CSAE Working Paper WPS/2013-17

Transportation Technology and Economic Change: The Impact of Colonial Railroads on City Growth in Africa*

Remi Jedwab† and Alexander Moradi‡

Abstract: What is the impact of modern transportation technology on long-run economic change in poor countries with high trade costs? Rail construction in colonial Sub-Saharan Africa provides a natural experiment: 90% of African railroad lines were built before independence, in a context where headloading was the dominant transportation technology. Using new data on railroads and cities over one century within one country, Ghana, and Africa as a whole, we find large permanent effects of transportation technology on economic development. First, colonial railroads had strong effects on commercial agriculture and urban growth before independence. We exploit various identification strategies to ensure these effects are causal. Second, using the fact that African railroads fell largely out of use post-independence, due to mismanagement and lack of maintenance, we show that colonial railroads had a persistent impact on cities. While colonial sunk investments (e.g., schools, hospitals and roads) partly contributed to urban path dependence, evidence suggests that railroad cities persisted because their early emergence served as a mechanism to coordinate contemporary investments for each subsequent period. Railroad cities are also wealthier than non-railroad cities of similar sizes today. This suggests a world where shocks to economic geography can trigger an equilibrium in which cities will emerge to facilitate the accumulation of factors, and thus have long-term effects on economic growth.

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

*We are grateful to Jean-Marie Baland, Robin Burgess, Denis Cogneau, Melissa Dell, Taryn Dinkelman, Dave Donaldson, Marcel Fafchamps, James Fenske, Richard Hornbeck, Elise Huillery, Lakshmi Iyer, Stelios Michalopoulos, Guy Michaels, Edward Miguel, Nathan Nunn, Henry Overman, Jean-Philippe Platteau, James Robinson, Adam Storeygard, Leonard Wantchekon and seminar audiences at U.C. Berkeley, Bocconi (CEPR/AMID), Cambridge (RES), Carlos Tercero, Dartmouth (NEUDC), DIAL, Geneva (AEHW), Georgetown, George Mason, George Washington, Harvard, Houston, 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, University of Barcelona and University of Virginia, for helpful comments. We thank 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, Arts E511, 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.

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

The railway is purported to be one of the most important technological innovations of the modern era. The first inter-city railway in the world was opened between Liverpool and Manchester in 1830. By the late nineteenth century, most European and American cities were connected to the rail network. While railways may have facilitated the transport of labor and goods, and, as a side effect, the circulation of ideas, it is still debated today whether railroads were "indispensable" to economic growth pre-1900 (Fogel, 1964; Hawke, 1970; Atack et al., 2010; Donaldson & Hornbeck, 2013). Since then, 1,400,000 km of rail has been built in the world. Developing countries are building new high-speed lines or upgrading old lines. These countries, and international organizations, are also spending a large share of their budget on roads. In 2011, transportation accounted for 20% of World Bank lending.

While huge sums are invested in railroads and roads in developing countries, little is known on their economic impact. It is difficult to identify a causal effect of transportation technology, as the placement of infrastructure is often endogenous. Second, even less is known on the mechanisms by which a transportation revolution can durably transform an economy. Third, the impact is likely to be higher in poor countries with basic infrastructure and high trade costs, hence the need to know more about the effects of such investments in such contexts. Unfortunately, these countries are not only poor, but they also suffer from data shortage. This limits our understanding of the relationship between transportation and poverty.

To address these difficulties, we use a natural experiment and a new data set on railroads and city growth at a fine spatial level over one century in Sub-Saharan Africa, and our main country of analysis, Ghana. Almost 90% of African railroad lines were built before independence. Although profitable, railroads fell largely out of use in the post-independence period. Yet they had long-term effects on urban and development patterns. In this paper, we document that colonial railroads shaped the economic geography of the continent, and use this setting to show how transportation technology can promote long-run economic change in poor countries.

One century ago, transportation costs were extremely high South of the Sahara (Chaves, Engerman & Robinson, 2012). Africa lacked navigable waterways. Draft animals were not used due to the Tsetse fly transmitting trypanosomiasis. There were only a few well-cleared tracks, that did not become roads before the 1920s. Africa only exported high value goods that were headloaded on short distances, or slaves who walked longer distances (Nunn, 2008). Economic change was limited to the coast. The hinterland was poor, and devoid of cities (Bairoch, 1988). Colonial powers sought to build ways of penetration to the hinterland, to ensure military domination and boost the export of natural resources (Chaléard, Chanson-Jabeur & Béranger, 2006). They built railroads, hoping they would transform the continent.[1] These investments were massive, costing the lives of many construction workers and around one third of colonial budgets. Although profitable, post-independence governments ceased investing in the rail, and many lines collapsed in the 1970s

[1]Our data indicates that 88.3% of total rail mileage in Sub-Saharan Africa was built before independence. Military domination was given as motivation in 35.5% of the cases, (European) mining and commercial agriculture were mentioned in 36.0% and 42.4% of cases respectively.

(Gwilliam, 2011). Political and economic instability caused a lack of maintenance and mismanagement in the rail sector. Moreover, rent-seeking favored large construction projects prone to embezzlement, such as building new roads (Burgess et al., 2013). Passenger and goods traffic thus dropped, and decades of investments were wasted. However, this paper shows that railroads transformed Africa.

As a motivational exercise, we examine the correlation between proximity to a colonial railroad, i.e. a line that was built before 1960, and the current location of cities, here defined as localities above 10,000 inhabitants, for 39 sub-Saharan African countries. Figure 1 shows these railroads and cities in 1960 and 2000. Our units of observation are 194,000 cells of 0.1x0.1 degrees (11x11 km), and our dependent variable is the total urban population of each cell in 2000. We regress urban population in 2000 on our measure of *colonial transportation*, a dummy equal to one if the cell is within 20 km from a rail line in 1960; and three measures of *pre-colonial transportation*: the Euclidean distance to the coast, and two dummies equal to one if the cell is within 20 km from the coast or a navigable river.[2] Colonial railroads have a strong effect on urban population today, as seen in column (1) of Table 1: We estimate that the rail dummy explains 50% of urban patterns today, when excluding the two largest cities of each country. Even if railroads collapsed post-1960, and were replaced by roads at nearby sites, we find that changes in transportation technology do not explain why railroad cells are more developed now. The railroad effect is almost unchanged when controlling for whether the cell is within 20 km from a paved or improved road in 2000 (column (2)), i.e *post-colonial transportation*. The rail effect is then canceled when controlling for urban population in 1960, which is highly significant (column 3). The correlations motivate the following hypothesis: Colonial cities emerged with rail building. Then, colonial railroads had a long-lasting effect on urbanization, due to *path dependence in urban patterns*. Areas that had an initial advantage pre-independence are more developed today.

The previous analysis is not intended to show causation, as the placement of railroads may be endogenous. The paucity of data for the continent also limits our understanding of the mechanisms at play. Why did railroads explain the emergence of cities before 1960? Why did these cities persist after 1960? To address these empirical difficulties, we examine our hypothesis in one country, Ghana, for which results are similar (columns (4)-(6)). This setting is attractive for four reasons.

First, rail building in colonial Ghana (1896-1957) provides a natural experiment. Two lines were built by the British colonizer to link the coast to mining areas and the hinterland. These lines went through dense and lowly-populated tropical forests. We show that the decrease in internal trade costs had a strong (and unexpected) effect on the local cultivation of cocoa for export, making Ghana the world's largest exporter as soon as 1911. Rural population increased along the lines because cocoa cultivation required more labor in producing villages. Urban population increased because villages used towns as trading stations. Since we follow a simple difference-in-difference strategy whereby we compare connected and non-connected cells of 11x11km over time, we provide evidence that the placement of the lines and the timing of rail building were exogenous to the cocoa and population booms.

Second, even if placement was not exogenous, Ghana's history gives us various identification strategies we exploit to confirm our effects are causal. We find no

[2]Regressions include country fixed effects and cell-level controls (see the footnote of Table 1).

effects for a set of placebo lines that were planned but not built. As cocoa trees take five years to produce, we verify there are no effects for lines that were not built in time to affect production the year for which we have data. Lastly, we instrument for rail connectivity with straight lines between the two ports and the hinterland capital, thus using the fact that being on a straight line between two large cities makes it more likely to be connected. These strategies all give similar results.

Third, Ghana's railroads also fell largely out of use post-independence (1957-2000). In addition, cocoa production has disappeared from the old producing areas due to the shifting cultivation process characteristic of this crop. As a result, locations along the railroad lines have lost their initial advantage in terms of both transportation and cash crop production. Yet today, locations along the railroad lines remain relatively more urban and economically developed.

Fourth, we were able to create a new panel data set at a fine spatial level over one century (2,091 cells of 0.1x0.1 degrees, or 11x11km, in 1901-2000). Ghana's 1901 population census was the first geocoded census in Africa. We then use data for later years to study the short- and long-term effects of rail building on urban growth, and economic change more generally. As argued by Lucas (1988), cities are the main engines of growth. Urbanization is our primary measure of development, in line with the literature (Acemoglu, Johnson & Robinson, 2002; Dittmar, 2011). Night lights and structural change are used as alternative measures. We have also collected large amounts of data on colonial and post-colonial infrastructure. We can thus study the mechanisms by which railroad cities persisted post-1957.

What we find is striking. In the early colonial period (1901-1931), the fall in trade costs made cocoa production for export markets profitable. Population increased along the lines, and cities emerged. At independence (1957), locations along the railroad lines were more economically developed, and they still were in 2000. This effect is not explained by changes in transportation technology, as measured by roads today, but is due to urban path dependence. Indeed, the effects of colonial railroads on long-run development are explained by colonial urbanization. As explained by Bleakley & Lin (2012), path dependence could be due to *sunk investments* or the *coordination problem* of contemporary factors. Given that fixed costs are a source of increasing returns, colonial *sunk investments* (e.g., schools, hospitals and roads) could account for urban persistence. Given returns-to-scale in production, factors need to be co-located in the same locations. There is a *coordination problem* as it is not obvious which locations should have the contemporary factors. Then it makes sense to locate factors in locations that are already developed. While colonial *sunk investments* matter, we show that railroad cities mostly persisted because they solved the *coordination problem* early on. Transportation technology may have large permanent effects in poor regions with high trade costs, by creating cities where no mechanisms of coordination (and no or few cities) existed before.

Our findings advance the literature on transportation technology and economic change. Transportation infrastructure can facilitate the circulation of goods, people and ideas. Colonial railroads have boosted exports in Africa, in line with the literature on transportation and trade (Michaels, 2008; Duranton & Turner, 2012; Faber, 2013; Donaldson, 2013). Colonial railroads have encouraged the movement of workers and firms, in line with the literature on transportation and population and employment growth (Baum-Snow, 2007; Atack et al., 2010; Banerjee, Duflo

& Qian, 2012; Baum-Snow et al., 2012; Ghani, Goswami & Kerr, 2012). Colonial railroads have promoted the diffusion of innovations, here the adoption of a new crop, in line with the literature on information and communication technology and development (Jensen, 2007; Aker, 2010; Dittmar, 2011). However, our main contribution is to show how transportation technology can trigger an equilibrium in which cities emerge to facilitate the accumulation of factors, where no such equilibrium existed before, and thus have long-term effects on economic growth. Our context is unique in that we cleanly estimate the long-run effects of modern transportation technology vs. no technology at all. Africa lacked navigable waterways, and draft animals were not used due to the Tsetse fly. Goods were headloaded and slaves walked. Africa differed from Europe (Bosker, 2013), North America (Atack et al., 2010; Atack & Margo, 2011; Donaldson & Hornbeck, 2013), India (Burgess & Donaldson, 2010; Donaldson, 2013) or China (Banerjee, Duflo & Qian, 2012). The impact of a new technology depends on the previously used technologies (e.g., railroads vs. rivers and canals in the antebellum U.S.). The less efficient the old technology is relative to the new one, the larger this impact will be. In Africa, modern transportation technology had large permanent effects because it solved the coordination problem of contemporary factors early on. The paper is thus closely related to the literature on path dependence, and Bleakley & Lin (2012) in particular. They show how a temporary natural advantage in transportation can have a persistent effect by creating a stimulus that shifts local population density to a higher equilibrium. Increasing returns can give rise to multiple urban equilibria.[3] Our objective is different, as we show how a temporary man-made advantage may trigger an urban equilibrium vs. a rural equilibrium, which has implications for long-run growth. Little is known about the channels of path dependence in developing countries (apart from Dell (2012)), especially in Africa where industrial agglomeration effects may be limited. Finally, trade costs are very high in Africa (Atkin & Donaldson, 2012). Storeygard (2012) is the only paper that measures a causal effect of trade costs on growth for the continent, but it uses recent variations in world oil prices, instead of studying rail or road building.[4] Burgess et al. (2013) show how road building is driven by political considerations (ethnic favoritism) instead of economic considerations. This may lower the returns to such investments.

Our focus on colonial railroads also connects with the literature on the impact of colonization on development. We innovate in three ways. First, the literature has mostly focused on the impact of colonial institutions (Acemoglu, Johnson & Robinson, 2001, 2002; Banerjee & Iyer, 2005; Feyrer & Sacerdote, 2009; Dell, 2010; Iyer, 2010; Bruhn & Gallego, 2012), while the effects of colonial investments have been overlooked. Second, the few studies that examined colonial investments highlighted the role of human capital (Glaeser et al., 2004; Huillery, 2009). However, the effects of colonial investments in physical capital (e.g., transportation infrastructure) may have been as large (or even larger). Lastly, we are the first paper to use African population panel data at a fine spatial level over one century.[5]

[3]Redding, Sturm & Wolf (2011) and Bleakley & Lin (2012) find that initial advantages durably alter the location of economic activity given increasing returns. On the contrary, Davis & Weinstein (2002, 2008) and Miguel & Roland (2011) find no long-run impact of the U.S. bombing of Japanese and Vietnamese cities respectively. Their results suggest little path dependence in urban patterns.

[4]There are few studies on urbanization in Sub-Saharan Africa: Storeygard (2012), Henderson, Roberts & Storeygard (2013), Gollin, Jedwab & Vollrath (2013) and Jedwab (2013).

[5]We also contribute to the literature on the historical roots of African underdevelopment: Nunn

The paper is organized as follows. Section 1 presents the historical background of rail building and economic change in Ghana and the data used. Section 2 explains the methodology and shows the results for the colonial period. Section 3 studies why these effects persist over time in Ghana, and Africa. Section 4 concludes.

1. RAILROADS AND CITIES: BACKGROUND AND DATA

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

1.1 New Data on Ghana, 1891-2000

In order to analyze the effect of rail construction on development, we construct a new data set of 2,091 grid cells of 0.1x0.1 degrees (11x11 km) for the following years: 1891, 1901, 1931, 1960 (three years after independence) and 2000 (the year of the latest available census). We choose a high resolution grid since we have very precise GIS data on railroads, population, and cocoa production. We obtain the layout of rail lines in GIS from *Digital Chart of the World*. We then use various documents to recreate the history of rail construction. We know when each line was finished and when each station was opened. We also located lines that were planned but not built. For each line, we create dummies equal to one if the Euclidean distance of the cell centroid to the line is 0-10, 10-20, 20-30, 30-40 or 40-50 km. Our main analysis focuses on the rail network in 1918. We also create a dummy equal to one if the cell contains a railroad station in 1918. We proceed similarly to construct a GIS database on rivers and roads in 1901-2000.

We use census gazetteers to reconstruct a GIS database of localities above 1,000 inhabitants. The number of these localities increased from 144 in 1891 to 2,975 in 2000. Since our analysis is at the cell level, we use GIS to construct the urban population for each cell-year observation. While we have exhaustive urban data in 1891, 1901, 1931, 1960 and 2000, we only have georeferenced population data for Southern Ghana in 1901 and the whole territory in 1931, 1970 and 2000. We calculated rural population by subtracting urban population from total population. All cells have the same area, so population levels are equivalent to population densities. Lastly, we have data on infrastructure provision at the gridcell level in 1901, 1931 and 2000. We also use census data on employment for each cell in 2000.

The data on cocoa land suitability was derived from maps of cocoa soils in Ghana. A cell is defined as *suitable* if it contains cocoa soils. It is *highly suitable* if more than 50% of its area consists of forest ochrosols, the best cocoa soils. It is *very highly suitable* if more than 50% of its area consists of first class or second class ochrosols, the best types of ochrosols. Production data was digitized from a contemporary map and we use GIS to calculate the amount produced (tons) for each cell in 1927. Production was almost zero around 1901 and we know where it was exactly located. We also have data on cocoa tonnages brought to each rail station in 1918.

(2008), Nunn & Wantchekon (2011), Michalopoulos & Papaioannou (2011), Nunn & Puga (2012), Michalopoulos & Papaioannou (2012) and Heldring & Robinson (2012).

1.2 The Railroad Age in Ghana

1.2.1 Railroads Built

Infrastructure investments are driven by the economic potential that justifies them. Hence, a simple comparison of connected and non-connected cells is likely to overstate the output created by it. The railroad age in Ghana provides us with a natural experiment to identify the causal effect of transportation technology on economic change. This summary draws on Gould (1960), Tsey (1986) and Luntinen (1996). The British established the Gold Coast colony in the south and extended their domination to what is now Ghana in 1896. Improving transportation infrastructure was on the agenda, to permit military domination and boost trade historically constrained by high transport costs. Draft animals were not used. Ghana also lacked navigable waterways. Headloading was the main means of transport, although cocoa was also rolled in barrels along a few tracks. Owing to the thick primary forest, there were only a few well-cleared tracks. Railroads were the transportation technology of the time, but the British had to choose between a western, central or eastern route. Figure 2 shows the geographic location of the mentioned lines.

The first line followed the western route (see (W) on Fig. 2). Strong interest groups of British capitalists lobbied to connect the gold fields (Tarkwa, Obuasi) in the hinterland. Mines needed heavy machinery and large quantities of firewood or coal. Headloading made gold production prohibitively costly. The colonial administration gave in to the pressure, turning down alternative lines, for which surveys attested a greater potential for agricultural exports (palm oil). The Governorship of Maxwell (1895-97) was instrumental in the decision-making process. He previously worked in the Malay States where railroads served the tin mines, and he supported the same model of "mining first" for Ghana. There were also military reasons to connect the Ashante capital Kumasi. The British fought four wars before they annexed the Ashante Kingdom in 1896. The railroad was meant to allow the quick dispatch of troops. Construction begun in 1898. The line started from Sekondi on the coast and reached the mines of Tarkwa and Obuasi in 1901 and 1902 respectively, and Kumasi in 1903. Much of the line went through virgin forest. Gold mining accounted for two thirds of the line's traffic (in volume) in 1904-1912. The line had a strong effect on cocoa cultivation, as argued by Tsey (1986, p.303-306). Cocoa freight on this line increased from 0 tons in 1904 to 19,191 tons in 1915.

The second line followed the eastern route (see (E) on Fig. 2). Colonial governors long favored a central route (e.g., from Saltpond or Apam, see (C) on Fig. 2), but a series of unexpected events led to the governorship of Rodger (1904-1910) who thought that the capital Accra had to be the terminus of this second line to Kumasi. By 1905, several additional motivations were cited for its construction (Tsey 1986, p.56-63): the export of palm oil, rubber, and cocoa, the exploitation of the Eastern Akim Goldfields around Kibi, and the development of tourism. Construction started in 1909, but Tafo station was opened in 1917. Rail construction had to stop due to wartime shortages, and Kumasi was connected in 1923. A potential concern is whether the placement was exogenous. Cocoa cultivation originally spread out in the Eastern province from Aburi Botanical Gardens, where the British distributed cocoa seedlings (see Fig. 2, and Section 1.3 below). The province's production was already growing before construction began: around 1,000 tons in 1901, 15,000 tons in 1910 and 100,000 tons in 1925. As cocoa trees take five years to produce,

production before 1914 cannot be attributed to the line.[6] Growing in advance could be a cause of rail construction. For this reason, it will be important to show that: (i) transport was prohibitively costly before, so production would have remained limited to pre-railroad levels, (ii) both lines have similar effects, (iii) results are robust to controlling for the diffusion of cocoa from Aburi, (iv) no positive effects are found for placebo lines, and (v) results are robust to instrumentation.

1.2.2 Reduction in Transportation Costs

Rail permitted a massive decrease in trade costs. While the freight rate per ton mile was 5 shillings (s) for headloading, 3.2s for canoe, 2.5s for lorries (1910, against 1s from 1925), 1.9s for cask rolling, and 1s for steam launch, it was only 0.4-0.6s for railroads. This comparison underestimates the magnitude of trade costs for various reasons: (i) it only concerns headloaders that walked along the few tracks, (ii) cask rolling required good quality roads, (iii) the few navigable waterways did not serve the areas where cocoa could be grown, and (iv) roads were of poor quality until 1924 when the "Tarmet Program" made roads suitable for motor traffic throughout the year. Until the late 1920s, rail was by far the best transportation technology.

We verify that pre-railroad trade costs were prohibitively high without modern transportation infrastructure. Using a GIS map of transportation networks in 1900 (rivers and forest tracks), we estimate for each cell the minimal transport cost of one ton of cocoa to any port. We then use historical estimates of production costs and coastal producer prices to measure the average profitability of cocoa cultivation for each cell, without railroads. We find that cocoa production would have been limited to a narrow coastal strip (Online Appendix Figure 1 displays the cells where cultivation was profitable). As in Donaldson & Hornbeck (2013), the reduction in trade costs must have expanded production in the *feasible* region (along the coast), where production is now even more profitable, and in the *infeasible* region, where production has become profitable. Railroads were thus essential to the colonization of the hinterland. According to Luntinen (1996, p.107), "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 railway reached them." From 1912 on, the share of cocoa transported by rail was around 80%, as shown in Figure 3.

Roads were first complementary to the rail as they were feeders to it. The first lorry was imported in 1903, but there were only two lorries in 1914 and roads were of poor quality until 1924). Roads later became serious competitors for the rail and opened new areas to cocoa. Even if no railroad had been built, roads would have permitted the boom.[7] Our goal is not to compare the respective impacts of railroads and roads. We focus on railroads because it provides us with a natural experiment to identify the impact of modern transportation technology vs. no technology.

[6] The line was first officially opened to traffic in 1912. There is some evidence that farmers went there as soon as 1909 to grow cocoa, expecting the railroad to be opened that year. Given that cocoa trees take five years to produce, we should not see any effect on total production before 1914.

[7] However, since roads were of poor quality until 1924, the coordination failure highlighted in the paper would not have been solved as early as with railroads. Railroads were thus more important than roads because they permitted the adoption of modern transportation technology 30 years before. The right counterfactual here is "no modern transportation technology" before 1924.

1.2.3 Placebo Lines

Five alternative routes were proposed before the first line was built. We can address concerns regarding endogeneity by using these lines as a placebo check of our identification strategy (see Fig. 2). The aim was to ensure military domination and increase trade. Judged by observables, the proposed lines were influenced by soil quality and population density in a similar way as the actual lines built. Only random changes explain why construction did not go ahead. First, the Cape Coast-Kumasi line (1873) was proposed to link the capital Cape Coast to Kumasi to send troops to fight the Ashante. The project was dropped because the war ended too quickly. Second, Governor Griffith advocated the construction of a central line from Saltpond to Oda and Kumasi (1893) to tap 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 the Western Line. Third, Maxwell also thought that the colony needed a central line. There were two competing projects with two different terminus, Apam-Oda-Kumasi (1897) and Accra-Oda-Kumasi (1897). A conference was to be held in London in 1897 to discuss the proposals but unexpectedly Maxwell died before reaching London. Fourth, Maxwell was replaced by Governor Hodgson who favored Accra. However, he thought that the Accra line should be built to Kpong (1898), so as to boost palm oil and cotton production there. Construction was approved in 1903 but Governor Nathan retired in 1904 before works even begun. Fifth, Governor Rodger did not see any interest in a line to Kpong and he proposed the Eastern Line. Construction started in 1909. The rail network was subsequently expanded. Hence, we also consider lines that were not built in time to affect production in 1927. Note that cocoa is a perennial crop. Pod production starts after 5 years (Jedwab 2013). 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 (1923) is a good counterfactual for the Accra-Tafo line (1918). Another line was built from Huni Valley to Kade in 1927, to connect the diamond mines at Kade and encourage agricultural exports. We verify that there are no effects for these lines in 1927.

1.3 Patterns of Economic Change in Ghana

Cocoa has been the main motor of Ghana's development (Austin 2008). Cocoa was introduced by missionaries in 1859, but it took 50 years before cocoa was widely grown, making Ghana the world's largest exporter as early as 1911. Figure 3 shows the aggregate production and export share of cocoa from 1900-1927. Figure 4 shows cells that are suitable or highly suitable for cultivation and production in 1927. Cocoa originally spread out in the Eastern province from Aburi, where the British distributed cocoa seedlings (Hill 1963, p.173-176). As Ghanaians realized how profitable cocoa was, more and more people specialized in it. Why did production boom in Ashanti, around Kumasi, and not in the South-West, closer to Sekondi? The South-West is simply characterized by poor soils and too much rainfall.[8]

Ghana has experienced sustained population growth after 1901. Its population increased from 1.9 million in 1901 to 3.2, 6.7 and 18.9 millions in 1931, 1960

[8]The South-West consists of oxysols or intergrades, which are very poor cocoa soils. The lack of soil minerals causes low yields and premature tree aging. Annual rainfall often exceeds 2,000mm, with a very wet dry season, which favors cocoa diseases.

and 2000 respectively. While Ghana was almost unurbanized at the turn of the 20th century, it is now one of the most urbanized countries in Africa. It started its urban transition earlier than most countries, due to the cocoa boom (Jedwab 2013). Arbitrarily defining as urban any locality with more than 1,000 inhabitants, Ghana's urbanization rate increased from 23.5% in 1901 to 48.6% in 1931, 52.6% in 1960 and 68.5% in 2000. Figure 5 shows the spatial distribution of these towns over one century (Online Appendix Figure 2 shows the patterns for total population). The two largest cities are Accra, the national capital, and Kumasi, the hinterland capital. Before 1901, towns were kingdom capitals or trading centres (Dickson, 1968). Most of the latter were on the coast (see 1901 on Fig. 5), where European merchants would meet local merchants from the interior. In the 20th century, most of urban growth took place in the forest zone (see 1931, 1960 and 2000 on Fig. 5), with the development of modern transportation, cocoa production and mining (Dickson, 1968). Many towns grew because they were cocoa buying centers, the homes of wealthy cocoa farmers, or market towns where they spent their income.

2. RAILROADS AND ECONOMIC CHANGE PRE-INDEPENDENCE

In this section we show that railroads led to economic change during the colonial period. We focus on the railroad age (1901-1931), as it allows us to cleanly estimate the effects of modern transportation infrastructure vs. no technology at all. In particular, we test if connected cells experience a boom in cocoa production, population growth, and urban growth. We explain the various strategies we implement to obtain causal effects. We show that these effects persisted at independence.

2.1 Main Econometric Specification

The main hypothesis we test is whether rail connectivity drove cocoa production and population growth during the railroad age. We follow a simple difference-in-difference strategy where we compare connected and non-connected cells over time. We run the following model for cells c and years $t = [1901, 1931]$:

$$Cocoa_{c,t} = \alpha + Rail_{c,t}\beta + \gamma_t + \delta_c + X_c\zeta_t + u_{c,t} \quad (1)$$

$$Pop_{c,t} = \alpha' + Rail_{c,t}\beta' + \gamma'_t + \delta'_c + X_c\zeta'_t + \nu_{c,t} \quad (2)$$

where our dependent variables are the production (tons) and total population (inhabitants) of cell c in year t. $Rail_{c,t}$ are cell dummies capturing rail connectivity: being 0-10, 10-20, 20-30, 30-40 or 40-50 km away from a line. The dummies are equal to zero in 1901. We include cell and year fixed effects. We expect rail connectivity to have a positive and significant effect on production ($\beta > 0$) and population ($\beta' > 0$). We then include $Cocoa_{c,t}$ in model (2) to see if cocoa captures the effect of railroads on population. If that is the case, it means that rail connectivity has an effect on population growth through more production along the lines. There could be an independent railroad effect on population, so our goal is not to instrument production with railroads, but to highlight one of the mechanisms at play.

We have a panel of 2,091 cells. Our analysis is performed on the restricted sample of suitable cells. If we use the full sample, we run the risk of comparing the southern and northern parts of Ghana, whose geography and history differ. We will show

that results hold when using the full sample. We also restrict our sample to those cells for which we have total population in 1901. We end up with 554 cells, and we believe these restrictions give more conservative estimates. We argue in Section 1.2 that the placement of railroads was not endogenous to production and population. We now describe the tests we perform to ensure these effects are causal.

2.2 Exogeneity Assumptions and Controls

In our analysis, we also include controls at the cell level interacted with year dummies ($X_c\zeta_t$) to account for potentially contaminating factors. We control for economic activity in 1901, such as cocoa production in 1901 and through a dummy equal to one if the cell has a mine.[9] We control for demography in 1901, by including urban and rural populations. We add physical geography variables such as the shares of suitable, highly, and very highly suitable cocoa soils, the mean and standard deviation of altitude (m), and average annual rainfall (mm) from 1900-1960. We control for economic geography by including dummies for bordering another country or the sea, and Euclidean distances (km) to Accra, Kumasi, Aburi (the town of origin of cocoa production), a port in 1901, a navigable river, and the coast.

Second, we test if connected cells and non-connected cells differ in 1901, using the variables above. Even if we control for these factors in our analysis, a significant difference could arise because line placement was endogenous. We regress each control on a dummy equal to one if the cell is less than 20 km from a 1918 rail line. The results reported in Online Appendix Table 1 show that treated cells have a larger rural population and are closer to main cities, which could suggest an upward bias, and have worse cocoa soil quality, which could suggest a downward bias. It is thus not obvious in which direction coefficients could be biased. We can also compare the connected cells with cells that would have been connected if the placebo lines had been built. This guarantees that treatment and control cells are similar in terms of economic potential. The same biases exist when comparing treated cells and all placebo cells. We can also compare treated cells with each placebo line, as some of them could prove a better counterfactual. When compared to cells along the placebo Cape Coast-Kumasi line (1873), treated cell are worse (soil quality, altitude and distance to Accra or the coast) or similar across all dimensions. Using these cells as a control group should lead to a downward bias and give more conservative estimates. We will show later that results are the same whichever control group is selected. Besides, since we have data for 11x11 km cells, neighboring locations are unlikely to differ in terms of unobservables. Cell area is 122 sq km, only 40% more than Manhattan's area. Cells less than 50 km from the lines are all similar in terms of observables.[10] If the placement is truly exogenous, the effect should steeply decrease as we move away from the line, which is what we will show.

Third, even if the placement was endogenous, production would have remained small before the lines were built because trade costs were prohibitively high (see section 1.2.2). This is similar to arguing that the timing of line construction was

[9]There were five mines in 1931: three gold mines, one diamond field, and one manganese mine. Mining exports amounted to 24.2% of exports and the number of Africans engaged in mining was only 12,048. Cocoa and mining accounted for 94.5% of exports.

[10]We regress each control on the rail dummies using the 40-50 km cells as the omitted group. There are no significant differences, except for rural population and, obviously, having a mine for the 0-10 km cells, but these effects are small.

exogenous. We should observe no effect before the lines are built. We run the same model as model (2), except we consider cocoa production and urban population in 1891 and 1901. We have no data on rural population in 1891. Effects are almost nil in 1891-1901, while we will show there are strong effects in 1901-1931.[11]

2.3 Main Results on Economic Change

Table 2 shows the main results for cocoa production and population growth. Column 1 reports the results for model (1), while columns (2)-(8) display the results for model (2). All regressions include cell and year fixed effects and controls. We find a strong effect of rail connectivity on production, but this effect decreases as we move away from the rail line and is zero after 40 km (column (1)). There is a strong effect on population growth up to 20 km (column (2)). People tend to live in the vicinity of the line, although there is some production beyond 20 km. Interestingly, the rail effect is picked up by cocoa production and the cell dummy for having a rail station in 1918 (column (3)). But the rail station effect also becomes lower and non-significant when we include the amount of cocoa produced brought to the station (column (4)). This means that the railroads have a strong effect on population growth, and that this growth is coming from opportunities in the cocoa sector, and other sectors if there are intersectoral linkages.[12]

The railroads have two effects on population growth. We call the first effect, the number of additional inhabitants per ton of cocoa produced, the *labor effect* (1.51***, column (4)), as more cocoa production requires more labor. The comparison of columns (4), (6) and (8) indicates that most of the labor effect takes place in villages (1.15***, column (6)). This is logical as cocoa is produced on farms surrounding producing villages (Jedwab 2013). We call the second effect, the number of inhabitants per ton of cocoa transported, the *trade effect* (0.86**, column (4)), as more cocoa being transported requires larger rail stations. The trade effect occurs in towns only (0.92**, column (8)). When using the urbanization rate as the main outcome, we find positive effects but only significant at 15% (not shown). Indeed, rural population increases almost as much as urban population. To conclude, railroads induced a cocoa boom, which drove both rural and urban growth.[13]

[11]The coefficients (p-values) of the 0-10, 10-20, 20-30, 30-40 and 40-50 km dummies in 1891-1901 are: -2 (0.62), 14 (0.37), -6 (0.51), 3 (0.50) and 5 (0.40) for tons of production, and 90 (0.76), 40 (0.88) and -155 (0.63), 70 (0.71) and 9 (0.96) for urban growth.

[12]Including population in model (1) does not change the railroad effects on production. This confirms that the relationship is not from railroads to population and then to production. Jedwab (2013) explains why the causality does not run from population to cocoa. Settlement was limited in tropical forests due to thick vegetation, high humidity and disease incidence. Farmers overcame these constraints when they could grow cocoa. Besides, cocoa production did not depend on cities for the provision of inputs, as it only required forested land, axes, machetes, hoes and labor. This traditional mode of production was not conducive to a role for cities in the diffusion of innovations.

[13]The 1931 census indicates that 48.5% of the urban male workforce worked in agriculture. Wealthy farmers settled in towns as they offered better living conditions (Hill, 1963). Second, towns served as trading stations for exports and imports. Trade accounted for 75% of rail traffic in 1904-1931, and 20.6% of urban male employment. Third, cocoa generated an income surplus that was spent more on "urban" goods and services (Jedwab 2013). For example, consumption goods (clothing, spirits, etc.) amounted to two thirds of imports then. Some of these goods were also produced locally, as manufacturing and services accounted for 30.9% of urban male employment.

2.4 Alternative Identification Strategies and Robustness

Table 3 displays the results when we implement various identification strategies. Column (1) replicates our main results from Table 2 (see columns (1) and (2)). For the sake of simplicity, we only focus on the 0-20 and 20-40 km dummies for production and the 0-20 km dummy for population, as there are no effects beyond.

Western Line vs. Eastern Line. Even if the Eastern Line was potentially endogenous, the Western Line was built for mining and military domination. Endogeneity is not a concern if we find similar effects for both lines. We actually find lower effects for the Western Line (column (2)), but this is explained by the fact that it goes through poorly suitable cells. Indeed, the aggregate effects are stronger for highly suitable cells (not shown). If we restrict the comparison to these highly suitable cells only, the effects are not significantly different (column (3)).

District-Year Fixed Effects. We can also include district-year fixed effects to control for time-variant heterogeneity. Using the district boundaries in 2000, we have 554 cells in 62 districts, or 9 cells per district. We then compare connected cells with the neighboring non-connected cells of the same district time over time. Column (4) shows results are robust to including the 124 district-year fixed effects. Results are also robust if we restrict the control group to the cells in the 40-50 km range, so as to compare neighboring cells using another sample restriction (not shown).

Placebo Regressions. We find no significant positive effects for the seven placebo lines in 1901-1931. For each line, we create a placebo treatment dummy equal to one if the cell is less than X km from the line. For the sake of simplicity, we only use 0-20 km dummies, so we test whether there are positive effects just along the placebo lines. Results are reported in Online Appendix Table 2. First, we expect no effect for the placebo cells. One issue here is that some of the placebo lines intersect with the area of influence (e.g., 0-20 km) of the existing lines, so that there may be a correlation between the treatment and placebo dummies. Therefore, we verify that there are no significant positive effects for the segments of these lines that do not intersect with existing lines. The connected cells now belong to the control group, which leads to significant negative placebo effects. We thus drop the railroad cells to conduct our preferred placebo test, which only compares the placebo cells and the other control cells. Coefficients are small or not significant. Columns (5) of Table 3 show that the main results hold when all the placebo cells are used as control cells. In column (6), the control cells are the cells along the Cape Coast-Kumasi placebo line (20 km), which may prove a better counterfactual (see section 2.2).

Instrumentation. We instrument the treatment with the distance from the straight lines between the two main ports, Sekondi and Accra, and the hinterland city of Kumasi. This strategy echoes the works of Michaels (2008) who instruments U.S. highways with the distance from the straight line joining two major cities, exploiting the fact that transportation networks tend to connect large cities. The instrument is valid as long as it is uncorrelated with any uncontrolled variable that affects the outcome. The Western Line linked Sekondi to the mines of Tarkwa and Obuasi and was extended to Kumasi for military reasons. It went through dense tropical forest and the random location of the mines explained why this line was built from Sekondi to Kumasi. Regarding the Eastern Line, Accra was the administrative capital of Southern Ghana while Kumasi performed the same role for the hinterland. It was obvious

that the two cities would be connected at one point. In column (7), we instrument the rail dummies by dummies for being 0-20 and 20-40 km from the straight lines. In both cases, the IV F-statistic is strong enough (see the notes below the table), and results are unchanged. Results hold if we use the distance from the straight lines Sekondi-Tarkwa-Obuasi and Accra-Kibi instead (see Fig. 2), to exploit the fact that mining was a major motivation behind rail building (not shown).[14]

Robustness. Results hold if we use the distance to rail stations instead, if we drop the controls, if we drop the nodes of the network (including the mines Tarkwa and Obuasi), if we drop the cells neighboring a node to account for spatial spillovers, if we use the full sample of 2,091 cells, if we use a log-linear functional form, or if we use Conley standard errors with a distance cut-off of 200 km to account for spatial autocorrelation (the tests are reported in Online Appendix Table 3).[15]

2.5 Additional Results on Economic Change

We now discuss several additional results on economic change.

Magnitude of Effects. We multiply the railroad effects by the numbers of treated cells, and estimate that the total railroad effects amount to 44.5% of the total change in cocoa production in 1901-1927 and 46.5% of population growth in 1901-1931. Another way to assess the magnitude of the railroad effects is to test whether they were large enough to produce a reversal of fortune in Southern Ghana. 56 cells already contained a town in 1891. These towns were kingdom capitals or trading centres for the slave trade. In 1901, the railroad cells were half as populated as the cells with a town in 1891. We run the same regression as in column (7) of Table 2 except we add a second treatment variable, a dummy equal to one if the cell already had a town in 1891 and the year is 1931. This allows us to compare the railroad effects with the effect of having a town in 1891, relative to the other control cells. While the railroad effects remain the same (2,167** for 0-10 km), the effect of having a town in 1891 was only 1,340*. We find that the railroad effects were strong enough so as to cancel the initial difference in 1901.

Historical Factors. We use the same model as model (2) to examine whether railroad cells had better (non-railroad) infrastructure over time, although no significant difference is observed in 1901. For each cell, we know the number of government and non-government schools, European and African hospitals, and churches, and whether the cell was crossed by a class 1, class 2 or class 3 road, in both 1901 and 1931.[16] Results are displayed in Online Appendix Table 4. There are strong positive effects on the number non-government schools (+0.68** for 0-10 km, given a mean of 0.22) and African hospitals (+0.13* for 0-10 km, given a mean of 0.01), and the

[14] Another advantage of instrumentation is that it solves the classical measurement error problem. In the presence of non-classical measurement errors, the IV estimator is biased upward. Measurement errors are endogenous if production was better measured along the lines. It is not a concern here as total production was 218,200 tons in the 1927 map, against 210,600 tons that were registered at the ports for that year. We then use exhaustive census data for population.

[15] As the railroads were built to connect the coast to mining areas, it is important to verify that the results hold when dropping the cells that contain a mine. The total population of the mining towns, 13,690 inhabitants in 1931, was not entirely exogenous to rail construction. However, 13,690 was small relative to a total urban population of 532,000 for the 554 cells, which minimizes this issue.

[16] Class 1 roads are *suitable for motor traffic throughout the year*, class 2 roads are *suitable for motor traffic but occasionally closed* and class 3 roads are *suitable for motor traffic in dry season only*.

probability of being crossed by a class 1 or class 2 road (+0.20*, +0.29*** and +0.22** for 0-10, 10-20 and 20-30 km respectively, given a mean of 0.24). We find no significant effects for European government schools or hospitals, and churches. These effects strongly decrease when we control for population. This confirms that railroads increased population density, and public goods were created as a result.

General Equilibrium Effects. Population growth was due to in-migration from the non-forested areas, mostly from the South-East and the North. Hill (1963) describes how the "migration involved individual Akwapim, Krobo, Shai, Ga and other Ghanaian farmers from south of the forest belt, in buying forest land which, at the time of purchase, was hardly inhabited." If railroads reallocate labor across space, does overall welfare increase? People migrate because they expect a higher income at the destination location (Harris & Todaro, 1970). Migration leads to a more efficient spatial allocation of resources. In Ghana, the railroads gave access to a new factor of production – forested land – that made people more productive, as it was used to grow cocoa for export. Using data on production and trade costs in 1930, we find that cocoa farmers are 45-90% wealthier than subsistence farmers. Subsistence farmers accounted for 90% of employment in 1901. The employment share in the cocoa sector increased from almost 0% in 1901 to one third in 1931.[17] Thus, one third of the population became more productive as they gained access to the new factor. This allowed Ghana to become one of the wealthiest African countries at independence (Hill 1963; Austin 2008; Jedwab 2013). More than half of population growth happened in cities, another sign of economic change. Railroads thus caused a qualitative change in population, by increasing overall density.

Social Savings. They are calculated as the cost difference between railroads r and the next-best transportation alternative a: Social savings $= (c_a - c_r) \times R$, where c is the marginal cost of the transportation technology and R is the total volume transported by rail. Thus, social savings are the savings to society if the goods are transported using the new technology. We consider headloading as the main alternative to railroads. The social savings are equivalent to 27% of GDP in 1927 vs. 5% for the U.S. in 1890 (Fogel, 1964) and 9.7% for India (Donaldson, 2013).

2.6 Economic Change Post-1931

Has the level effect on economic development in 1931 narrowed over time? We investigate this hypothesis by studying the relative growth of connected cells after 1931. We focus on urban growth, as we are particularly interested in the emergence and growth of cities in Africa, in line with Table 1. Besides, we do not have total population data for the year 1960. We use urban population panel data over one century to study the dynamics of path persistence. We run the following model for the 554 suitable cells c and years $t = [1891, 1901, 1931, 1960, 2000]$:

$$zUPop_{c,t} = \mu + Rail_c\beta_t + \eta_t + \theta_c + X_c\phi_t + \nu_{c,t} \quad (3)$$

with $zUPop_{c,t}$ being the standard score of urban population in year t. We standardize the urban variable to account for demographic growth post-1931. Ghana's

[17]The value of total exports per capita was multiplied by 8. Cocoa explains the export boom. While production was tiny in 1901, it amounted to 80% of exports by 1927. We use aggregate and export data to verify that the production of other cash crops was unaffected.

population was multiplied by 6 in 70 years, and the size of its cities increased as a result of natural increase and advances in urban housing and transportation technologies. $Rail_c$ are the rail dummies, using the railroad lines in 1918, as the analysis in sections 2.2 and 2.4 has confirmed their exogeneity. The lines that were built after 1918 are more endogenous, so there is a trade-off between identifying causal long-term effects and using all the lines that were built pre-1960. For each year t, β_t are the estimates of the rail effects relative to 1901. We include cell and year fixed effects, and the same controls as before interacted with year dummies.

Figure 6 displays the effects β_t. It confirms that connected cells were not different from non-connected cells before railroads were built (1891). The significant positive effects (0-20 km) in 1931 were unchanged over time. In other words, locations that had an initial advantage during the railroad age remained relatively more developed at independence (1960) and today (2000). These long-term effects are robust to controlling for road investments in 1901-2000, restricting the control group to placebo cells, or the cells with a city in 1891, and using log urban population as the dependent variable instead of the standard score (see Online Appendix Figure 3). The fact that the results remain the same when using the placebo cells as control cells confirm that the long-term effects are also causal. The railroad cells are more developed today, because of the initial shock to economic geography.

The railroad lines that were built after 1918 had lower effects (not shown, but available upon request). Section 2.4 showed that the two placebo lines that were built in 1923 (Tafo-Kumasi) and 1927 (Huni Valley-Kade) had no effects on cocoa production and population in 1931. We then find positive effects in 1960 and 2000 for the extension of the Eastern line from Tafo to Kumasi. These effects are twice smaller than for the lines built in 1918. No effect is found for the other placebo line, and the few branch lines that were built thereafter. The lines that were built from 1927 clearly suffered from the competition of colonial roads. Besides, as we will argue later, many of the coordination failures had already been solved by 1931. In other words, the marginal returns to new investments were strongly decreasing.

3. COLONIAL RAILROADS AND LONG-RUN ECONOMIC CHANGE

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

3.1 Evidence on the Decline of Colonial Railroads

By 1931, 500 miles of track had been laid, and rail transported 1,500 tons of goods per mile. At independence, the network reached its maximum size of 700 miles. From 1944 to 1974, rail transported on average 2,500 tons of goods per mile. Traffic collapsed after 1974. In 1984, rail only transported 500 tons of goods. Traffic never fully recovered, and rail transported 900 tons of goods in 2000. Similarly, while railroads accounted for more than 70% of cocoa transport until 1970, this share decreased to 30% in the 1980s and 7% in 2000. Railroads transports manganese and bauxite now, these commodities being too bulky for road transport.

What caused the obsolescence of rail? Luntinen (1996) describes how underinvestments and management issues in the rail sector and considerable road investments

produced a significant decline of the former. First, political and economic instability had a damaging effect on past public investments.[18] By 1980, track, motive power, and rolling stock were in desperate physical condition. There were also management issues. In 1974, the Ghana Railway Corporation (GRC) employed 15,000 workers, twice as many as in 1958 although traffic was the same. Payroll absorbed 70% of expenditure and GRC had been in deficit since 1966. Service quality was poor, which reduced traffic and freight revenues, thus delaying the maintenance and accelerating the decline of the network. Second, the first governments of Ghana massively invested in the road network. Roads were three times cheaper to build. Yet maintenance costs were much lower for railroads. Ghana's total road network increased from 840 km in 1901 to 6,700 km in 1931, 13,400 km in 1960 and 40,000 km in 2000. The railroad lines were obviously replaced by roads, so that railroad cities did not lose their access to transportation. We run the same regression as in model (2) except we consider as the dependent variable a dummy equal to one if the cell is crossed by a paved (improved) road for the years 1960 and 2000. We find that railroad cells are more (less) likely to be crossed by a paved (improved) road. Many improved roads have thus been bitumenized between 1960 and 2000. However, as discussed in section 2.6, we do not find that road expansion explain why railroad cities persisted post-1931. The fact that numerous roads were built in the rest of the country implies that the railroad cells lost their initial advantage in terms of transportation, and other factors must account for urban persistence.

Lastly, an agronomic feature of cocoa is that it is produced by "consuming" the forest. Cocoa farmers go to a patch of virgin forest and replace forest trees with cocoa trees. Pod production peaks after 25 years, and declines thereafter. When trees are too old, cocoa farmers start a new cycle in a new forest. Removing forest trees alters the original environmental conditions and replanted cocoa trees die or are much less productive. Jedwab (2013) uses district panel data from 1901-2010 to describe how cocoa production has disappeared from the original areas of cultivation. Production density in the Eastern province, along the Eastern Line, peaked in 1938 (12.9 tons per sq km of forested land) and decreased afterwards (4.4 on average in 1960-2000). Production density in the Ashanti province, along the Western Line, peaked in 1964 (12.1) before decreasing afterwards (3.9 on average in 1980-2000). Farmers have replaced the old cocoa farms with food crops for urban markets. While we do not have production data at the cell level, we run the same regression as model (3), using the standard score of rural population as the dependent variable for the years 1901, 1931, 1970 and 2000. The population effect in 1931 has narrowed over time (see Online Appendix Figure 4): for example, the coefficient for the 0-10 km dummy has decreased from 0.68*** in 1931 to 0.23*** in 2000. While the rural population of connected cells kept growing post-1931, it grew relatively less than in the rest of the forest, where new cocoa booms occurred.

3.2 Colonial Railroads and The Channels of Path Dependence

In Ghana, railroad cities persisted post-1931 (see Figure 6). The effects were unchanged between 1960 and 2000, although rail traffic collapsed after 1974. In the rest of Africa, railroad cities also persisted after 1960 (see Table 1). Urban persis-

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

tence is not explained by road investments. Then, what could explain it? Bleakley & Lin (2012), who examine the long-term effects of portage sites on population patterns in the U.S., contrast the respective roles of historical and contemporary factors. First, (colonial) *sunk investments* could induce people to stay at these locations. If schools and hospitals are expensive to build, people are less mobile and initial advantages have long-run effects. The long-term effects of historical factors will depend on how fast sunk capital depreciates. The decline of railroads dates from only 30 years ago. Railroad cells are thus likely to be over-supplied with such factors, at least in the short run, and it may take time before population moves to other locations. Second, if there are returns-to-scale in production, factors need to be co-located in the same locations. There is a *coordination problem* as it is not obvious which locations should have the contemporary factors. In this case, it makes sense to locate factors in locations that are already developed, for example the railroad cells. The location of contemporary factors (including people) today then depends on past population density, without it being explained by historical factors. We study how the railroad effects on urban population today vary as we control for the various channels of path dependence, i.e. the historical factors and contemporary factors. We run the following model for 552 suitable cells c:

$$zUPop_{c,2000} = a + Rail_{c,1918}\kappa + X_c\varphi + \omega_{c,t} \tag{4}$$

with $zUPop_{c,2000}$ as the standard score of urban population in 2000 and $Rail_{c,1918}$ as the rail dummies. We use the lines in 1918 because it allows us to estimate causal long-term effects. X_c are the same controls as before. We drop the cells including the nodes. Results are reported in Table 4. Column (1) shows that railroad cells are more developed today. Urban persistence is not explained by road expansion, as controlling for roads in 2000 does not remove the rail effects on urban population today (column (2)). The effects disappear when we control for the standard scores of urban and rural population in 1931 (column (3)). The long-run effects are strongly explained by the population effects in 1931. A one standard deviation in urban population in 1931 leads to a 0.6 standard deviation in urban population in 2000. Rural population in 1931 has a significant positive effect on urban population today, as many villages that boomed then became towns later.

Historical Factors. In section 2.5, we showed that connected cells had better infrastructure by 1931. The number of African schools and hospitals increased, as well as the probability of being crossed by a class 1 or 2 road. These effects were mostly explained by rising population densities by 1931. If these historical factors had an independent long-term effect on urban population today, including them in the regression of column (3) should capture some of the effects of urban and rural population in 1931 on urban population in 2000. We actually find that their inclusion reduces the coefficient of urban population in 1931 by 25% (from 0.60*** to 0.45***). Thus, 25% of urban persistence post-1931 can be explained by sunk investments. These results are in line with the Table III of Bleakley & Lin (2012), where the inclusion of historical factors appears to reduce the long-term effects of portage in the U.S. by similar magnitudes. While historical factors matter to explain urban patterns today, they are not the main channel of path dependence.

Contemporary Factors. We first verify that railroad cells have higher densities of contemporary factors today (2000). We run the same regression as model (3), except we use as a dependent variable the density of various contemporary factors.

Results are displayed in Table 5. In Columns (1), (2), (3), (4), (5) and (8), we show that the inhabitants of railroad cells live closer to a primary school, a junior secondary school, a senior secondary school, a clinic, a hospital and a post-office respectively. The inhabitants of railroad cells are also more likely to have access to clean water (column (6)), and a paved or improved road (column (7)). They are also more likely to live in a house with solid walls (column (8)). These variables control for the different types of factors: human capital, transportation and communication infrastructure, and housing. These effects are strongly reduced when controlling for urban and rural population today. In other words, cells that are more populated today also have higher densities of contemporary factors. These results validate the coordination failure hypothesis. As a second test, we corroborate that urban population in 1931 has a large effect on urban population 2000, because it leads to higher urban densities in 1960 (column (5) of Table 4). In other words, high urban densities at one period leads to high urban densities the next period, and, similarly, in the next periods, because it repeatedly solves the coordination failure for each period. In column (6) of Table 4, we show that controlling for contemporary factors does not strongly modify the relationship between urban population in 1960 and urban population in 2000. These results point to the following story: railroad cells have higher urban densities today, because people co-locate where there are more people in the previous period, and other contemporary factors "follow" people. There were then more people in the previous period because of the population effect in 1931 and repeated co-location decisions. The fact that the long-term effects were also smaller for railroads built later indicate that many of the coordination failures had already been solved by 1931 (see section 2.6).

To conclude, if historical factors matter to explain urban patterns today, railroad cities persisted over time mostly because they solved the coordination failure of contemporary factors as early as 1931, and for each subsequent period then.

3.3 Colonial Railroads and Economic Development Today

We have shown that railroad cells have higher densities of contemporary factors today. However, these positive effects are strongly reduced once we control for urban density. In per capita terms, railroad cities are not better endowed in factors than non-railroad cities of similar sizes. However, these cities could still be wealthier than the other cities. We run the same regression as model (4) except we now use various contemporary measures of economic development as the dependent variable. Results are displayed in Table 6. Given the paucity of income data at a fine spatial level, we use satellite data on night lights as an alternative measure of development, in line with Henderson, Storeygard & Weil (2012). Our dependent variable is average light intensity for each cell in 2000-01. Column (1) shows that the rail effects remain positive and significant when controlling for contemporary population (Panel B vs. Panel A). Henderson, Storeygard & Weil (2012) shows that the elasticity of *ln(lights/area)* to *ln(GDP)* is 0.3. We modify our dependent variable to be in line with their paper, and we find the following railroad effects on *ln(lights/area)* when controlling for *ln(population)*: 1.0*** for 0-10 km and 0.5** for 10-20 (not shown).[19] These cells are thus 30% and 15% wealthier per capita

[19] In column (1) of Table 6, we did not use the same outcome as in Henderson, Storeygard & Weil (2012), as there are cells for which no light is observed by satellite. Using the log of night light intensity mechanically drops these observations. Then, all the cells have the same area in our study.

than other cells. Another measure of development is structural change. In columns (2)-(4), we show that railroad cities have relatively less farmers and relatively more industrial and service workers than non-railroad cities in 2000, for the same city size. These effects are driven by three subsectors (columns (5)-(7)): manufacturing, construction and wholesale and retail trade. We also find more mining in railroad cities (not shown), which is logical since the railroads served two mines. These cities are not administrative centres (column (8)). Lastly, we find that these cities are more diversified, when using the Herfindhal index (column (9)).

The evidence suggests that modern transportation technology can promote long-run economic change. Why are railroad cities wealthier than non-railroad cities of similar sizes? Since we are already controlling for population density, this difference is not due to a higher density of (observable) contemporary factors. But there could be unobservable contemporary factors that were repeatedly co-located along the railroads. For example, railroad cities initially specialized in the trade of cocoa for export, and the trade of imported goods to satisfy the needs of farmers. Seventy years later, these cities still have a comparative advantage in the production and distribution of manufactured goods and services, and are wealthier as a result. Thus, railroads also solved a coordination failure for specific economic activities.

3.4 Long-Run Economic Change: From Ghana to Africa

First, most railroads in the rest of Africa also fell largely out of use in the 1970s (Chaléard, Chanson-Jabeur & Béranger, 2006; Gwilliam, 2011). Second, African countries considerably expanded their road networks during the 1960s and 1970s. Many roads were upgraded, while new roads were built to open new land for development. The road network size is now 1,500,000 km vs. 70,000 km for rail (of which 15,000 km stopped being used). Roads dominate the transport sector, carrying as much as 90% of passenger and freight traffic. However, column (2) of Table 1 has shown that controlling for roads today does not remove the effect of colonial railroads (pre-1960) on cities today (2000) in Africa. As in Ghana, urban persistence is not explained by road expansion post-independence.

The paucity of data for Africa does not allow us to precisely examine the channels of path dependence as we did for Ghana, but we use the information we have on the year of "connection" of each cell. Rail construction in colonial Africa can be separated into four episodes: 1890-1914, 1915-1929, 1930-1939 and 1940-1960. 50% of the cells were connected during the scramble for Africa (1890-1914), and World Wars I and II and the Great Depression all led to a multi-year fall in the number of connections, due to budget restrictions in European countries (see Online Appendix Figure 5). We use these events as natural experiments to compare the long-term effects of lines that were built relatively earlier than others. Whether a line is built in 1890-1914, 1915-1929 or 1930-1939 should not make a major difference in terms of historical factors. All these lines were built more than 60 years ago, and sunk capital is probably equally depreciated for all these periods now. We run the same model as in column (2) of Table 1 (thus controlling for roads in 2000, and including the controls listed in the notes below Table 1), and interact the rail dummy (0-20 km) with four dummies for each episode of rail building. The rail effect is much stronger for earlier periods (not shown): 4,850** before 1914, 4,153*** in 1915-1929, 2,271*** in 1930-1939 and -775 in 1940-1960. The effects disappear when controlling for urban population in 1960, as in column (3)

of Table 1. This suggests the following story, in line with the coordination failure hypothesis: the cells connected earlier became large cities at independence (1960), as they solved a coordination failure earlier, and increasing returns explain why these cities persisted post-1960. No effect is found for lines built post-1940. It could well be that these lines were different in terms of economic returns, or many of the coordination failures had already been solved by then.

African cities along the old railroad lines are also wealthier today. We run the same model as in column (2) of Table 1 (thus controlling for roads in 2000), except the outcome is our measure of average light intensity at the cell level in 2000-2001. When using *ln(light/area)* as the dependent variable, as in Henderson, Storeygard & Weil (2012), and controlling for *ln(urban population)* in 2000, the effect of the 0-20 km rail dummy is 0.5***, given a mean of 0.1 (not shown). Using an elasticity of light intensity with respect to GDP of 0.3 , we find that railroad cells are 15% wealthier per urban capita than other cells. When interacting the railroad dummy (0-20 km) with the four dummies for each episode of rail building, we find the following effects (not shown): 0.6*** (18% in terms of GDP per urban capita) in 1890-1914, 0.5*** (15%) in 1915-1929, 0.3 in 1930-1939 and 0.1 in 1940-1960. This indicates that the cells that were connected earlier are wealthier today, even when controlling for urban population in 2000. As in Ghana, railroad cities are wealthier than non-railroad cities of similar sizes, and this effect is stronger the earlier the cell was connected, and the earlier the city was created. Therefore, railroad cities must have accumulated more unobservable factors than non-railroad cities over time, as their earlier emergence gave them a clear initial advantage.

4. DISCUSSION

We now discuss the implications of our results for economic theory and public policy. First, transportation infrastructure is a man-made advantage that can launch a process of endogenous growth. Second, modern transportation technology can promote long-run economic change, but the effects will depend on the context.

4.1 Urban Persistence as "Growth-Enhancing Path Dependence"

In this section, we reinterpret the conceptual framework of Bleakley & Lin (2012) (see section VIII.B), and the literature on path dependence more generally, to describe how modern transportation technology produced economic change in Africa. They consider an economy with many locations characterized by increasing returns and congestion costs, which ensures that cities only grow until a certain point. Increasing returns can give rise to multiple spatial equilibria, as a dense location is likely to grow further. In this case, a temporary natural advantage has a persistent effect by creating a stimulus shifting local population density to a higher equilibrium. Our objective is different, as we want to see how a temporary man-made advantage may trigger a new equilibrium in which cities emerge to facilitate the accumulation of factors, where no cities existed before, and thus have long-term effects on economic growth. In other words, we do not compare two (urban) locations, but the "countryside" and a "city". Our analysis is dynamic too.[20]

[20] Our model only considers the countryside and one city, instead of multiple cities. The city serves as a proxy for the urban sector, i.e. the aggregate set of urban locations. Another way to interpret the model is that it characterizes a local economy separated from the rest of the country.

Figure 7 shows indirect utility V for a marginal mobile agent as a function of the size, X, of a city she could choose to live in. We define V* as the utility the agent can receive in the countryside. We posit an inverted-U shape relationship between utility V and city size X, as congestion costs dominate static increasing returns for higher urban densities. We assume that the utility-density relationship, in both the countryside and the city, also depends on the stock of technology in the economy – the "A" in the standard production function – as it makes everyone more productive. We implicitly assume that A has a disproportionate effect on larger cities. This A increases in two ways. First, an exogenous man-made advantage α, such as transportation infrastructure, raises overall productivity (and utility). Second, A depends on endogenous technological progress, which we make a function of city size X in the previous period. Dynamic increasing returns thus imply that the larger the city, the more technology is accumulated. The equilibrium is defined by the comparison of V* and V, the respective utilities for the countryside and the city.

Initially, at period 0, the economy is poor and rural. In other words, V_0* is higher than V_0, and there are no cities ($X = X_0$). In period 1, transportation infrastructure is built for exogenous reasons ($\alpha = 1$), and both V_1 and V_1* increase. If V_1 starts above V_1*, people migrate to the city, and city size increases until X_1. In period 2, the value of the man-made advantage remains the same ($\alpha = 1$), but the fact that a city was created in the previous period accelerates technological progress, in both the city and the countryside, and both V and V* shift upward. With V_2 higher than V_2*, even more people migrate to the city, and city size increases until X_2.[21] Now, what happens if the man-made advantage becomes obsolete ($\alpha = 0$)? Both V and V* shift downward. However, both V_3 and V_3* will be higher than V_0 and V_0* respectively, because urbanization produced endogenous technological progress after period 1. In other words, the temporary man-made advantage has triggered a process in which cities emerge, and the dynamic increasing returns permitted by rising urban densities caused the long-term effects on economic growth.

The literature on path dependence often compares how a spatial equilibrium can shift across two urban locations. Our context is thus different in that there was no urban equilibrium initially. Most of Africa was under-urbanized around 1900, while Ghana only had a few cities in 1901. By solving the coordination failure of contemporary factors, the temporary man-made advantage can launch a process of endogenous growth, that rapidly becomes independent of the initial advantage.

4.2 Implications for Transportation Technology and Development

In this paper, we use a natural experiment and a new data set on railroads and city growth at a fine spatial level over one century in Ghana, and Africa as a whole. Most railroad lines were built before independence, to serve the interests of the colonial powers. Although profitable, railroads fell largely out of use after independence. Yet they had long-term effects on urban and development patterns. While colonial sunk investments partly contributed to urban path dependence, evidence suggests that railroad cities persisted because their early emergence served as a mechanism

[21] The man-made advantage and technological progress will increase productivity (and utility) in both the countryside and the city. We do not specify the mechanisms by which this is happening. For example, we could imagine that transportation infrastructure leads to the commercialization and the modernization of agriculture. It could also lead to industrialization in the city, and the increase in manufacturing wages would attract rural workers, reduce rural labor supply, and raise rural wages.

to coordinate contemporary investments for each subsequent period. Railroad cities are also wealthier than non-railroad cities of similar sizes today. We use this setting to show how modern transportation technology can promote long-run economic change in poor countries. Our results have several implications.

First, transportation technology can produce economic change by reducing trade costs, integrating markets and facilitating the circulation of ideas. In poor and rural countries, transportation infrastructure and the cities it contributes to create can also serve as a mechanism to coordinate contemporary investments. Therefore, more observable factors (e.g., human capital) and unobservable factors (e.g., technology) can be accumulated, which may increase long-run productivity.

Second, the economic impact of a new transportation technology will depend on the previously used technologies. The less efficient the old technology is relative to the new one, the larger this impact will be. In colonial Sub-Saharan Africa, railroads were a modern transportation technology, compared to headloading, a very basic technology representative of the under-urbanization and underdevelopment of the continent pre-1900. They had large permanent effects as a result. Road networks were considerably expanded post-independence. However, the effects of these investments were smaller. Many of the spatial coordination failures had already been solved by then. These results suggest that there are decreasing marginal returns to new investments in transportation. What really matters is the initial transition from a mostly rural equilibrium to an urban equilibrium. Nevertheless, new investments could still have strong positive effects in poor, remote regions with high trade costs. They could permit increased commercialization of agriculture, start an urbanization process, and lay out the foundations of future industrialization. Moreover, compared to institutions that are difficult to change, transportation technology offers an easily implementable yet potentially powerful policy instrument.

Third, the effects of new investments in transportation infrastructure in poor, remote regions is likely to depend on their intrinsic economic potential. Railroads had strong effects in Southern Ghana because they opened new land to commercial agriculture. The same investments could have had different effects in the North of Ghana, where the land is more arid. As road building in Sub-Saharan Africa was often driven by political considerations (e.g., ethnic favoritism) instead of economic considerations in the past (Burgess et al., 2013), it could well be that there are still many poor and rural regions with a high economic potential and high trade costs. All in all, this suggests that the effects of investments in transportation infrastructure are conditional on the context in which they take place. Shifting an equilibrium across two urban locations may not have as strong and permanent effects as triggering an urban equilibrium when no such equilibrium existed before, and that is why modern transportation technology had a long-lasting impact on the continent.

GEORGE WASHINGTON UNIVERSITY
UNIVERSITY OF SUSSEX, AND CENTER FOR THE STUDY OF AFRICAN ECONOMIES, UNIVERSITY OF OXFORD

REFERENCES

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

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

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

Atack, Jeremy, and Robert Margo. 2011. "The Impact of Access to Rail Transportation on Agricultural Improvement: The American Midwest as a Test Case, 1850-1860." *The Journal of Transport and Land Use*, 4(2): 5–18.

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

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

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.

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

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

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

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

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.

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

Bosker, Maarten. 2013. "From Baghdad to London: Unravelling Urban Development in Europe and the Arab World 800-1800." *Review of Economics and Statistics, Forthcoming*.

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

Burgess, Robin, and Dave Donaldson. 2010. "Can Openness Mitigate the Effects of Weather Shocks? Evidence from India's Famine Era." *American Economic Review*, 100(2): 449–53.

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

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

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

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

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

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

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

Dell, Melissa. 2012. "Path Dependence in Development: Evidence from the Mexican Revolution." Unpublished manuscript, Harvard University, Department of Economics.

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

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

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

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

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

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

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

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.

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

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2013. "Urbanization with and without Industrialization." Unpublished manuscript, Department of International Development, University of Oxford.

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.

Harris, John R, and Michael P Todaro. 1970. "Migration, Unemployment & Development: A Two-Sector Analysis." *American Economic Review*, 60(1): 126–42.

Hawke, Gary Richard. 1970. *Railways and Economic Growth in England and Wales, 1840-1870*. Oxford: Oxford University Press.

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

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

Henderson, Vernon, Mark Roberts, and Adam Storeygard. 2013. "Is Urbanization in Sub-Saharan Africa Different?" Unpublished manuscript, Department of Economics, Brown Univerisity.

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

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

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

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

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

Lucas, Robert Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3–42.

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

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

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

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

Miguel, Edward, and Gérard Roland. 2011. "The Long-Run Impact of Bombing Vietnam." *Journal of Development Economics*, 96(1): 1–15.

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

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

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

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.

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.

Tsey, Christian. 1986. *Cold Coast Railways: The Making of A Colonial Economy 1879-1962*. Glasgow: PhD Thesis, University of Glasgow, Faculty of Social Sciences.

TABLE 1: COLONIAL RAILROADS AND CITY GROWTH FOR 39 SELECTED AFRICAN COUNTRIES (2000)

Dependent Variable:	**Urban Population in 2000 (Number of Inhabitants in Localities ≥ 10,000 Inh.)**					
Sample:	SSA (1)	SSA (2)	SSA (3)	Ghana (4)	Ghana (5)	Ghana (6)
Colonial Transportation:						
Dummy Rail 1960, 0-20 km	3,917***	3,162***	-361	3,060***	2,400**	-717
	(1,083)	(925)	(382)	-913	-1020	-465
Pre-Colonial Transportation:						
Dummy Coast, 0-20 km	-474	-464	-595	5,375*	3735	-343
	(1,463)	(1,413)	(596)	-2966	-2853	-1058
Distance to the Coast (km)	-0.5	-0.4	-0.1	1.4	0.3	-0.9
	(0.4)	(0.3)	(0.1)	-5.3	-4.8	-2.1
Dummy Navigable River, 0-20 km	295	71	-171	-355	-725*	-285
	(454)	(477)	(326)	-399	-400.7	-417
Post-Colonial Transportation:						
Dummy Paved Road 2000, 0-20 km		1,333***	483***		2,110***	1,252***
		(391)	(122)		-267	-411
Dummy Improved Road 2000, 0-20 km		1,040***	386***		634	333
		(356)	(101)		-579	-292
Path Dependence in Urban Patterns:						
Urban Population in 1960 (Inh.)			5.5***			3.4***
			(0.4)			-0.3
Country Fixed Effects	Y	Y	Y	N	N	N
Number of Observations (Cells)	194,000	194,000	194,000	2,091	2,091	2,091
(Adj.) R-squared	0.41	0.41	0.79	0.97	0.97	0.99

Notes: OLS cross-sectional regressions using urban population data on 194,000 0.1x0.1 degree (11x11km) cells *c* for 39 selected Sub-Saharan African (SSA) countries in 2000 (see Figure 1). The dependent variable, *Urban Population in 2000*, is the population of cell *c* that resides in localities above 10,000 inhabitants in 2000. *Dummy Rail 1960, 0-20 km* is an indicator variable whose value is one if cell *c* is within 20 km from a railroad line built before 1960, the year most countries became independent. *Dummy Coast, 0-20 km*, *Dummy Navigable River, 0-20 km*, *Dummy Paved Road 2000, 0-20 km*, and *Dummy Improved Road 2000, 0-20 km* are indicator variables whose value is one if cell *c* is within 20 km from the coast, a navigable river, a paved road in 2000 or an improved road in 2000 respectively. *Distance to the Coast (km)* is the Euclidean distance from cell *c* to the coast. *Urban Population in 1960* is the population of cell *c* that resides in localities above 10,000 inhabitants in 1960. Robust standard errors are reported in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Standard errors are corrected for spatial autocorrelation using the approach of Conley (1999), with a distance cut-off of 200 km. All regressions include country fixed effects and the following controls: three indicator variables whose value is one if the cell contains the largest city, the second largest city or the capital city of the country, Euclidean distance (km) to the largest city, mean and standard deviation of altitude (m), average annual rainfall in 1900-1960 (mm) and the shares of class 1, class 2, class 3, undetermined, sparsely vegetated and submerged soils in the cell. In columns (1)-(3), we use data for 39 countries. In columns (4)-(6), we use data for Ghana only. See Online Data Appendix for data sources.

TABLE 2: COLONIAL RAILROADS, COCOA PRODUCTION, AND POPULATION GROWTH (1901-1931)

Dependent Variable:	**Cocoa 1901-1927** (Tons Produced)	**Population 1901-1931** (Number of Inhabitants)			**Rural Pop. 1901-1931** (In Loc. < 1,000 Inh.)		**Urban Pop. 1901-1931** (In Loc. ≥ 1,000 Inh.)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dummy Rail 1918, 0-10 km	673***	3,615***	407	623	1,383***	324	2,232**	298
	(114)	(1,135)	(633)	(622)	(377)	(480)	(1,013)	(433)
Dummy Rail 1918, 10-20 km	511***	1415***	465	614	919**	323	495	291
	(111)	(511)	(503)	(482)	(378)	(329)	(306)	(310)
Dummy Rail 1918, 20-30 km	344***	800	154	251	760**	358	40	-107
	(100)	(611)	(575)	(556)	(359)	(350)	(380)	(342)
Dummy Rail 1918, 30-40 km	250**	488	14	116	209	-87	279	202
	(112)	(437)	(427)	(431)	(299)	(280)	(293)	(301)
Dummy Rail 1918, 40-50 km	59	226	83	113	-18	-92	244	206
	(93)	(392)	(389)	(382)	(240)	(-251)	(258)	(236)
Cocoa (Tons Produced)			1.76***	1.51***		1.15***		0.36
			(0.41)	(0.39)		(0.24)		(0.27)
Dummy Rail Station 1918			4,101*	1,529		735		793
			(2,099)	(1,753)		(590)		(1,525)
Cocoa at Rail St. (Tons) 1918				0.86**		-0.06		0.92**
				(0.44)		(0.06)		(0.43)
Cell FE and Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cell Controls	Y	Y	Y	Y	Y	Y	Y	Y
Number of Cells	554	554	554	554	554	554	554	554
Number of Observations	1,108	1,108	1,108	1,108	1,108	1,108	1,108	1,108
Adj. R-squared	0.61	0.73	0.77	0.81	0.70	0.74	0.67	0.80

Notes: OLS panel regressions using cocoa production and population data on 554 0.1x0.1 degree (11x11km) Ghanaian cells for the years 1901 and 1931. Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In column (1), the dependent variable is the production (tons) of cocoa of cell c in year t. In columns (2)-(4), the dependent variable is the population of cell c in year t. In columns (5) and (6), the dependent variable is the population of cell c that resides in localities whose population is superior to 1,000 inhabitants in year t. In columns (7) and (8), the dependent variable is the population of cell c that resides in localities whose population is inferior to 1,000 inhabitants in year t. All regressions include cell fixed effects (N = 554), year fixed effects (N = 2), and cell controls interacted with year dummies: 1931 mine dummy, cocoa production in 1901, urban and rural populations in 1901, share (%) of soils suitable / highly suitable / very highly suitable for cocoa cultivation, mean and standard deviation (m) of altitude, average annual rainfall (mm), and Euclidean distances (km) to Accra, Kumasi, Aburi, a port in 1901, a navigable river and the coast. The 554 cells are the Southern cells that are suitable for cocoa cultivation. See Online Data Appendix for data sources.

TABLE 3: ALTERNATIVE IDENTIFICATION STRATEGIES (1901-1931)

Regression:	*Main*	*West vs. East*		*District-Year FE*	*Control Group:*		*IV Straight Lines*
			Highly Suitable		Placebo Cells	C.Coast Kumasi	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Dependent Variable = Cocoa (Tons Produced) in 1901-1927							
Dummy Rail 1918, 0-20 km	560***	314***	712***	431***	775***	743***	596***
	(85)	(79)	(215)	(104)	(121)	(150)	(168)
Dummy Rail 1918, 20-40 km	277***	122*	321***	228***	408***	431***	44
	(79)	(72)	(119)	(82)	(106)	(147)	(311)
Rail 1918 x Eastern Line 0-20 km		824***	429				
		(228)	(308)				
Rail 1918 x Eastern Line 20-40 km		498**	162				
		(202)	(270)				
Panel B: Dependent Variable = Population (Number of Inhabitants) in 1901-1931							
Dummy Rail 1918, 0-20 km	2,052***	1,464**	1,855*	1,553***	2,142***	1,839**	2,770**
	(611)	(613)	(995)	(552)	(732)	(730)	(1,228)
Rail 1918 x Eastern Line 0-20 km		2,165	1,093				
		(1,398)	(1,598)				
Cell FE, Year FE, Cell Controls	Y	Y	Y	Y	Y	Y	Y
Number of Observations	1108	1108	708	1108	628	490	1108

Notes: OLS panel regressions using production and population data on 554 0.1x0.1 degree (11x11km) cells for the years 1901 and 1931. Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In Panel A, the dependent variable is the production of cocoa (tons) of cell c in year t. In Panel B, the dependent variable is the population of cell c in year t. In columns (2) and (3), we compare the effects for the Western and Eastern lines. In column (3), we restrict the sample to highly suitable cells for the comparison. In column (4), we include district-year fixed effects (N = 62 x 2 = 124). In columns (5) and (6), the control group is restricted to all placebo cells and cells along the Cape Coast-Kumasi placebo line (1873) respectively. In column (7), we instrument the rail dummies by the distance to the straight lines Sekondi-Kumasi and Accra-Kumasi. In Panel A, the two instruments are dummies equal to one if the minimal Euclidean distance of the cell centroid to the straight lines is 0-20 or 20-40 km. The IV F statistics are 48.8 (0-20 km) and 7.1 (20-40 km). In Panel B, the instrument is a dummy equal to one if this distance is 0-20 km. The IV F statistic is 56.6. All regressions include cell fixed effects (N = 554), year fixed effects (N = 2) and controls interacted with year dummies. The controls are described in the footnote of Table 2. The 554 cells are the Southern cells that are suitable for cocoa cultivation. See Online Data Appendix for data sources.

TABLE 4: COLONIAL RAILROADS AND URBAN GROWTH (1901-2000)

Dependent Variable:	**Urban Population (Z-Score), 2000**					
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy Rail 1918, 0-10 km	0.56**	0.48**	0.13	0.05	0.09	-0.00
	(0.22)	(0.21)	(0.14)	(0.13)	(0.09)	(0.08)
Dummy Rail 1918, 10-20 km	0.28***	0.21**	0.10	0.14*	0.13**	0.10**
	(0.10)	(0.10)	(0.09)	(0.08)	(0.06)	(0.05)
Dummy Rail 1918, 20-30 km	0.13	0.09	0.10	0.10	0.06	0.05
	(0.11)	(0.11)	(0.11)	(0.10)	(0.07)	(0.06)
Dummy Rail 1918, 30-40 km	0.09	0.08	0.03	0.06	0.05	0.03
	(0.10)	(0.09)	(0.07)	(0.07)	(0.07)	(0.06)
Dummy Rail 1918, 40-50 km	0.13	0.09	0.12	0.15*	0.18***	0.14***
	(0.11)	(0.11)	(0.08)	(0.08)	(0.06)	(0.05)
Urban Population (Z-Score), 1931			0.60***	0.45***	0.06	0.05
			(0.10)	(0.07)	(0.08)	(0.08)
Rural Population (Z-Score), 1931			0.16***	0.10*	0.00	-0.01
			(0.06)	(0.05)	(0.05)	(0.05)
Urban Population (Z-Score), 1960					0.66***	0.55***
					(0.09)	(0.10)
Rural Population (Z-Score), 1970					0.06*	0.07**
					(0.03)	(0.03)
Cell Controls	Y	Y	Y	Y	Y	Y
Controlling for: Roads (2000)	N	Y	Y	Y	Y	Y
Historical Factors (1931)	N	N	N	Y	Y	Y
Contemporary Factors (2000)	N	N	N	N	N	Y
Number of Observations (Cells)	552	552	552	552	552	547
R-Squared	0.49	0.52	0.70	0.74	0.83	0.85

Notes: OLS cross-sectional regressions using urban population data on 552 0.1x0.1 degree (11x11km) cells *c* for the year 2000. Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the population of cell *c* that resides in localities above 1,000 inhabitants in 2000. The variable is standardized using a standard score (Z-Score) to account for demographic growth post-1931. In columns (2)-(6), we include indicator variables whose value is one if cell *c* is crossed by a paved road, an improved road or an earthen road in 2000 respectively. In columns (4)-(6), we control for various historical factors in both 1901 and 1931: numbers of government schools and non-government schools in cell *c*, four indicator variables whose value is one if there is an European hospital or an African hospital in cell *c*, number of churches in cell *c*, and six indicator variables whose value is one if cell *c* if crossed by a class 1, class 2 or class 3 road respectively. In column (6), we control for the contemporary factors of Table (5). All regressions include controls at the cell level. The controls are described in the footnote of Table 2. We also use z-scores for urban population and rural population in 1901 and 1960-70. We use rural population in 1970 as cell rural population is missing for the year 1960. The sample is the same as in Table 2, except we drop the nodes Sekondi and Kumasi. Accra was not in the sample of 554 cells. We have 547 cells in columns (6): we use the 10% *Population and Housing Census* in 2000, and we do not have enough individual observations to correctly measure the contemporary factors for 5 cells. See Online Data Appendix for data sources.

TABLE 5: COLONIAL RAILROADS AND CONTEMPORARY FACTORS IN 2000

Dependent Variable:	Primary School (%)	JSS (%)	SSS (%)	Health Clinic (%)	Hospital (%)	Clean Water (%)	Paved or Improved Road	Post Office (%)	Solid Walls (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: ***Railroads in 1918 and contemporary factors***									
Dummy Rail 1918, 0-10 km	8.0**	12.4***	12.8**	10.4**	19.4***	6.6*	0.2**	9.3*	13.0***
	(3.9)	(4.0)	(5.1)	(4.8)	(4.9)	(3.6)	(0.1)	(4.9)	(3.3)
Dummy Rail 1918, 10-20 km	8.1**	10.1***	9.3**	6.0	10.9***	0.7	0.1*	6.5	4.0
	(3.2)	(3.3)	(4.2)	(4.0)	(3.5)	(2.5)	(0.0)	(4.4)	(2.5)
Dummy Rail 1918, 20-30 km	4.8	6.9**	2.0	0.7	7.3**	2.3	0.1	-2.2	1.1
	(2.9)	(3.3)	(3.7)	(4.0)	(3.3)	(2.7)	(0.1)	(3.9)	(2.2)
Dummy Rail 1918, 30-40 km	5.2*	6.2**	-0.1	-0.3	5.5*	3.1	0.1	1.4	2.8
	(2.7)	(3.1)	(3.5)	(4.0)	(3.1)	(2.3)	(0.1)	(4.2)	(2.2)
Dummy Rail 1918, 40-50 km	5.3*	6.5*	-0.2	1.6	4.8	4.8*	0.1	5.0	4.6*
	(2.9)	(3.3)	(3.8)	(4.2)	(3.3)	(2.6)	(0.1)	(4.4)	(2.8)
Panel B: ***Railroads in 1918 and contemporary factors, conditioned on contemporary population***									
Dummy Rail 1918, 0-10 km	2.2	5.5	6.1	3.8	10.8**	3.5	0.1	2.4	9.7***
	(4.3)	(4.4)	(4.9)	(5.0)	(4.4)	(3.3)	(0.1)	(4.9)	(3.0)
Dummy Rail 1918, 10-20 km	3.9	5.6	6.9*	3.0	6.5**	-0.5	0.1	3.9	3.2*
	(3.4)	(3.4)	(3.5)	(3.9)	(3.2)	(2.5)	(0.1)	(3.9)	(1.9)
Dummy Rail 1918, 20-30 km	1.6	3.7	1.9	-0.3	5.2*	2.2	0.0	-2.6	1.6
	(3.0)	(3.3)	(3.0)	(3.6)	(2.7)	(2.5)	(0.1)	(3.5)	(1.7)
Dummy Rail 1918, 30-40 km	3.4	4.3	-0.5	-1.2	4.0	2.9	0.0	0.9	2.9
	(2.7)	(3.0)	(2.9)	(3.6)	(2.4)	(2.3)	(0.1)	(3.7)	(1.8)
Dummy Rail 1918, 40-50 km	4.7*	5.6*	-2.3	-0.1	2.9	3.8*	0.1	3.0	3.4
	(2.8)	(3.2)	(3.1)	(3.8)	(2.7)	(2.3)	(0.1)	(3.8)	(2.2)
Mean (2000)	84.0	74.4	22.2	43.2	14.3	14.1	0.45	29.6	20.1
Observations	552	552	552	552	552	547	552	552	547

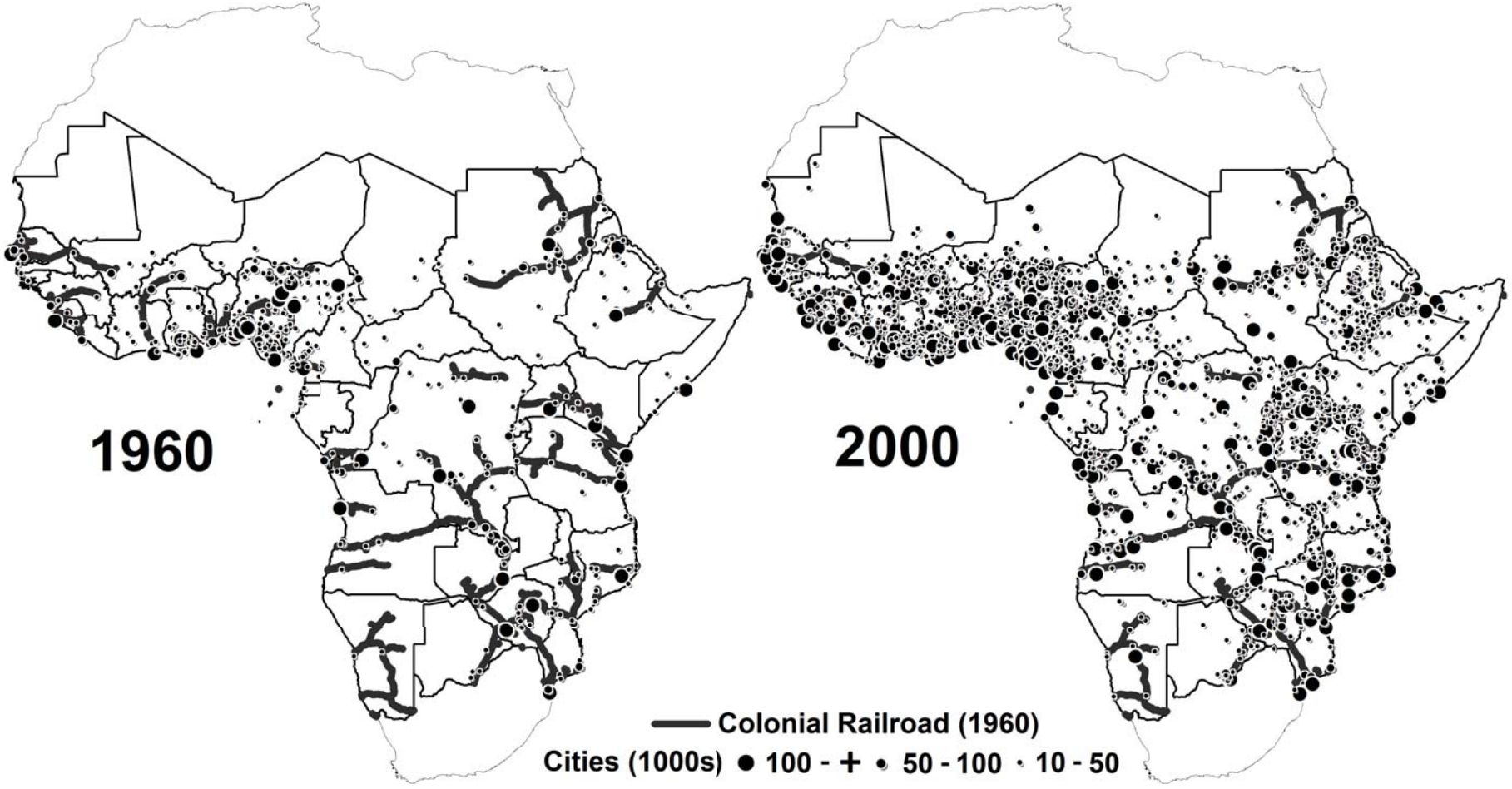
Notes: OLS cross-sectional regressions using data on 552 0.1x0.1 degree (11x11km) cells *c* for the year 2000. Robust standard errors are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In Panel A, we show the effects of the rail dummies on nine contemporary factors (2000). In Panel B, we control for urban population (pop. in localities $\geq$ 1,000 inh.) and rural population (pop. in localities $\leq$ 1,000 inh.) in 2000. Both variables are standardized to account for demographic growth post-1931. We use the following outcomes. Columns (1), (2), (3), (4), (5) and (8): share of inh. (%) living less than 5 km from a: primary school, junior secondary school (JSS), senior secondary school (SSS), health clinic, hospital or post office. Column (6): share of inh. (%) with access to clean water. Column (7): indicator variable whose value is one if cell *c* is crossed by a paved or improved road. Column (9): share of inh. (%) in a residence with solid walls. All regressions include controls at the cell level. The controls are described in the footnote of Table 2. We also use the z-scores for urban population and rural population in 1901. The sample is the same as in Table 2, except we drop the nodes Sekondi and Kumasi. Accra was not in the sample of 554 cells. We have 547 cells in columns (6) and (9): we use the 10% *Population and Housing Census* in 2000, and we do not have enough individual observations to correctly estimate these shares for 5 cells. See Online Data Appendix for data sources.

TABLE 6: COLONIAL RAILROADS AND ECONOMIC DEVELOPMENT IN 2000

Dependent Variable:	Lights (%)	Columns (2)-(8): Employment Share of (%)							Herfindahl Index
		Agri.	*Indu.*	*Serv.*	*Manuf.*	*Constr.*	*Trade*	*Admin.*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: ***Railroads in 1918 and contemporary measures of development***									
Dummy Rail 1918, 0-10 km	3.2***	-14.0***	5.4***	8.6***	2.7***	1.0***	6.1***	0.3**	-0.16***
	(0.8)	(2.7)	(1.1)	(1.8)	(0.8)	(0.2)	(1.1)	(0.1)	(0.03)
Dummy Rail 1918, 10-20 km	0.7**	-8.3***	3.4***	4.9***	2.1***	0.4**	3.6***	0.1	-0.10***
	(0.3)	(2.3)	(0.9)	(1.5)	(0.7)	(0.2)	(1.0)	(0.1)	(0.03)
Dummy Rail 1918, 20-30 km	-0.0	-6.0***	2.2***	3.9***	1.5**	0.3*	3.1***	0.0	-0.08***
	(0.20)	(2.0)	(0.8)	(1.3)	(0.6)	(0.2)	(0.9)	(0.1)	(0.03)
Dummy Rail 1918, 30-40 km	0.2	-3.4*	1.6**	1.7	0.9	0.2	1.6**	-0.1	-0.04*
	(0.2)	(1.9)	(0.8)	(1.2)	(0.6)	(0.2)	(0.7)	(0.1)	(0.02)
Dummy Rail 1918, 40-50 km	0.3	-4.9**	2.4***	2.5**	1.4**	0.2	1.7**	0.1	-0.06**
	(0.2)	(2.0)	(0.9)	(1.2)	(0.6)	(0.2)	(0.7)	(0.1)	(0.03)
Panel B: ***Railroads in 1918 and contemporary measures of development, conditioned on contemporary population***									
Dummy Rail 1918, 0-10 km	2.7***	-8.8***	3.9***	4.9***	1.5*	0.7***	4.0***	0.2	-0.11***
	(0.8)	(2.3)	(1.1)	(1.4)	(0.8)	(0.3)	(0.9)	(0.1)	(0.03)
Dummy Rail 1918, 10-20 km	0.4*	-6.0***	2.8***	3.1***	1.6***	0.3*	2.6***	0.0	-0.08***
	(0.2)	(1.8)	(0.8)	(1.1)	(0.6)	(0.2)	(0.8)	(0.1)	(0.02)
Dummy Rail 1918, 20-30 km	-0.1	-5.2***	2.1***	3.1***	1.4**	0.3	2.6***	-0.0	-0.07***
	(0.6)	(1.6)	(0.7)	(1.1)	(0.6)	(0.2)	(0.7)	(0.1)	(0.02)
Dummy Rail 1918, 30-40 km	0.2	-2.7*	1.5**	1.2	0.7	0.2	1.3**	-0.1	-0.03*
	(0.2)	(1.4)	(0.7)	(0.9)	(0.5)	(0.1)	(0.6)	(0.1)	(0.02)
Dummy Rail 1918, 40-50 km	0.1	-3.5**	1.9**	1.6*	1.1**	0.2	1.2**	0.0	-0.05**
	(0.2)	(1.5)	(0.8)	(0.9)	(0.5)	(0.1)	(0.5)	(0.1)	(0.02)
Mean	1.2	76.2	9.0	14.7	6.0	1.4	7.3	0.4	0.63
Observations	552	547	547	547	547	547	547	547	547

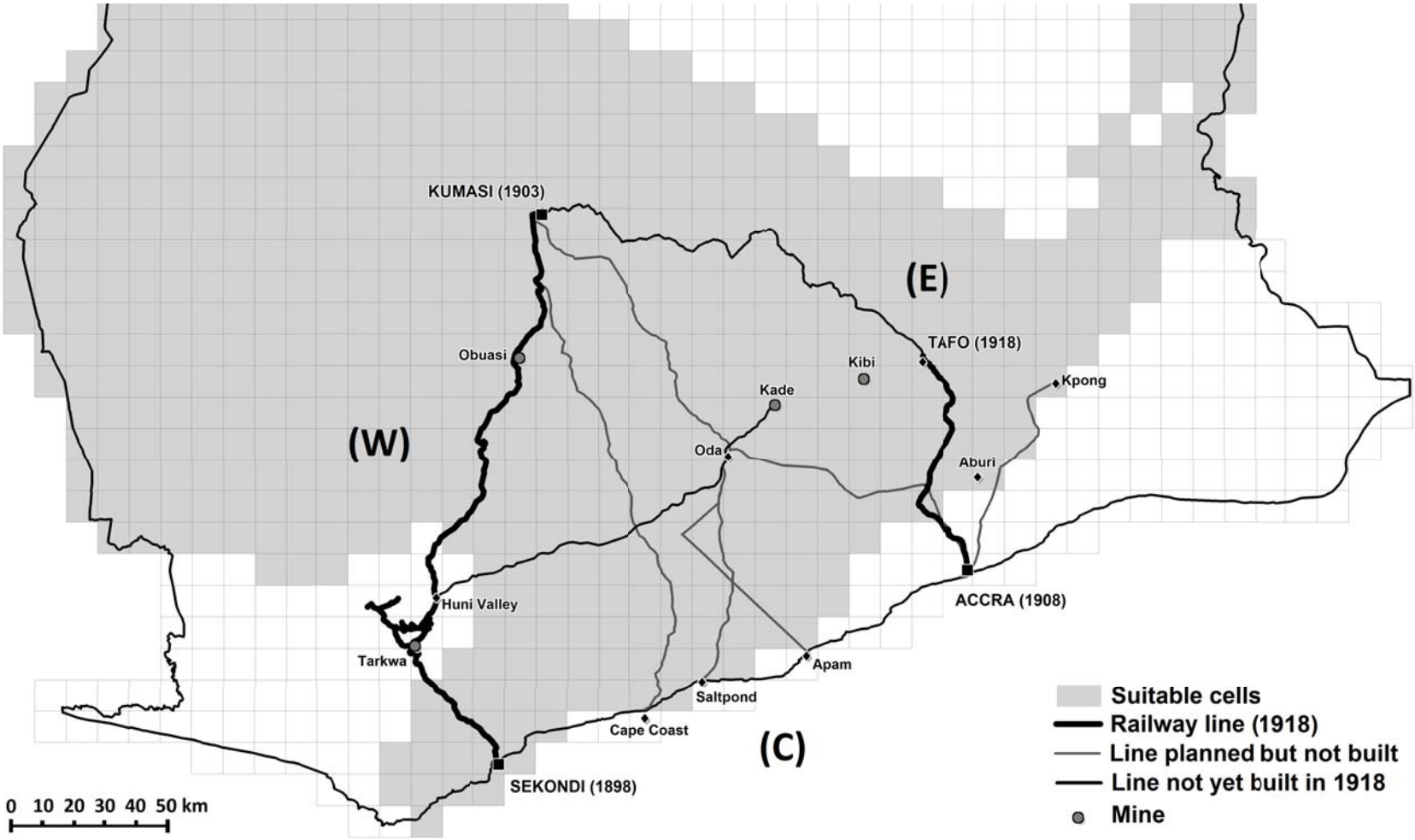
Notes: OLS cross-sectional regressions using data on 552 0.1x0.1 degree (11x11km) cells *c* for the year 2000. Robust standard errors are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. In Panel A, we show the effects of the rail dummies on nine contemporary measures of development (2000). In Panel B, we control for urban population (pop. in localities ≥ 1,000 inh.) and rural population (pop. in localities ≤ 1,000 inh.) in 2000. Both variables are standardized to account for demographic growth post-1931. We use the following outcomes. Column (1): share of cell area (%) for which a light is observed by satellite (1992-2010). Column (2)-(4): employment shares of (%) agriculture (*agri.*), industry (*indu.*) and services (*serv.*). Column (5)-(8): employment shares of selected industrial and service subsectors (%): "manufacturing" (*manuf.*), "construction" (*constr.*), "wholesale and retail trade" (*trade*) and "public administration" (*admin.*). Effects are small and/or not significant for other subsectors (with the exception of mining): "public utilities", "mining", "transport, storage and communication", "finance, insurance, real estate and business services", "education and health" and "other services". Column (9): Herfindahl Index using the 11 subsectors. All regressions include controls at the cell level (see the footnote of Table 2). The sample is the same as in Table 5. We have 547 cells in columns (2)-(9): we use the 10% *Population and Housing Census* in 2000, and we do not have enough individual observations to correctly estimate these shares for 5 cells. See Online Data Appendix for data sources.

Figure 1: Colonial Railroads and City Growth for 39 Selected African Countries in 1960-2000



Notes: This map shows colonial railroads and cities for 39 selected Sub-Saharan African countries in 1960 and 2000. 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 1960, when most African countries became independent. Cities are localities where population is superior to 10,000 inhabitants in 1960 (N = 436) and 2000 (N = 2,191). We do not have data for localities below the 10,000 population threshold, except for Ghana, our country of analysis. See Online Data Appendix for data sources.

Figure 2: Colonial Railroads, Placebo Lines and Cocoa Suitability in Southern Ghana



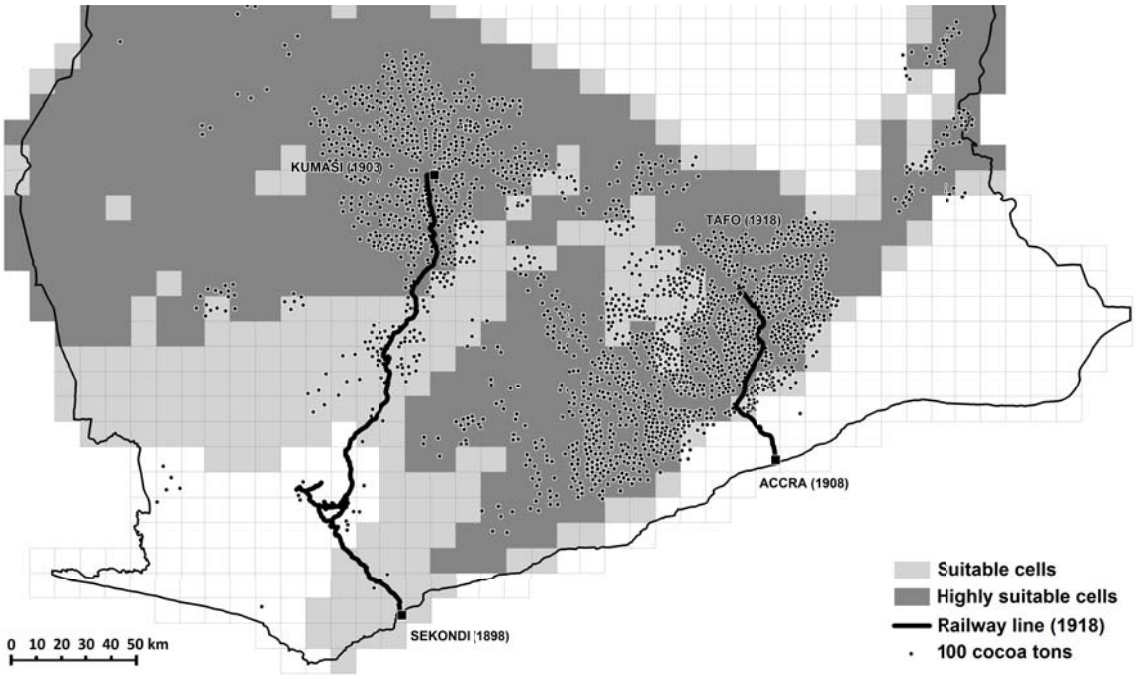
Notes: A 0.1x0.1 degree (about 11x11 km) cell is defined as suitable if it contains cocoa soils. The map displays the rail lines in 1918, the years line construction begun and finished, the main rail stations and the placebo lines: lines that were planned but not built (Cape Coast-Kumasi 1873, Saltpond-Kumasi 1893, Apam-Kumasi 1897, Accra-Kumasi 1897 and Accra-Kpong 1898) and lines that were not built early enough to affect cocoa production in 1927 (Tafo-Kumasi 1923 and Huni Valley-Kade 1927). (W), (C) and (E) show the western route, central route and eastern route respectively. The map also shows gold mines (Tarkwa, Obuasi and Kibi) and diamond fields (Kade). See Online Data Appendix for data sources.

Figure 3: Cocoa Production, Exports and Transportation, 1900-1927



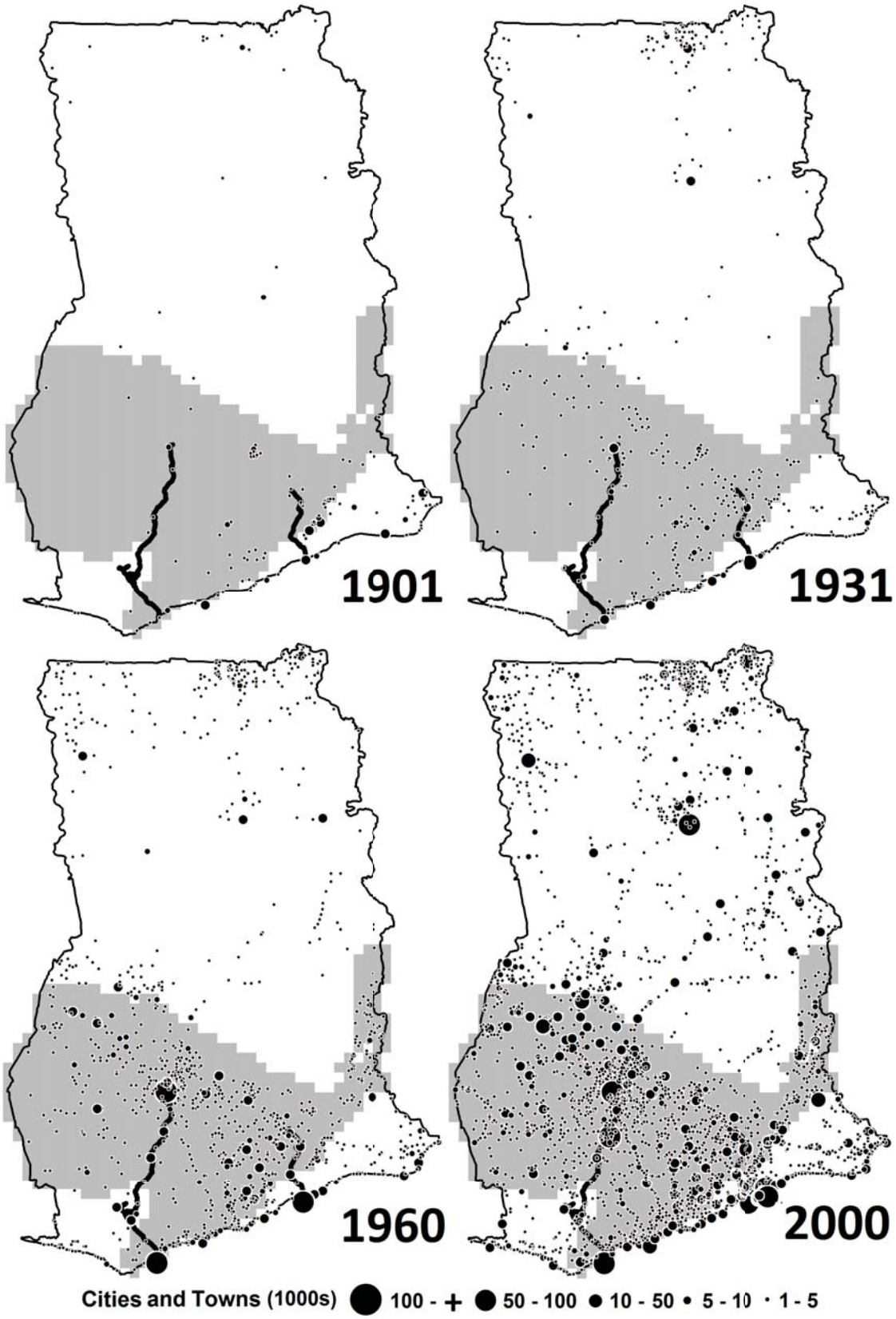
Notes: The figure displays three-year moving averages for cocoa production, cocoa tonnages transported by rail to a coastal port, and the share of cocoa exports of total exports from 1900 to 1927. See Online Data Appendix for data sources.

Figure 4: Railroads in 1918, Cocoa Suitability and Production in 1927



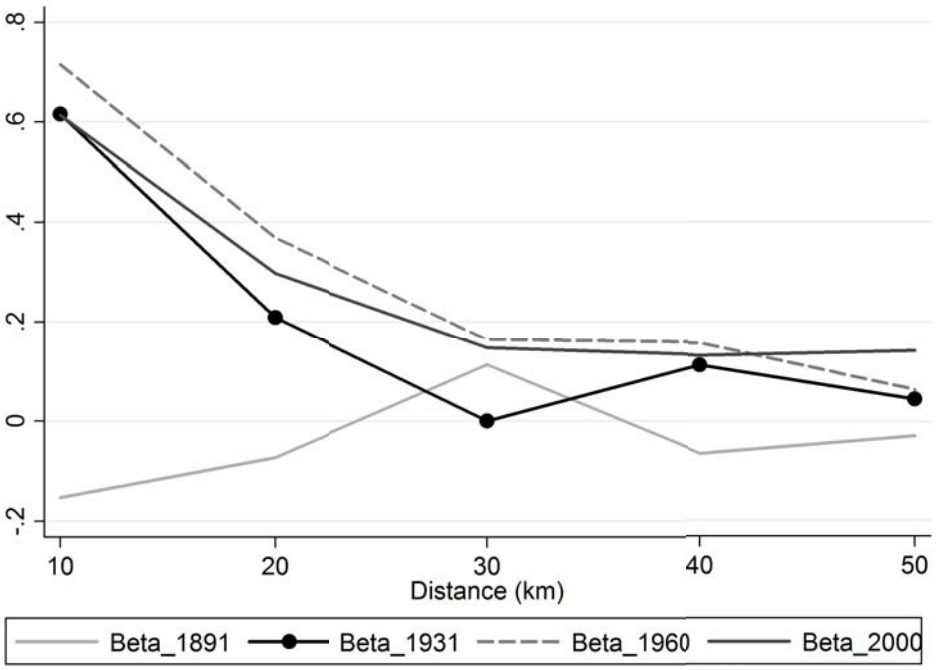
Notes: The map only shows Southern Ghana. A 0.1x0.1 degree (about 11x11 km) cell is defined as suitable if it contains cocoa soils, and highly suitable if more than 50% of its area consists of forest ochrosols, the best cocoa soils. The map displays the railroad lines in 1918, suitable cells, highly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

Figure 5: Railroads in 1918 and City Growth in 1901-2000



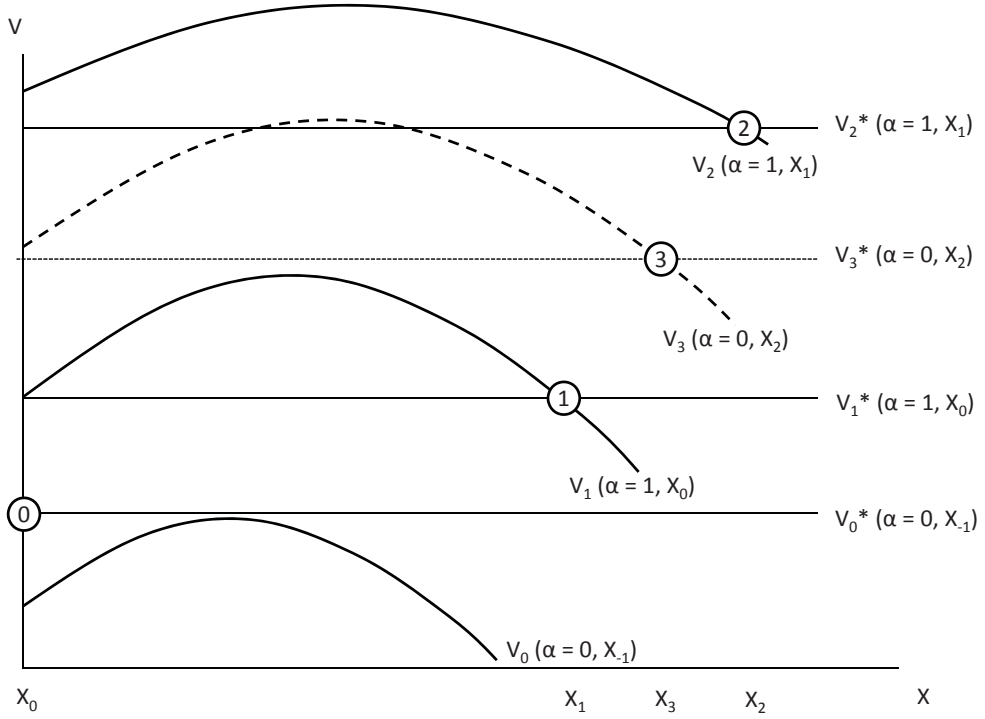
Notes: The maps display the railroad lines in 1918 and cities and towns in 1901-2000. Cities and towns are localities where population is superior to 1,000 inhabitants in 1901 (N = 143), 1931 (N = 438), 1960 (N = 1,100) and 2000 (N = 2,991). See Online Data Appendix for data sources.

Figure 6: Effects (Relative to 1901) of Railroads in 1918, 1891-2000



Notes: The graph displays estimates of Equation (3) for each distance threshold (km) = [0-10, 10-20, 20-30, 30-40, 40-50] and each year = [1891, 1931, 1960, 2000], using 1901 as the reference year. The dependent variable is the standard score of urban population. The effects are all significant at 5% until 20 km for the years 1931, 1960 and 2000. See Online Data Appendix for data sources.

Figure 7: Differences in Density with Dynamic Increasing Returns



Notes: The graph show indirect utility V for various locations. V* is the utility in the countryside. V is the utility in a city of size X. Congestion costs dominate static increasing returns for higher values of X. We posit that both V* and V depend on an exogenous man-made advantage α (e.g., transportation infrastructure), and endogenous technological progress, which we make a function of city size X in the previous period. Thus, there are thus dynamic increasing returns in the economy.

FOR ONLINE PUBLICATION: DATA SOURCES

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

Spatial Units for Ghana:
We assemble data for 2,091 grid cells of 0.1x0.1 degrees (11x11 km) from 1891 to 2000. We choose a high resolution grid because we have precise GIS data on railroads, agriculture and population in 1891-2000. Each cell has the same size, except those cells that are coastal or crossed by a border. We create dummies equal to one if the cell is coastal or bordering another country to control for this issue. The cells belong to 110 districts (2000).

Railway Data for Ghana:
We obtain the layout of railway lines in GIS from *Digital Chart of the World*. We use Gould (1960), Dickson (1968), Tsey (1986), and Luntinen (1996) to recreate the history of railway construction. For each line, we know when it was surveyed, planned, started, and finished and when each station was opened. From the same sources, we know lines that were built but not planned. Most of those placebo lines follow historical trade routes and became roads later. Using the GIS road network also available from *Digital Chart of the World*, we recreate those placebo lines in GIS. We calculate for each grid cell the Euclidean distance (km) from the cell centroid to each real or placebo line. Lastly, we create a set of cell dummies equal to one if the cell centroid is less than X km away from the line: 0-10, 10-20, 20-30, 30-40 and 40-50 km. We create a cell dummy equal to one if the grid cell contains a rail station in 1918. We also know how many tons of cocoa were brought to each station in 1918.[22] Data on railway traffic was obtained from various sources.[23] Lastly, we construct the instruments by using GIS to obtain the Euclidean distance (km) from each cell centroid to the straight lines Sekondi-Kumasi, Accra-Sekondi, Sekondi-Tarkwa-Obuasi and Accra-Kibi.

Commercial Agriculture Data for Ghana:
A very precise map of cocoa production in 1927 was obtained from the *1927 Yearbook of the Gold Coast* and digitized. This map displays dots for each 100 tons of cocoa production.[24] We then use GIS to reconstruct total cocoa production (tons) for each grid cell in 1927. Cocoa production was limited to 500 tons in 1901, and we know which grid cell they belonged to. We use Bateman (1965) to obtain the international and producer prices.

Population and Urban Data for Ghana:
We collect population data from the gazetteers of the *Population and Housing Censuses* 1891, 1901, 1931, 1960, 1970, and 2000. They list localities and their population size. Defining as a city any locality with more than 1,000 inhabitants, we obtain a geospatialized sample of 3,057 different cities for all these years. We used the GeoNet data base to retrieve the geographical coordinates of each city. Using GIS, we recalculate total urban population for each grid cell. We are then able to recreate rural population for each gridcell in 1901, 1931, 1970 and 2000. From the census gazetteers, we know the population size of each village (locality with less than 1,000 inhabitants). It was impossible to find the geographical coordinates of all of them. Yet, the 1901 census was exhaustively conducted and geospatialized in the South of Ghana (N = 756 cells). We know for each cell the number of *large towns*, *towns* (more than 500 inhabitants), *head chief towns*, *large villages* (100-500 inhabitants) and *villages* (less than 100 inhabitants). Using GIS, we can deduce for each cell the number of villages that are less than 100 inhabitants, the number of villages that have between 100 and 500 inhabitants and the number of villages that have between 500 and 1,000 inhabi-

[22] This information was retrieved from the *Administration Report of Gold Coast Railways (1921)*.

[23] These sources are: *Government Gazette Supplement of Accra* 1914, *Administration Report of Gold Coast Railways* 1920, 1921, 1929-30 and 1931-32, Gould (1960), Tsey (1986) and Luntinen (1996).

[24] Aggregating all the dots, we obtain 209,100 tons of cocoa production in 1927, which is very comparable with what we find from national estimates (see Gunnarsson 1978).

tants. From the census, we know the average settlement size for each category and we can reconstruct total rural population for each cell in 1901. For 1931, we have a map of the distribution of population for the whole country.[25] This map displays at a very fine spatial level settlements that have less than 500 inhabitants and settlements that have between 500 and 1,000 inhabitants. From the census, we know the average settlement size for each category, and we can reconstruct total rural population for each cell in 1931. We use the 2000 Facility Census which has population data on all settlements in 1970 and 2000. We know the enumeration area the rural locality belongs to. As we have a GIS map of all enumeration areas (N = 12,556) in 1970-2000, we simply reallocated all the villages to the different grid cells, and estimated total rural population in 1970 and 2000. The boundaries of the enumeration areas were significantly modified between 1960 and 1970, and we do not have a GIS map of them in 1960. Thus, we cannot estimate rural population in 1960, and use rural population in 1970 instead. Another issue is that some locations have a total or urban population of 0 for the early years. Some cells were historically sparsely inhabited tropical forest. Besides, not all cells have a locality above 1,000 inhabitants. When using logs, we add one inhabitant to each cell to ensure they are not dropped from the analysis.

Mining Production Data for Ghana:
We have the geographical coordinates of each mine (gold, manganese, and diamonds) in 1901 and 1931. We thus create a dummy equal to one if the cell contains a mine.[26]

Geographical and Economic Geography Data for Ghana:
Forest data comes from GIS data compiled by Globcover (2009). The data displays those areas with virgin forest or mixed virgin forest/croplands, which were areas with virgin forest before they were cleared for cash crop production. Soil data comes from the *1958 Survey of Ghana Classification Map of Cocoa Soils for Southern Ghana*. The map was digitized in GIS, and we calculated for each cell the share of land that is suitable for cultivation. We know the respective shares of land which consists of ochrosols (first class, second class, third class, unsuitable), oxysols, and intergrades. A cell is defined as suitable if it contains cocoa soils. It is then highly suitable if more than 50% of its area consists of forest ochrosols, the best soils for cultivation. It is very highly suitable if more than 50% of its area consists of class 1 and class 2 ochrosols. Although the map is for 1958, it is a good indication of cocoa suitability in 1901, as soil types are time-invariant. Climate data comes from *Terrestrial Air Temperature and Precipitation: 1900-2007 Gridded Monthly Time Series, Version 1.01*, 2007, University of Delaware. We estimate for each cell average annual precipitations (mms) in 1900-1960. Topography comes from SRTM3 data. We estimate for each cell the mean and standard deviation of altitude (meters). For each cell, we use GIS to get the Euclidean distances (km) to Accra, the capital city, Kumasi, the largest hinterland city, Aburi, the city where cocoa production originated, a port in 1901, a navigable river and the coast.

Other Transportation Networks Data for Ghana:
Transportation networks in 1901 are obtained from Gould (1960) and Dickson (1968), and *Colonial Annual Reports of the Gold Coast* 1903, 1904 and 1907.[27] We use various sources to reconstruct a GIS database of roads in 1901 and 1931: Gould (1960) and *Map of The Gold Coast with Togoland Under British Mandate*, published in 1930. Those road maps have a consistent legend showing class 1 roads ("*roads suitable for motor traffic throughout the year*"), class 2 roads ("*roads suitable for motor traffic but occasionally closed*"), and class 3 roads ("*roads suitable for motor traffic in dry season only*"). Other roads are not suitable for motor traffic and are not considered here. We use Michelin paper maps to recreate the 1965

[25] The map was obtained from the 1960 *Ghana Population Atlas*.

[26] Mining data is collected from the following documents: *The Mineral Industry of the British Empire and Foreign Countries 1913-1919*; *Reports of the Mines Department of the Gold Coast* 1931-1958.

[27] We use these sources, Maxwell (1923) and Tsey (1986) to obtain the trade cost associated with each transportation technology in 1900. Cardinall (1932) is our main source for production costs.

and 1998 road networks in GIS (which we use as proxies for 1960 and 2000 respectively), distinguishing paved (bitumenized), improved (laterite), and earthen roads.

Non-Transportation Infrastructure Data for Ghana:
We have data on government and non-government schools (missions) in 1902 (which we use as an approximation for 1901) and 1930-31. The data was compiled from education reports: *Report on the Education Department of the Gold Coast* for the years 1902 and 1930-31. They list all the schools in the country, which we then geospatialized. We also have data on European and African hospitals in 1902 and 1930. The data was compiled from health reports: *Report on the Medical and Sanitary Department of the Gold Coast* for the years 1902 and 1929-30. Data on the locations in churches in 1902 and 1929-30 comes from the *Ecclesiastical Returns* of the *Blue Books* of the same years. Besides, we use the 2000 *Facility Census* and the 2000 *Population and Housing Census* to recreate data on infrastructure provision at the cell level in 2000: share of inhabitants (%) living less than 5 km from a primary school, junior secondary school (JSS), senior secondary school (SSS), health clinic, hospital, or post office, and share of inhabitants (%) with access to clean water.

Economic Development Data for Ghana:
Data on urban employment in 1931 comes from the *Population and Housing Census*. We then obtain the share of cell area (%) for which a light is observed by satellite at least once in 1992-2010. The source of the satellite data on night lights is NOOA (2012). We follow the approach of Henderson, Storeygard and Weil (2012) and estimate average light intensity for each cell, for the year 2000-01. Second, we use the 10% sample of the *Population and Housing Census* to recreate data on employment at the cell level in 2000. We estimate the employment shares of 3 aggregate sectors "agriculture", "industry", and "services", and 11 subsectors: "agriculture", "mining", "public utilities", "construction", "manufacturing", "wholesale and retail trade, hotels and restaurants", "transport, storage and communications", "finance, insurance, real estate and business services", "education and health", "government services" and "other services". Since we only have data for 10% of the whole census, the most rural cells of our sample do not have enough observations to correctly estimate these shares. Data is thus missing for them. We then use the employment shares of the 11 subsectors to compute the Herfindhal index of diversification in 2000.

Railway Data for Sub-Saharan Africa:
We assemble data for 194,000 grid cells of 0.1x0.1 degrees (11x11 km) in Sub-Saharan Africa from 1960 to 2000. We obtain the layout of rail lines in GIS from *Digital Chart of the World*. We use Wikipedia and various studies available on the internet to recreate the history of each line. We know when each line was built, whether it was before the country became independent, and the main motivations behind its construction. There are 66,491 km of railway lines in 2000, but 58,716 km were built before independence (57,872 km if we use the year 1960 instead). There are three groups of motivation: military domination (against natives or other colonial powers), mining, and commercial agriculture. Data on colonial budgets in French West Africa and Kenya was compiled by Huillery (2012) and Burgess et al. (2013). We also create dummies whose value is one if the cell is less than 20 km from the coast, a navigable river, a paved road or an improved road in 2000. We also estimate the Euclidean distance (km) from the cell centroid to the coast.

Urban and Satellite Data on Night Lights for Sub-Saharan Africa:
We obtain from various sources a geospatialized database of localities above 10,000 inhabitants for 39 Sub-Saharan African countries in 1960 and 2000. We exclude from the analysis four small island countries and four Southern African countries for which urban data could not be obtained. Africapolis (2010) reports the data for 15 West African countries, while Africapolis (2012) reports the data for 18 Central and East African countries. Data for the remaining 6 countries was obtained using Wikipedia (2013) and the reports of their *Population and Housing Censuses* for the closest years to 1960 and 2000: Botswana, Malawi,

Mozambique, Namibia, Zambia and Zimbabwe. After obtaining a list of all localities above 10,000 inhabitants, we use GeoNet to retrieve their geographical coordinates. Using GIS, we recalculate total urban population for each grid cell. The source of the satellite data on night lights is NOOA (2012). We follow the approach of Henderson, Storeygard and Weil (2012) and estimate average light intensity for each cell for the year 2000-2001.

Controlling Variables for Sub-Saharan Africa:

We create dummies whose value is one if the cell contains the largest city, the second largest city or the capital city of the country. We compute the Euclidean distance (km) to the largest city. Climate data comes from *Terrestrial Air Temperature and Precipitation: 1900-2007 Gridded Monthly Time Series, Version 1.01*, 2007, University of Delaware. We estimate for each grid cell average annual precipitations (mms) in 1900-1960. Topography comes from SRTM3 data. We estimate for each cell the mean and standard deviation of altitude (meters). Lastly, we use FAO (2011) to estimate for each cell the shares of of class 1 (the most suitable), class 2, class 3, undetermined, sparsely vegetated and submerged soils.

REFERENCES

Africapolis. 2010. *Africapolis I: Western Africa*. Paris: SEDET.

Africapolis. 2012. *Africapolis II: Central and Eastern Africa*. Paris: SEDET.

Bateman, Merril. 1965. *Cocoa in the Ghanaian Economy*. Ph.D. dissertation, MIT.

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

Cardinall, Allan W. 1932. *The Gold Coast, 1931: a Review of Conditions in the Gold Coast in 1931 as Compared with Those of 1921*. Government Printer: Accra.

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

FAO. 2001. *Permanent Crop and Arable Land of Sub-Saharan Africa*. Rome: Food and Agricultural Organization.

Globcover. 2009. *Globcover Regional: Africa*. Louvain: European Space Agency Ionia.

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

Gunnarsson, Christer. 1978. *The Gold Coast Cocoa Industry 1900-1939: Production, Prices and Structural Change*. Lund: Published Thesis.

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

Huillery, Elise. 2012. "The Black Man's Burden: The Cost of Colonization of French West Africa." Unpublished manuscript, Department of Economics, Sciences Po.

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

Maxwell, John. 1923. *Gold Coast Handbook 1923*. Accra: Government Printer.

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

Tsey, Christian E. 1986. *Gold Coast Railways: the Making of a Colonial Economy, 1879-1929*. PhD University of Glasgow, Glasgow.

Wikipedia. 2013. *Lists of Cities by Country*. San Francisco, CA: California.

FOR ONLINE PUBLICATION: APPENDIX TABLES

APPENDIX TABLE 1: OBSERVABLES FOR TREATED CELLS VS. CONTROL CELLS IN 1901

Variable of Interest:	**Dummy Rail 1918, 0-20 km**		
Group of Control Cells:	All Control Cells	Placebo Cells Only	Cape Coast - Kumasi Only
Dependent Variable:	(1)	(2)	(3)
Panel A: Economic Variables			
Mine dummy	0.02**	0.01	0.02
Cocoa production in 1901	9.6**	9.6	9.6
Panel B: Demographic Variables			
Urban population in 1901	249	-276	231
Rural population in 1901	647***	379**	245
Panel C: Physical Geography Variables			
Share soils suitable for cocoa (%)	-0.11***	-0.14***	-0.14**
Share soils highly suitable (%)	-0.21***	-0.26***	-0.32***
Share soils very highly suitable (%)	-0.01	-0.04	0.01
Altitude: mean (m)	-18.0*	-1.3	48.0***
Altitude: standard deviation (m)	0.4	1.5	18.4***
Average annual rainfall (mm)	-22.4	21.6	77.6**
Panel D: Economic Geography Variables			
Distance to Accra (km)	-42.1***	38.8***	31.6***
Distance to Kumasi (km)	-11.5**	2.1	-5.6
Distance to Aburi (km)	-39.4***	39.1***	18.8*
Distance to a port in 1901 (km)	-33.3***	5.0	29.2***
Distance to a navigable river (km)	11.9	40.2***	5.3
Distance to the coast (km)	-32.1***	5.9	28.6***
Number of Treated Cells:	104	104	104
Number of Control Cells:	450	152	44

Notes: OLS regressions using data on 16 outcomes for 554 cells in 1901. These are the main controlling variables we use in our empirical analysis. This table tests that the treated and control cells are not significantly different in terms of observable characteristics in 1901, for various groups of control cells. We regress each control variable on a dummy equal to one if the cell is less than 20 km from a 1918 railroad line. Robust standard errors (not reported): * p<0.10, ** p<0.05, *** p<0.01. There are 16 different regressions for each column. In column (1), all control cells are included (N = 450). In column (2), the control cells are the cells less than 20 km from a placebo line (N = 152). In Column (3), the control cells are the cells less than 20 km from the Cape Coast-Kumasi placebo line (1873; N = 44). See Online Data Appendix for data sources.

APPENDIX TABLE 2: EFFECTS FOR PLACEBO LINES (1901-1931)

Type of Placebo Line:	**All Lines**	**Planned But Never Built (From West to East)**					**Not Built Yet**	
Placebo Line:		C.Coast Kumasi 1873	Saltpond Kumasi 1893	Apam Kumasi 1897	Accra Kumasi 1897	Accra Kpong 1898	Tafo Kumasi 1923	H.Valley Kade 1927
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Dependent Variable = Cocoa (Tons Produced) in 1901-1927							
Dummy Placebo 1918, 0-20 km	13	-188*	-56	-94	302**	541	-13	-209*
	(106)	(108)	(108)	(106)	(122)	(357)	(126)	(112)
	Panel B: Dependent Variable = Cocoa (Tons Produced) in 1901-1927							
Dummy Placebo 1918, 0-20 km	-354***	-544***	-241**	-285***	-9.6	72	-278**	-245*
x Dummy Rail 1918, 0-20 km = 0	(90)	(101)	(110)	(108)	(137)	(422)	(129)	(139)
	Panel C: Dependent Variable = Cocoa (Tons Produced) in 1901-1927							
Dummy Placebo 1918, 0-20 km	-102	-369***	-109	-167	114	482	-70	-206
Drop if Dummy Rail 1918, 0-20 km = 0	(120)	(104)	(119)	(119)	(145)	(391)	(143)	(144)
	Panel D: Dependent Variable = Population (Number of Inhabitants) in 1901-1931							
Dummy Placebo 1918, 0-20 km	52	-133	-123	-172	655	889	789	-1,238***
	(500)	(534)	(522)	(534)	(781)	(2,330)	(741)	(450)
	Panel E: Dependent Variable = Population (Number of Inhabitants) in 1901-1931							
Dummy Placebo 1918, 0-20 km	-1,341**	-1,634***	-1,104*	-1,162**	-1,067*	445	-400	-1,663***
x Dummy Rail 1918, 0-20 km = 0	(582)	(508)	(578)	(561)	(584)	(3,706)	(606)	(511)
	Panel F: Dependent Variable = Population (Number of Inhabitants) in 1901-1931							
Dummy Placebo 1918, 0-20 km	117	-763	-226	-400	-561	3,620	389	-1,271***
Drop if Dummy Rail 1918, 0-20 km = 0	(520)	(472)	(500)	(495)	(545)	(2,977)	(681)	(471)
Cell FE, Year FE, Cell Controls	Y	Y	Y	Y	Y	Y	Y	Y
Number of Observations	1,108	1,108	1,108	1,108	1,108	1,108	1,108	1,108

Notes: OLS panel regressions using production and population data on 554 0.1x0.1 degree (11x11km) cells for the years 1901 and 1931. This table tests that there are no spurious effects for seven placebo lines in 1901-1931. In Panel A, B and C, the dependent variable is the production of cocoa (tons) of cell c in year t. In Panel D, E and F, the dependent variable is the population of cell c in year t. For each placebo line, we create an indicator variable whose value is one if cell c is less than 20 km from the line. In column (1), we consider all placebo lines. In columns (2)-(5), we consider placebo lines that were planned but not built. In columns (6) and (7), we consider lines that were not built early enough to affect cocoa production in 1927. Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. All regressions include cell fixed effects, year fixed effects and cell controls interacted with year dummies (see the footnote of Table 2). In Panels A and D, we compare the placebo cells with the non-placebo cells (i.e., the railroad cells and the other control cells). In Panels B and E, we compare the placebo cells that do not intersect with a 1918 railroad line (i.e., the placebo cells that are not within 20 km from a railroad line in 1918) with the other control cells. In Panels C and F, we drop the cells less than 20 km from a railroad line in 1918, in order to compare the placebo cells with the other control cells. See Online Data Appendix for data sources.

APPENDIX TABLE 3: ROBUSTNESS CHECKS (1901-1931)

Regression:	*Main*	*Distance To Rail Station*	*No Controls*	*No Railroad Nodes*	*No Neighbors of Nodes*	*Full Sample*	*Logs*	*Conley Standard Errors*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Dependent Variable = Cocoa (Tons Produced) in 1901-1927								
Dummy Rail 1918, 0-20 km	560***	575***	544***	562***	304***	427***	2.1***	560***
	(85)	(88)	(113)	(86)	(102)	(75)	(0.4)	(162)
Dummy Rail 1918, 20-40 km	277***	263***	373***	278***	284***	196***	1.0**	277**
	(79)	(81)	(94)	(79)	(79)	(67)	(0.4)	(106)
Panel B: Dependent Variable = Population (Number of Inhabitants) in 1901-1931								
Dummy Rail 1918, 0-20 km	2,052***	2,039***	1,939***	1,588***	1,553***	_	0.7*	2,052***
	(611)	(626)	(563)	(445)	(531)		(0.3)	(696)
Cell FE, Year FE, Cell Controls	Y	Y	Y, Y, N	Y	Y	Y	Y	Y
Number of Observations	1,108	1,108	1,108	1,102	1,044	4,182	1,108	1,108

Notes: OLS panel regressions using production and population data on 554 0.1x0.1 degree (about 11x11km) cells *c* for the years 1901 and 1931. This table shows various robustness checks for the main results of Tables 2 and 3. In Panel A, the dependent variable is the production of cocoa (tons) of cell *c* in year *t*. In Panel B, the dependent variable is the total population of cell *c* in year *t*. Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. All regressions include cell fixed effects, year fixed effects and cell controls interacted with year dummies (see the footnote of Table 2). In column (2), the rail dummies are generated using the distance to rail stations (km). In column (3), we drop the controls. In column (4), we drop the railroad nodes (Kumasi, Obuasi and Sekondi in our sample of 554 cells). In column (5), we drop the railroad nodes and all the cells neighboring a cell containing a railroad node, to account for spatial spillovers of the nodes. In column (6), we use the full sample, but data is not exhaustive for total population. However, the data set being exhaustive for urban population, we verify that the results of Table 2 hold when we use this alternative outcome. While the coefficient of the 0-10 km rail dummy is 2,232** when using the restricted sample of 554 cells (see column (7) of Table 2), it is 2,017* when using the full sample. In column (7), we use a log-linear functional form. In column (8), standard errors are corrected for spatial autocorrelation using the approach of Conley (1999), with a distance cut-off of 200 km. See Online Data Appendix for data sources.

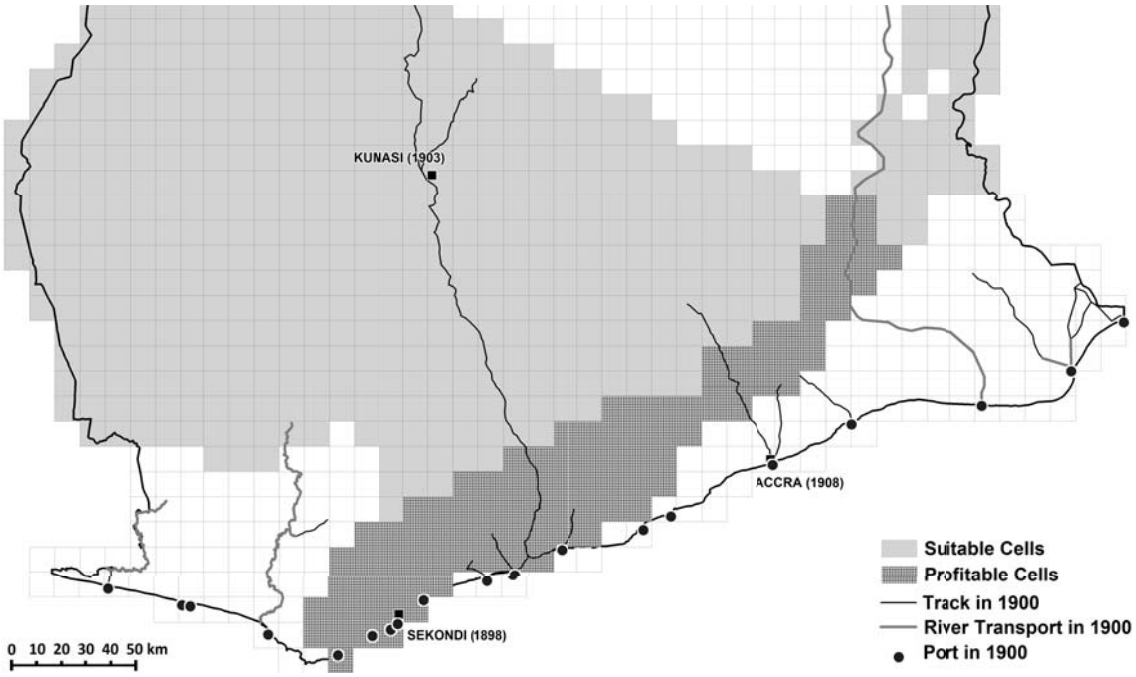
APPENDIX TABLE 4: COLONIAL RAILROADS AND HISTORICAL FACTORS (1901-1931)

Dependent Variable:	Number of Gvt Schools	Number of Non-Gvt Schools	European Hospital Dummy	African Hospital Dummy	Number of Churches	Class 1 Road Dummy	Class 2 Road Dummy	Class 3 Road Dummy	Class 1 or 2 Road Dummy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: ***Railroads in 1918 and historical factors (1901-1931)***									
Dummy Rail 1918, 0-10 km	0.03	0.68**	0.03	0.13*	0.51	0.13	0.25**	0.04	0.20*
	(0.05)	[0.343]	(0.03)	(0.07)	(0.58)	(0.11)	(0.11)	(0.16)	(0.12)
Dummy Rail 1918, 10-20 km	-0.02	0.18	0.00	-0.01	0.10	0.25**	0.27***	0.21	0.29***
	(0.02)	[0.147]	(0.01)	(0.02)	(0.55)	(0.10)	(0.10)	(0.14)	(0.10)
Dummy Rail 1918, 20-30 km	-0.00	0.05	0.00	0.00	0.22	0.21**	0.07	0.11	0.22**
	(0.03)	[0.130]	(0.00)	(0.03)	(0.43)	(0.10)	(0.10)	(0.12)	(0.10)
Dummy Rail 1918, 30-40 km	-0.01	0.10	0.00	-0.02	-0.02	0.07	0.03	-0.06	0.09
	(0.02)	[0.119]	(0.00)	(0.02)	(0.38)	(0.09)	(0.10)	(0.10)	(0.10)
Dummy Rail 1918, 40-50 km	0.01	0.04	0.00	-0.01	-0.20	0.06	0.06	-0.06	0.10
	(0.03)	[0.135]	(0.01)	(0.02)	(0.35)	(0.09)	(0.09)	(0.09)	(0.10)
Panel B: ***Railroads in 1918 and historical factors (1901-1931), conditioned on historical population (1901-1931)***									
Dummy Rail 1918, 0-10 km	-0.05	0.16	0.01	0.05	-0.29	0.11	0.17	0.05	0.12
	(0.04)	(0.24)	(0.01)	(0.05)	(0.57)	(0.12)	(0.11)	(0.16)	(0.12)
Dummy Rail 1918, 10-20 km	-0.05	-0.02	-0.00	-0.04*	-0.21	0.24**	0.24**	0.21	0.26**
	(0.03)	(0.16)	(0.01)	(0.02)	(0.55)	(0.10)	(0.10)	(0.14)	(0.10)
Dummy Rail 1918, 20-30 km	-0.02	-0.06	-0.00	-0.02	0.05	0.20**	0.05	0.11	0.20*
	(0.04)	(0.15)	(0.00)	(0.03)	(0.45)	(0.10)	(0.10)	(0.12)	(0.10)
Dummy Rail 1918, 30-40 km	-0.02	0.03	0.00	-0.03	-0.12	0.07	0.02	-0.06	0.08
	(0.02)	(0.13)	(0.00)	(0.02)	(0.37)	(0.09)	(0.09)	(0.10)	(0.10)
Dummy Rail 1918, 40-50 km	0.01	0.01	0.00	-0.02	-0.24	0.06	0.05	-0.06	0.09
	(0.03)	(0.12)	(0.00)	(0.02)	(0.31)	(0.09)	(0.09)	(0.09)	(0.10)
Cell FE, Year FE, Cell Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,108	1,108	1,108	1,108	1,108	1,108	1,108	1,108	1,108

Notes: OLS panel regressions using production and population data on 554 0.1x0.1 degree (about 11x11km) cells *c* for the years 1901 and 1931. This table shows the effects of the rail dummies on various measures of historical factors (1901-1931). In Panel A, we regress each measure of historical factors on the rail dummies. In Panel B, we control for total population (1901-1931). Robust standard errors clustered at the cell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. All regressions include cell fixed effects, year fixed effects and cell controls interacted with year dummies (see the footnote of Table 2). In columns (1) and (2), the dependent variable is the number of government ("Gvt") and non-government ("Non-Gvt") schools respectively. In columns (3) and (4), the dependent variable is an indicator variable whose value is one if there is an European hospital or an African hospital in cell *c* respectively. In column (5), the dependent variable is the number of churches. In columns (6)-(9), the dependent variable is an indicator variable whose value is one if cell *c* is crossed by a class 1 road, class 2 road, class 3 road, or a class 1 or class 2 road. See Online Data Appendix for data sources.

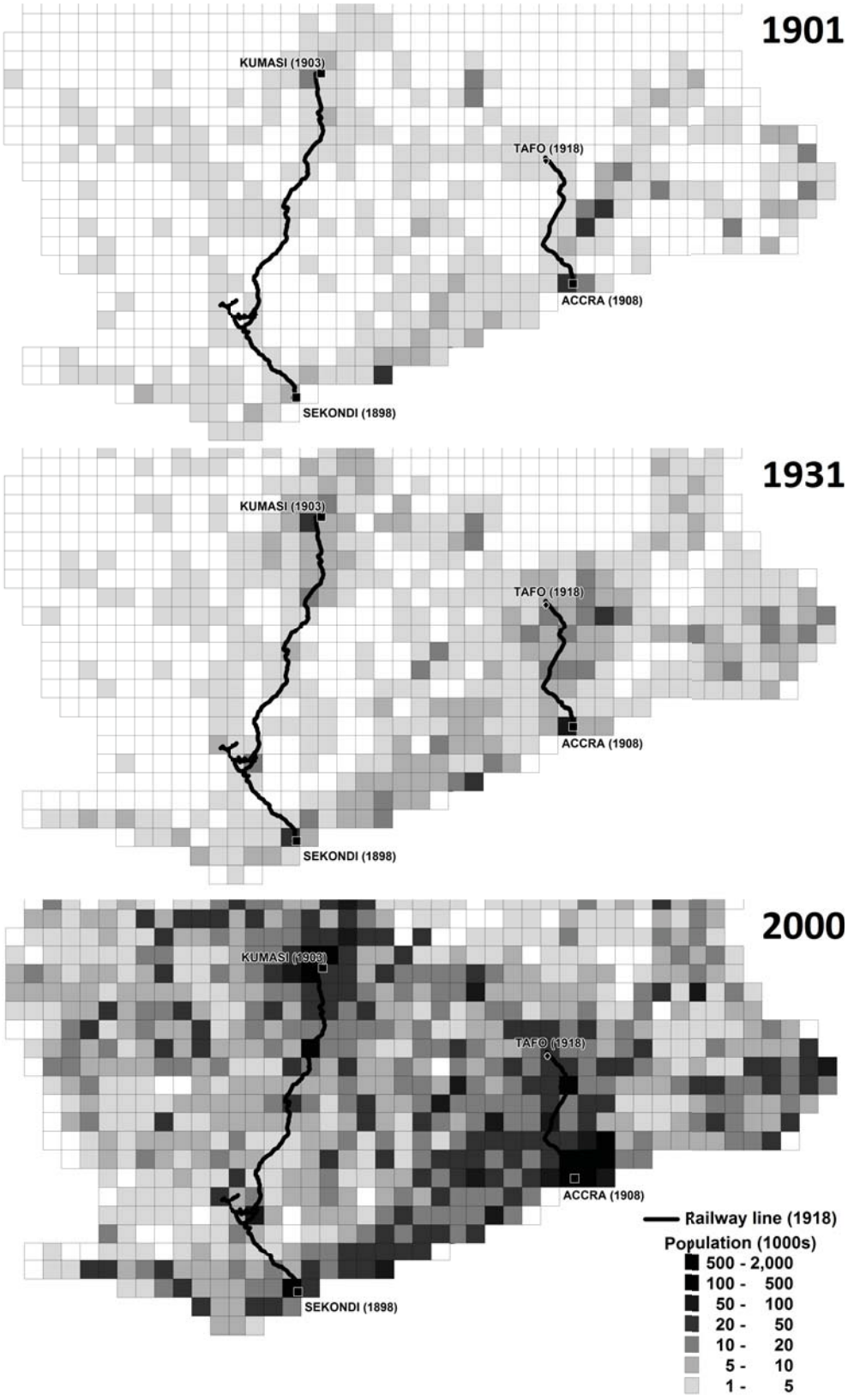
FOR ONLINE PUBLICATION: APPENDIX FIGURES

Appendix Figure 1: Transportation Networks in 1900 and
Area of Profitable Production Without Modern Transportation Technology.



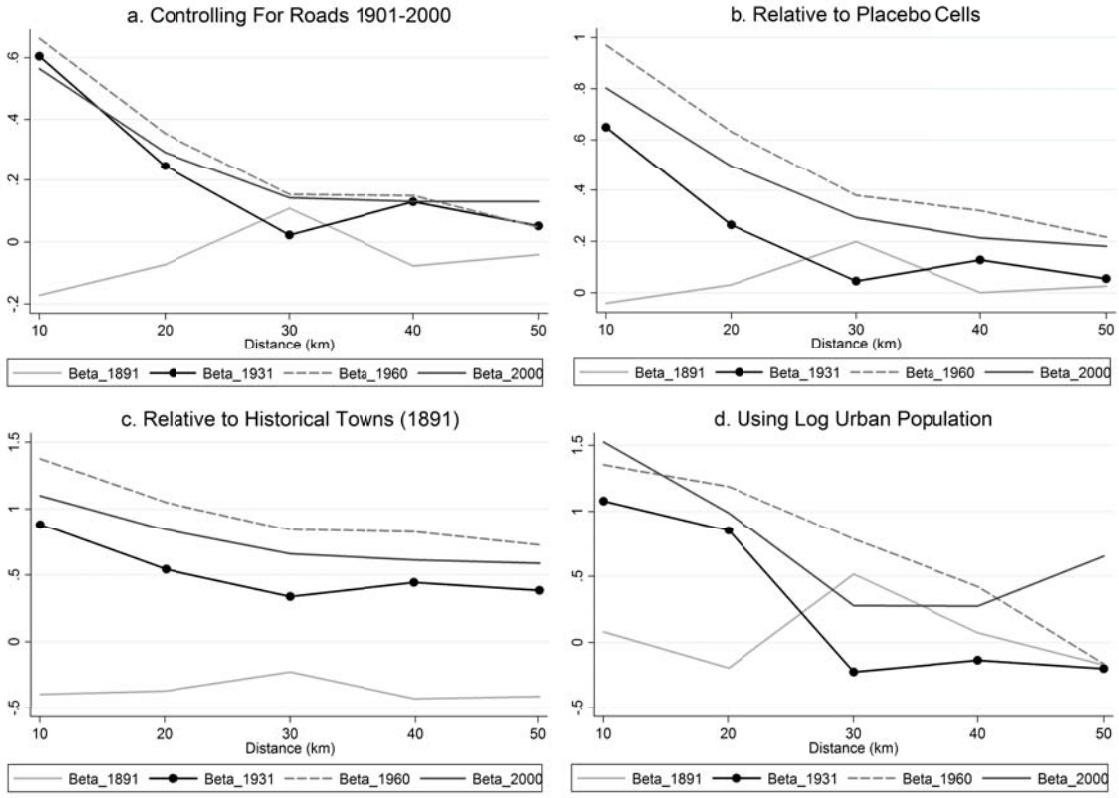
Notes: We want to know for which cells cocoa cultivation would have been profitable in 1927 if railroads (and roads) had not been built, i.e. what is the counterfactual distribution of cocoa production without modern transportation technology. The map shows transportation networks in 1900 (forest tracks, rivers and ports), the 0.1x0.1 degree (11x11 km) cells suitable for cultivation, and the cells for which cultivation is profitable in the counterfactual scenario. We estimate the following model for each cell in 1927 (per ton produced): Profit (1920s) = Producer price (1920s) - production cost (1920s) - transportation cost (1900). We consider the 1920s, as production in 1927 was determined by economic conditions five years before, in 1922, but also by the expectations of cocoa farmers for the coming decade. First, we use the average producer price offered at most ports in 1920-29. Second, we use data on production costs (purchase of forested land, clearing, planting, weeding, etc.) before 1930. Third, using the *Path Distance* function in ArcGIS we computed the lowest cost path from every cell to 19 coastal ports in 1900. We use data on trade costs given the available transportation technologies in 1900 and their associated cost per ton mile: head porterage on forest tracks (5 shillings, but adding a penalty for slopes), head porterage through the forest (8s), cask rolling for the few tracks on which casks could be rolled (1.9s), and canoe or steam launch for the cells along a river (1s-3.5s, depending on the river considered). The map shows the cells for which production was counter-factually profitable. We also verify that the rail permitted a dramatic expansion of the *feasible region*, by significantly reducing trade costs for most cells in the forest (0.25-0.75s, depending on the line and distance considered). See Online Data Appendix for data sources.

Appendix Figure 2: Railroads in 1918 and Population in 1901, 1931 and 2000.



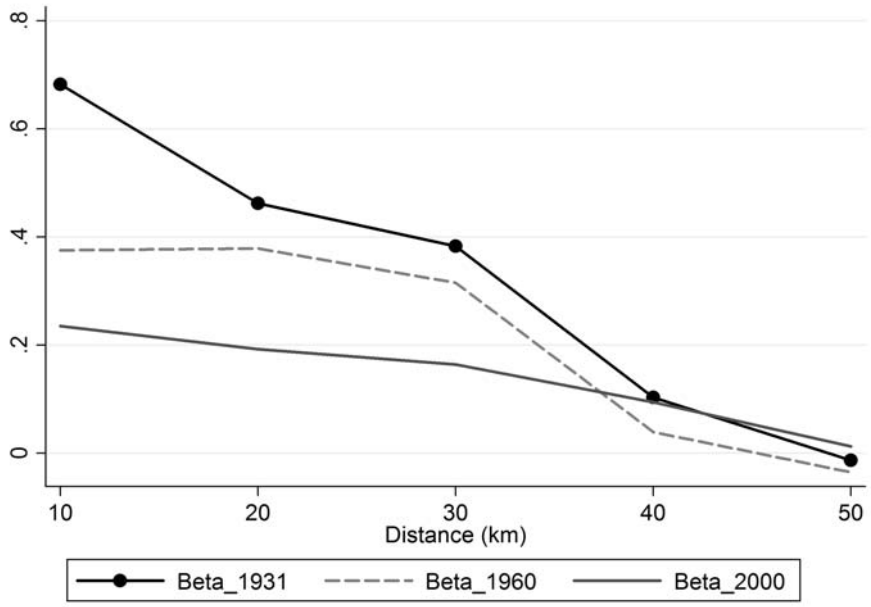
Notes: The map only shows Southern Ghana. The maps displays the total population of each 0.1x0.1 degree (about 11x11 km) cell in 1901, 1931 and 2000. Total population data is not available for the year 1960. Since all cells have the same area, total population is also a measure of population density. See Online Data Appendix for data sources.

Appendix Figure 3: Effects (Relative to 1901) of Railroads in 1918, Robustness and Specification Checks, 1891-2000



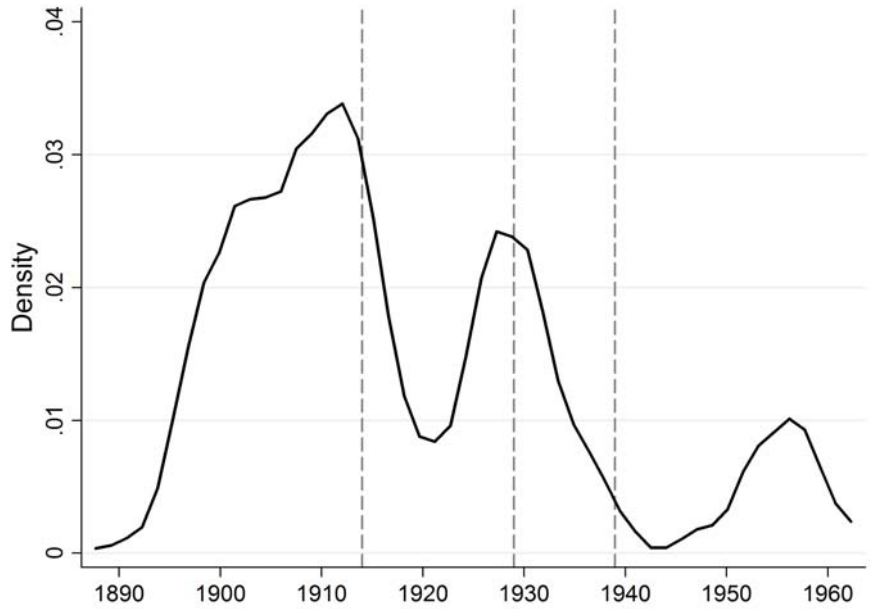
Notes: The graph displays estimates of Equation (3) for each distance threshold (km) = [0-10, 10-20, 20-30, 30-40, 40-50] and each year = [1891, 1931, 1960, 2000], using 1901 as the reference year. * p<0.10, ** p<0.05, *** p<0.01. The dependent variable is the standard score of urban population. In subfigure a., we control for road investments over time, by including the following variables: twelve indicator variables whose value is one if the cell is crossed by a class 1 road, class 2 road or class 3 road in 1901, or 1931, and a paved road, an improved road or an earthen road in 1960, or 2000. In subfigure b., the control group consists of placebo cells in 1918, using the lines planned but not built (see Fig. 2). In subgifure c., the control group consists of cells that already a town (a locality ≥ 1,000 in.) in 1891. In subifgure d., we use log urban population as the dependent variable, instead of the z-score of urban population. See Online Data Appendix for data sources.

Appendix Figure 4: Effects (Relative to 1901) of Railroads in 1918, Rural Population, 1901-2000



Notes: The graph displays estimates of Equation (3) for each distance threshold (km) = [0-10, 10-20, 20-30, 30-40, 40-50] and each year = [1931, 1970, 2000], using 1901 as the reference year. The dependent variable is the standard score of rural population. We use rural population in 1970 as a proxy for rural population in 1960, as rural population data is missing for the year 1960. * p<0.10, ** p<0.05, *** p<0.01. See Online Data Appendix for data sources.

Appendix Figure 5: Kernel Distribution of the Year of Connection for the Connected Cells of Colonial Sub-Saharan Africa, 1890-1960.



Notes: This graph shows the kernel distribution of the year of connection for each connected cell (0.1x0.1 degree, about 11x11 km) of 39 selected Sub-Saharan African countries during the colonial period 1890-1960. 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 1960, when most African countries became independent. 1890 is the first year a cell was reached by a railroad in our sample. 11,759 out of 194,000 cells were "connected" (here defined as being within 20 km of a railroad line) during the colonial period. The three vertical dashed lines represent the first year of World War I (1914), the Great Depression (1929) and World War II (1939) respectively. See Online Data Appendix for data sources.