for the cells within 10 km from a 1918 railroad line (col. (1)), the cells beyond 10 km from a 1918 railroad line (col. (2)), the cells between 10 and 20 km from a 1918 railroad line (col. (3)), and the cells within 10 km from a placebo line (col. (4)), respectively. Col. (2)-(4): We test whether the 0-10 km cells and the control cells (10+ km; 10-20 km; Placebo) are significantly different. We regress each variable on a dummy equal to one if the cell is within 10 km from a 1918 railroad line, and show whether the control cells are significantly different from the 0-10 km cells (robust SEs: * p<0.10, ** p<0.05, *** p<0.01). See *Web Appendix* for data sources.

TABLE 2: COLONIAL RAILROADS AND ECONOMIC DEVELOPMENT, 1901-1931

Dependent Variable (Z-Score; %):	Cocoa Prod. 1927	Rural Pop. 1931		Urban Pop. 1931		City 1/0 1931	Urb.Rate 1931
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rail 1918, 0-10 km	1.06***	0.68***	0.19	0.74***	-0.03	0.41***	6.23*
	(0.22)	(0.17)	(0.22)	(0.28)	(0.13)	(0.15)	(3.57)
Rail 1918, 10-20 km	0.83***	0.45**	0.17	0.09	-0.03	0.29*	2.34
	(0.22)	(0.19)	(0.14)	(0.10)	(0.10)	(0.18)	(4.98)
Rail 1918, 20-30 km	0.53***	0.38**	0.20	-0.07	-0.15	-0.21	-4.76
	(0.16)	(0.15)	(0.15)	(0.12)	(0.11)	(0.15)	(3.70)
Rail 1918, 30-40 km	0.39**	0.10	-0.03	0.06	0.02	-0.02	0.80
	(0.15)	(0.12)	(0.12)	(0.09)	(0.09)	(0.13)	(3.50)
Cocoa Prod. 1927			0.33***		0.13***		
(Z-Score)			(0.07)		(0.04)		
Cocoa at Rail Station 1918			-0.03		0.43***		
(Z-Score)			(0.02)		(0.15)		
Rural Pop. 1901	0.07*	0.51***	0.49***	0.16	0.11**	1.48	0.11**
(Z-Score)	(0.04)	(0.05)	(0.05)	(0.10)	(0.05)	(0.05)	(1.07)
Urban Pop. 1901	-0.02	-0.09	-0.08	0.49***	0.45***	2.72**	0.22***
(Z-Score)	(0.04)	(0.08)	(0.08)	(0.11)	(0.09)	(0.05)	(1.14)
Cell Controls	Y	Y	Y	Y	Y	Y	Y
Observations	554	554	554	554	554	554	554
R-squared	0.55	0.60	0.65	0.46	0.64	0.30	0.33

Notes: OLS regressions using data on 554 cells for the years 1901 and 1931. Robust SEs clustered at the district level (N = 62); * p<0.10, ** p<0.05, *** p<0.01. Table 1 lists the controls. Col. (1): The dependent variable is the z-score of cocoa production (tons) in 1927. Col. (2)-(3): It is the z-score of the rural population (inh.) in 1931. Col. (4)-(5): It is the z-score of the urban population (inh.) in 1931. Col. (6): It is the z-score of a dummy equal to one if the cell has a city (loc. ≥ 1,000 inh.) in 1931. Col. (7): It is the urbanization rate (%) in 1931 (we also control for the urbanization rate in 1901). Col. (3) and (5): We add a dummy equal to one if the cell contains a railroad station in 1918 (coeff. not shown). See *Web Appendix* for data sources.

TABLE 3: IDENTIFICATION STRATEGIES, 1901-1931-2000

Strategy:	*OLS*	*Ethnic*	*District*	*Long.Lat.*	*Min.Mil.*	*Placebo*	*C:Placebo*	*IV*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:	*Dependent Variable = Cocoa Production (Z-Score) in 1927, Ghana*							
Rail 1918, 0-40 km	0.65***	0.83***	0.44***	0.42***	0.32**	-0.04	1.03***	0.79***
(Col.(6): Placebo)	(0.14)	(0.11)	(0.11)	(0.10)	(0.14)	(0.18)	(0.12)	(0.22)
Panel B:	*Dependent Variable = Urban Population (Z-Score) in 1931, Ghana*							
Rail 1918, 0-10 km	0.72**	0.75**	0.69**	0.74**	0.58*	0.08	1.11**	1.24**
(Col.(6): Placebo)	(0.28)	(0.30)	(0.34)	(0.31)	(0.32)	(0.09)	(0.41)	(0.63)
Panel C:	*Dependent Variable = Urban Population (Z-Score) in 2000, Ghana*							
Rail 1918, 0-10 km	0.80**	0.81**	0.62*	0.83**	0.76*	-0.03	1.06**	1.74**
(Col.(6): Placebo)	(0.36)	(0.37)	(0.32)	(0.39)	(0.45)	(0.11)	(0.43)	(0.65)
Panel D:	*Dependent Variable = Urban Population (Z-Score) in 1960, Africa*							
Rail 1918, 0-10 km	0.37***	0.37***	0.28***	0.37***	0.34***	0.32***	0.37***	0.31***
(Col.(6): Placebo)	(0.05)	(0.05)	(0.04)	(0.05)	(0.07)	(0.05)	(0.08)	(0.09)
Cell Controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: OLS regressions using data on 554 Ghanaian cells for the years 1901, 1931 and 2000 and 193,923 African cells for the years 1900 and 1960. Robust SEs clustered at the district level; * p<0.10, ** p<0.05, *** p<0.01. Ghana: We include the z-scores of the urban and rural populations in 1901. Table 1 lists the controls. Africa: We include country FE and the z-score of the urban population in 1900. Footnote 12 lists the controls. We drop the capital, first and second largest cities. Col. (2)-(3): We add 9 (755) ethnic group FE and 62 (2,304) district FE for Ghana (Africa). Col. (4): We add first, second, third and fourth order polynomials of the long. and lat. of the cell's centroid. Col. (5): The rail dummy is equal to one if the cell is within 10 km from the Western line (a line built for mining/military domination) for Ghana (Africa). Col. (6): The rail dummy is equal to one if the cell is within X = [40; 10] km from a placebo line. Col. (7): The control group (C:) is restricted to placebo cells. Col. (8): For Ghana, the instrument is a dummy equal to one if the cell is within 40 km from the straight lines Sekondi-Tarkwa-Obuasi-Kumasi and Accra-Kumasi. For Africa, it is a dummy equal to one if the cell is within 40 km from a straight line of the EMST network based on the initial urban network in 1900. We always drop the nodes. Panels A, B-C and D: The coeff. (F-stat.) of the instrument (1st stage) is 0.69*** (164), 0.17*** (20) and 0.08*** (92), respectively. See *Web Appendix* for data sources.

TABLE 4: COLONIAL RAILROADS AND HISTORICAL FACTORS, 1901-1931

Dependent Variable:	(1) N.Schools 1931	(2) Hospital 1/0 1931	(3) N.Churches 1931	(4) Class 1 Road 1/0 1931
Panel A: Railroads and historical factors in 1931				
Rail 1918, 0-10 km	0.60***	0.12***	0.29	0.02
	(0.22)	(0.04)	(0.46)	(0.07)
Rail 1918, 10-20 km	0.12	-0.00	-0.06	0.18***
	(0.08)	(0.01)	(0.34)	(0.06)
Panel B: Railroads and historical factors in 1931, conditioned on population in 1931				
Rail 1918, 0-10 km	0.17	0.07**	-0.76	-0.00
	(0.16)	(0.03)	(0.46)	(0.07)
Rail 1918, 10-20 km	0.09	-0.01	-0.45	0.19***
	(0.09)	(0.01)	(0.33)	(0.07)
Cell Controls ; Mean	Y ; 0.33	Y ; 0.02	Y ; 1.98	Y ; 0.30

Notes: OLS regressions using data on 554 cells for the years 1901 and 1931. Robust SEs clustered at the district level; * p<0.10, ** p<0.05, *** p<0.01. Table 1 lists the controls. The 20-30 and 30-40 km dummies are included (coeff. not shown). We include lags of the dependent variable in 1901. Panel B: Z-scores of the urban and rural populations in 1931 are added. See *Web Appendix* for data sources.

TABLE 5: COLONIAL RAILROADS AND GENERAL EQUILIBRIUM EFFECTS, 1901-1931

Dependent Variable:	Urban Population (Z-Score) in 1931			
	(1) *X = 0-10*	(2) *X = 10-20*	(3) *X = 20-30*	(4) *X = 30-40*
Rail 1918, X km	0.72**	0.07	-0.09	0.04
vs. Rail 1918, > X km	(0.28) [554]	(0.08) [505]	(0.13) [450]	(0.09) [396]
	(5) *X = 40-50*	(6) *X = 50-60*	(7) *X = 60-70*	(8) *X = 70-80*
Rail 1918, X km	0.02	0.00	-0.02	0.10
vs. Rail 1918, > X km	(0.06) [341]	(0.09) [280]	(0.07) [221]	(0.10) [174]
Pre-Rail Dummy:	(9) *City1901*	(10) *Coastal*	(11) *Route1850*	(12) *Road1901*
Pre-Rail, Forest	0.54**	0.63	0.18***	0.31*
	(0.25)	(0.41)	(0.06)	(0.17)
Pre-Rail, Non-Forest	0.73***	0.71***	0.15***	0.04
	(0.24)	(0.24)	(0.05)	(0.08)

Notes: OLS regressions using data on 554 cells in columns (1)-(8) and 2,091 cells in columns (9)-(12) for the years 1901 and 1931. Robust SEs clustered at the district level; * p<0.10, ** p<0.05, *** p<0.01. Table 1 lists the controls. The 20-30 and 30-40 km dummies are added (coeff. not shown). Col. (1)-(8): The variable is a dummy equal to one if the cell is within X km from a line. We drop the cells below X km (number of obs. in brackets). Col. (9)-(12): We drop the 0-10 km railroad cells. "Forest" includes the main 554 cells. Col. (9): "Pre-Rail" is a dummy equal to one if the cell has a city in 1901. Col. (10): It is a dummy equal to one if the cell borders the sea. Col. (11)-(12): It is a dummy equal to one if the cell is within 10 km from a trade route in 1850, or a road in 1901. See *Web Appendix* for data sources.

TABLE 6: COLONIAL RAILROADS AND URBAN GROWTH, 1901-2000

Dependent Variable in 2000:	Columns (1)-(6): Urban Population (Z-Score)						Night Lights	Industry Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rail 1918, 0-10 km	0.87**	0.83**	0.32	0.24	0.25	0.23	1.68***	2.29*
	(0.37)	(0.37)	(0.24)	(0.23)	(0.17)	(0.16)	(0.46)	(1.31)
Rail 1918, 10-20 km	0.20**	0.18*	0.02	0.07	0.04	0.05	0.34*	3.20**
	(0.09)	(0.09)	(0.08)	(0.07)	(0.05)	(0.04)	(0.21)	(1.18)
Paved Road 2000, 0-10 km		0.22***	0.02	-0.04	-0.02	-0.05	-0.02	1.63*
		(0.07)	(0.09)	(0.06)	(0.04)	(0.04)	(0.25)	(0.86)
Improved Road 2000, 0-10 km		0.23**	0.06	0.04	-0.00	-0.03	0.28	1.08
		(0.11)	(0.12)	(0.11)	(0.06)	(0.05)	(0.22)	(0.87)
Urban Pop. (Z) 1931			0.64***	0.53***	-0.10	-0.12	-0.11	-0.09
			(0.21)	(0.18)	(0.11)	(0.10)	(0.25)	(0.69)
Rural Pop. (Z) 1931			0.19***	0.17**	0.10	0.11	0.39**	-1.65***
			(0.07)	(0.08)	(0.07)	(0.08)	(0.16)	(0.58)
Urban Pop. (Z) 1960					0.87***	0.83***	-1.06**	-2.01**
					(0.11)	(0.11)	(0.48)	(0.87)
Rural Pop. (Z) 1970					-0.03	-0.01	-0.30	0.17
					(0.05)	(0.05)	(0.16)	(0.58)
Urban Pop. (Z) 2000							2.85***	4.85***
							(0.62)	(0.76)
Rural Pop. (Z) 2000							-0.04	0.74
							(0.11)	(0.49)
Cell Controls	Y	Y	Y	Y	Y	Y	Y	Y
Historical Factors 1901, 1931	N	N	N	Y	Y	Y	Y	Y
Contemporary Factors 2000	N	N	N	N	N	Y	Y	Y
Observations	553	553	553	553	553	548	548	548
R-squared	0.43	0.45	0.66	0.69	0.86	0.88	0.80	0.87

Notes: OLS regressions using data on 553 cells for the years 1901, 1931, 1960 and 2000. Robust SEs clustered at the district level; * p<0.10, ** p<0.05, *** p<0.01. The 20-30 and 30-40 km rail dummies are added (coeff. not shown). The sample is the same as in Table 2, except we drop Kumasi. Table 1 lists the controls. Col. (4)-(8): Historical factors in 1901 and 1931: dummies equal to one if the cell has a school, a hospital, or a church, and the respective numbers of schools, hospitals and churches, and dummies equal to one if the cell is within 10 km from a class 1/class 2/class 3 road. Col. (6)-(8): Contemporary factors in 2000: percentages of inh. living less than 5 km from a primary school, a junior or senior secondary school (JSS or SSS), a health clinic, a hospital, a post-office, or a telephone, in a residence with solid walls, a solid roof or floor, or with access to clean water, and percentages of adults aged 25 and over that are literate, have ever been to school, and have finished primary school, JSS, or SSS. We use rural population in 1970 as rural population is missing in 1960. 5 cells are dropped in col.(6)-(8), due to missing data. Col. (7): The dependent variable is mean night light intensity in 2000. Col. (8): It is the employment share of industry and services (%) in 2000. See *Web Appendix* for data sources.

TABLE 7: COLONIAL RAILROADS AND URBAN GROWTH, AFRICA 1900-2010

Dependent Variable in 2010: (Z-Score, Except in Col.(6))	Columns (1)-(4): Urban Population				City 1/0	Night Lights
	(1)	(2)	(3)	(4)	(5)	(6)
Rail 1960, 0-10 km	0.34***	0.16**	0.13**	0.04	0.19*	0.60***
	(0.04)	(0.07)	(0.06)	(0.03)	(0.10)	(0.21)
Rail 1960, 10-20 km	-0.01	-0.01	-0.02***	-0.02***	-0.00	0.36***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.06)
Rail 1890-1918, 0-10 km		0.23***	0.16*	0.01	0.13	1.26***
		(0.09)	(0.08)	(0.04)	(0.11)	(0.28)
Rail 1919-1945, 0-10 km		0.14*	0.06	-0.02	0.09	-0.09
		(0.09)	(0.08)	(0.04)	(0.11)	(0.25)
Paved Road 2000, 0-10 km			0.20***	0.13***	0.37***	0.56***
			(0.02)	(0.02)	(0.02)	(0.05)
Improved Road 2000, 0-10 km			0.12***	0.08***	0.27***	0.15***
			(0.01)	(0.01)	(0.01)	(0.03)
Urban Pop. 1960				0.65***	0.21***	0.27***
(Z-Score)				(0.04)	(0.06)	(0.10)
Urban Pop. 2010						0.21*
(Z-Score)						(0.11)
Country FE, Cell Controls	Y	Y	Y	Y	Y	Y
Adj. R-squared	0.08	0.08	0.09	0.42	0.09	0.14

Notes: OLS regressions using data on 193,923 African cells for the years 1900 and 1960. Robust SEs clustered at the district level; * p<0.10, ** p<0.05, *** p<0.01. The 20-30 and 30-40 km rail dummies are added (coeff. not shown). We include country FE and the z-score of the urban population in 1900. Footnote 12 lists the controls. We drop the capital, first and second largest cities. Col. (2)-(6): We add two dummies equal to one if a cell connected in 1960 was connected either in 1890-1918 or in 1919-1945. Col. (3)-(6): We add two dummies equal to one if the cell is within 10 km from a paved/improved road in 2000. Col. (5): The dependent variable is the z-score of a dummy equal to one if there is a city in the cell in 2010. Col. (6): It is mean night light intensity in 2010. See *Web Appendix* for data sources.