Colonial Investments and Long-Term Development in Africa: Evidence from Ghanaian Railroads*

Remi Jedwab and Alexander Moradi

Abstract: What is the impact of colonial infrastructure investments on long-term development? Railroad construction in colonial Ghana provides a natural experiment. Using panel data at a fine spatial level over one century, we find strong effects of rail connectivity on cash crop production, population, and urban growth. We exploit various identification strategies to ensure our effects are causal. Colonial railroads had large welfare effects for Ghanaians before independence. They also had a persistent impact, despite a complete displacement by other transportation means. Capital accumulation and demographic growth account for path dependence. Colonial railroads thus shaped the economic geography of Ghana.

Keywords: Colonialism; Transportation Infrastructure; Multiple Spatial Equilibria, Sub-Saharan Africa
JEL classification: F54; O55; O18; R4; F1

---

*Remi Jedwab, Department of Economics, George Washington University and STICERD, London School of Economics (e-mail: jedwab@gwu.edu). Alexander Moradi, Department of Economics, University of Sussex and CSAE, University of Oxford (email: a.Moradi@sussex.ac.uk). We thank Robert Allen, Gareth Austin, Jean-Marie Baland, Sonia Bhalotra, Francois Bourguignon, Miriam Bruhn, Denis Cogneau, Paul Collier, Taryn Dinkelman, Robert Eastwood, Marcel Fafchamps, James Fenske, James Foster, Richard Hornbeck, Elise Huillery, Eliana La Ferrara, Molly Lipscomb, Stephan Litschig, Rocco Macchiavello, Karen Macours, William Masters, Stelios Michalopoulos, Branko Milanovic, Nathan Nunn, Henry Overman, Thomas Piketty, Jean-Philippe Platteau, Ilia Rainer, Ariell Reshef, Elizabeth Sadoulet, Sheetal Sekhri, Stephen Smith, Francis Teal, Pedro Vicente and seminar audiences at Bocconi (CEPR/AMID Development Conference), Oxford (CSAE and Economic History seminars, CSAE Conference), London School of Economics (STICERD and Economic History seminars, SERC Conference), GWU, RAND, University of Virginia, Dartmouth College (NEUDC), International Food Policy Research Institute, Paris School of Economics, Namur, Carlos Tercero, Sussex, Cambridge (RES Conference), Geneva (AEHW), Kent, University of Barcelona and Stellenbosch (WEHC) for helpful comments. We are grateful to 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.

1. INTRODUCTION

What is the impact of colonial infrastructure investments on long-term development? An extensive literature has studied the role of colonial institutions, relating economic outcomes today to the duration of colonization (Feyrer & Sacerdote 2009), the form of colonization (extraction versus settlement: Engerman & Sokoloff 2000, Acemoglu, Johnson & Robinson 2001, 2002, Dell 2010, Bruhn & Gallego 2012; direct versus indirect rule: Banerjee & Iyer 2005, Iyer 2010) or the identity of the colonizer (Porta et al. 1998; Bertocchi & Canova 2002). Fewer studies investigated channels other than institutions. Glaeser et al. (2004) argue that human capital mattered more. Banerjee & Iyer (2005) and Dell (2010) show that colonial institutions influenced public investments but they either look at current investments or they use proxies for colonial investments such as literacy and schooling. The effect of colonial investments in physical capital has been largely ignored.[1]

We look at rail construction in Africa. First, railroads were often the single largest expenditure item in colonial budgets. In Ghana, for example, railroad expenditures accounted for 31.4% of total public expenditure in 1898-1931.[2] Railroad construction represented two thirds of government development expenditures. Recurrent education expenditures in contrast amounted to a meagre 4%. Second, railroads are a canonical "colonial" transportation technology. 88.3% of total rail mileage in sub-Saharan Africa was built before independence. These lines served various purposes.[3] Military domination, against natives or other colonial powers, was given as motivation in 35.5% of the cases, mining and cash crop agriculture were mentioned in 36.0% and 42.4% of cases respectively. The lines were often intended to support European mining companies and farmers. Yet these lines also affected the economic lives of Africans by creating economic opportunities. Because infrastructure investments are highly localized, they shaped the economic geography of colonies, with long-term effects on development.

In this paper we study the impact of rail construction on development in colonial and post-colonial Ghana. Two lines were built by the British in 1898-1918 to link the coast to mining areas and the hinterland city of Kumasi. Using panel data at a fine spatial level over one century (11x11 km grid cells in 1891-2000), we find a strong causal effect of rail connectivity on the production of cocoa, the country's main export commodity, and

[1] Donaldson (2010), Burgess & Donaldson (2010), Burgess et al. (2011) and Chaves, Engerman & Robinson (2012) also analyze colonial investments in transportation infrastructure but they do not investigate their effects on long-term development.

[2] By comparison, in Kenya in 1896-1930 the share of railroad expenditure in total public expenditure was 19.3%. In French West Africa in 1910-1956, this share amounted to 30.0%.

[3] For each rail line in Africa, we know when it was built and the main motivation behind the construction of the line. More details can be found in Web Appendix 1.

development, which we proxy by population and urban growth.

First, we exploit various identification strategies. We document that the placement of the lines was exogenous to cocoa production and population growth. The *Western Line* connected the European gold mines to the coast and was extended to Kumasi for military domination. The *Eastern Line* connected the capital city of Accra to the network through Kumasi. First, we show that pre-railroad trade costs were prohibitively high limiting cocoa production to a narrow coastal strip. Second, we verify the parallel trend assumption using data before 1901. Third, having spatially refined data allows us to compare identical observations. Cells less than 50 km from the lines are all similar in terms of observables, yet the effects are much stronger just along the lines. Fourth, there are no effects for a set of placebo lines that were planned but not built. Fifth, 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. Sixth, cocoa was mentioned as one of the motivations of the Eastern Line, but we claim it was an ad hoc justification. This is confirmed by the fact we find the same effects for both lines. We also verify that our results are robust by employing two techniques used in the literature. We instrument for rail connectivity with straight lines between the two ports and the city of Kumasi, thus using the fact that being on a straight line between two large cities makes it more likely to be connected. In the spirit of nearest neighbor matching, we also only use placebo grids, those that would have been connected if the placebo lines had been built, as a control group.

Second, these investments had large welfare effects for Ghanaians during the colonial period. We find a strong effect of rail connectivity on cocoa production, population, and urban growth in 1901-1931. Trade costs decreased along the lines, which made cocoa production for export markets profitable. Rural population increased because cocoa cultivation required more labor in cocoa-producing villages. Urban population increased because producing villages used towns as trading stations, whether large railroad stations or small intermediary towns. We examine the welfare effects of rail construction. The social savings approach indicates that railroads account for 8.7% of GDP using cocoa only. As an alternative, we use our point estimate to show that railroads caused 44.5% of cocoa production in 1927. We find that two thirds of the surplus generated by cocoa production went to Ghanaians while the other third was captured by the colonizer. Colonization meant both extraction and development in this context.

Third, we show that these colonial investments had long-term effects on development. Railroads fell largely out of use in the 1970s, due to poor management, lack of maintenance, and competition from roads. Goods and passenger traffic collapsed after 1974, and railroads now transport one-third of what they did at independence. Moreover, Jedwab (2011) describes how cocoa production has disappeared from the old producing areas due to the shifting cultivation process characteristic of this crop. Although cells along the rail lines have lost their initial advantage in terms of both transportation and cocoa cultivation, we find that railroad cells are more urbanized, have larger manufacturing and service sectors, and better infrastructure today, showing evidence for path dependence. The dynamics of path dependence are studied using panel data on population growth in 1901-2000. We test several explanations and find that physical capital accumulation (structural change and contemporary infrastructure investments) and the demographic transition, and not sunk investments in 1931, account for this result.

Our focus on railroads also connects with the literature on transportation infrastructure, trade and development. This literature often mentions the conjunction of bad geography and poor infrastructure as the main obstacle to trade expansion in Africa (African Development Bank 2003, Buys, Deichmann & Wheeler 2010).[4] Despite this recent interest in transportation projects, little is known on their long-run effects, and more research is needed. The macroeconomic literature finds that better infrastructure diminishes trade costs and boosts exports (Radelet & Sachs 1998, Limão & Venables 2001). The micro-development literature shows that rural roads reduce poverty in connected villages by integrating labor and goods markets, thus providing new economic opportunities to their inhabitants (Jacoby 2000, Jacoby & Minten 2009, Mu & van de Walle 2011). Another strand studies the impact of large transportation projects, whether highways (Michaels 2008, Storeygard 2011) or railroads (Banerjee, Duflo & Qian 2012, Atack et al. 2010, Donaldson 2010, Burgess & Donaldson 2010, Atack & Margo 2011, Donaldson & Hornbeck 2011). They show there are significant gains from market integration.[5]

Our findings also advance the literature on the historical roots of African underdevelopment. Gennaioli & Rainer (2007) and Michalopoulos & Papaioannou (2012) emphasize the role of pre-colonial centralization. Nunn (2008), Nunn & Wantchekon (2011) and Nunn & Puga (2012) investigate the economic impact of slave trade, while Acemoglu, Johnson & Robinson (2001), Huillery (2009) and Michalopoulos & Papaioannou (2011) study the effects of colonial rule. This paper shows how colonial infrastructure in-

[4] In 2005, Africa had 0.002 km of railroad track per 1,000 sq km, and Ghana had around 4 km, while India had 21 km, the U.S. 25 km and Europe 51 km (World Bank 2010). Turning to roads, Africa had 85 km of roads per 1,000 sq km, and Ghana had 239 km, while India had 1,115 km, the U.S. 702 km and Europe 1,377 km.

[5] Those studies are more convincing in terms of identification. Michaels (2008), Atack et al. (2010) and Banerjee, Duflo & Qian (2012) use the fact that being on a straight line between two large cities makes it more likely to be connected to a highway or a railroad. Donaldson (2010) does not find any effect for rail lines that were approved but never built.

vestments one century ago shaped the economic geography of Ghana. This result is in line with the literature on the existence of spatial equilibria and the role of path dependence in the location of economic activity (e.g., Davis & Weinstein 2002, Redding, Sturm & Wolf 2011, Holmes & Lee 2012, and Bleakley & Lin 2012). Similarly to Bleakley & Lin (2012) who study portage sites in the U.S., we find no evidence that railroad cells, having lost their natural advantages, are in decline.

The paper is organized as follows. Section 2 presents the historical background of rail construction, cocoa and population growth in Ghana and the data used. Section 3 explains the methodology. Section 4 displays the results. Section 5 discusses the welfare effects during the colonial period. Section 6 studies why these effects persist over time. Section 7 concludes.

2. BACKGROUND AND DATA

We discuss the historical background and the data we use in our analysis. Web Appendix 1 contains more details on how we construct the data.

2.1 New Data on Ghana, 1891-2000

In order to analyze the effect of railroad construction on development in Ghana, we construct a new data set of 2091 grid cells of 0.1x0.1 degrees (11x11 km) from 1891 to 2000. We choose a high resolution grid since we have very precise GIS data on railroads, cocoa production, and population. 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 started and finished and when each station was reached and opened. We also located lines that were planned but not built. For each line, we create cell 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 transportation networks in 1901 (rivers and forest tracks) and motor roads in 1922.[6]

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

[6]There are three types of roads: class I (*roads suitable for motor traffic throughout the year*), class II (*roads suitable for motor traffic but occasionally closed*) and class III (*roads suitable for motor traffic in dry season only*).

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.

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.[7] 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, we only have georeferenced population data for Southern Ghana in 1901 and the whole territory in 1931, 1970, 1984 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 and human capital for each cell in 2000.

2.2 The Railroad Age

2.2.1 Railroads Built

Infrastructure investments are typically endogenous, 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 effect of reduced trade costs on development. This summary draws on Gould (1960), Tsey (1986) and Luntinen (1996). Figure 1 shows the geographic location of the mentioned lines.

The British established the Gold Coast colony in the south and extended their domination to what is now Ghana in 1896. Improving transport infrastructure was on the agenda, to permit military domination and boost trade historically constrained by high transport costs. Draft animals were not used due to the Tsetse fly transmitting trypanosomiasis. 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 in southern Ghana, there were only a few well-cleared tracks. Railroads were the transport technology of the time, but the British had to choose between a western, central or eastern route.

The first line followed the western route. Strong interest groups of British capitalists lobbied to connect the gold fields in the hinterland. Mines needed heavy machinery and large quantities of firewood (or coal). Headloading

[7] The number of 1,000 inhabitant localities was 144 in 1891, 143 in 1901, 353 in 1921, 437 in 1931, 627 in 1948, 1,110 in 1960, 1,262 in 1970, 1,895 in 1984, and 2,975 in 2000.

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). Web Appendix 2 describes how the Governorship of William E. Maxwell (1895-97) was instrumental in the decision-making process. Maxwell previously worked in the Malay States where railroads served the tin mines and thought that the same model of "mining first" should be applied to 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 gold mines of Tarkwa and Obuasi in May 1901 and December 1902 respectively (see Fig. 2). The line was further extended to Kumasi in September 1903. Much of the line went through virgin forest and was opened to traffic in 1904. This line was built by Europeans for Europeans, and gold mining accounted for two thirds of the line's traffic (in volume) in 1904-1912. Cocoa did not play any role in the location choice, but the line had a strong effect on its 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. Colonial governors long favored a central route (e.g., from Saltpond or Apam, see Fig. 2), but a series of unexpected events led to the governorship of J. P. 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 around Abetifi (see Fig. 2). Construction started in 1909 and the line reached Mangoase in late 1909. However, serious flooding in 1910 and 1911 meant that the line was not opened to traffic before 1912. It was extended to Koforidua in 1915 and Tafo in 1916, but Tafo station was only opened in July 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 2.3 below). The province's production was already growing before Mangoase was reached in 1909: around 1,000 tons in 1901, 5,000 tons in 1905, 15,000 tons in 1910, 40,000 tons in 1915, 65,000 tons in 1920 and 100,000 tons in 1925. As cocoa trees take five years to produce, production before 1914 (ca. 30,000 tons) cannot be attributed to the line.[8] Growing in advance could

[8]The line was officially opened to traffic in 1912, but it reached Mangoase three years before in 1909. There is some evidence that farmers went there as soon as 1909 to grow

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 spatial diffusion of cocoa from Aburi, (iv) no positive effects are found for placebo lines, and (v) results are robust to instrumentation or matching.

2.2.2 Reduction in Transport Costs

Railroads permitted a massive decrease in transportation costs. While the freight rate per ton mile was 5s 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. Yet this simple comparison of freight rates underestimates the magnitude of transport costs: (i) the cost above only concerns headloaders that walk along a forest track. There were only a few well-cleared tracks and headloaders often had to go through the dense tropical forest, which made it even more costly, (ii) cask rolling necessitated good quality roads and there were only a few of them then, (iii) Ghana lacked navigable waterways and these did not serve the areas where cocoa could be grown, and (iv) roads were of poor quality until 1924 when the government started the "Tarmet Program" which made roads suitable for motor traffic throughout the year (Gould 1960). Until the late 1920s, rail was by far the best transport technology.

Figure 2 displays the cells where cocoa cultivation was profitable *on average* using transportation networks in 1900. Web Appendices 3 and 4 describes how this map was created. Using a GIS map of networks in 1900 (see Fig. 3), we estimate for each cell the minimal transport cost of one ton to any port. We account for topography when calculating this cost as the topographic distance is higher than the Euclidean distance. From Cardinall (1931), we get an estimate of the production cost of one ton of cocoa, £20.9 on average.[9] Given the coastal producer price was £31.3 on average in 1920-29, production is profitable in a cell if 31.3 - 20.9 - transport costs $>$ 0, or transport costs $<$ £10.4.[10] As in Donaldson & Hornbeck (2011), the reduction in trade costs must have expanded production in the *feasible* region, where production was already profitable ex-ante but is now even more profitable, and in the *infeasible* region, where production was not profitable ex-ante but is now profitable. Figure 2 displays both regions. Pre-railroad trade costs

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.

[9] This includes the cost of establishing a farm in the forest (five years with no production) and the annual cost of labor when trees bear fruit, assuming a farm lifetime of 30 years.

[10] £10.4 is an upper bound as there are various costs that we cannot account for when a subsistence farmer adopts cocoa production: risk premium due to risk aversion, losses during transport due to theft or rotting, etc.

were prohibitively high, limiting cocoa production to a narrow coastal strip. 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% (see Fig. 4).

Roads were first complementary to the railroads as they were feeders to them. The first lorry was imported in 1903, but there were only two lorries in 1914 (Luntinen 1996). Besides, roads were of poor quality until 1924. Roads later became serious competitors for the railroad and opened new areas to cocoa. Even if no railroad had been built, roads would have permitted the boom. Our goal is not to compare the respective impacts of railroads and roads. We focus on the railroad age because it provides us with a natural experiment to identify the effect of colonial investments on development.

2.2.3 Placebo Lines

The British had to choose between a western, central or eastern route, and various lines were proposed and surveyed before the Western and Eastern Lines were built. We can address concerns regarding endogeneity by using these alternative railroad routes as a placebo check of our identification strategy. Web Appendix 2 gives a detailed background of each line.

Five alternative routes were proposed before the first line was built. 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. These proposed lines all had the same probability of being built as the two lines that were indeed built, and only random events (e.g., a change in the colonial Governor) 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 various 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 on the Volta river (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, who replaced Nathan, 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 of the type of cocoa predominantly grown in Ghana starts after 5 years (Jedwab 2011). Hence, to observe any impact on cocoa 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 cocoa, kola, palm oil, and timber exports. We verify that there are no effects for these lines in 1927.

2.3 The Cash Crop Revolution and Development in Ghana

Cocoa has been the main motor of Ghana's development, and this made it a leader of the African "cash crop revolution" (Austin 2008; Gollin, Jedwab & Vollrath 2012). Cocoa was introduced by missionaries in 1859, but it took 30 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 grid 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).[11] As Ghanaians realized how profitable cocoa was, more and more people specialized in the crop. Why did production boom in Ashanti, around Kumasi, and not in the South-West, around Sekondi? Trade costs were lower there. But the South-West is characterized by very poor cocoa soils and too much rainfall.[12]

Ghana has experienced sustained population growth after 1901. Its population increased from 1.9 million in 1901 to 3.2 million in 1931. Population growth accelerated in 1921-1931, when the annual growth rate was 3.6%, against 1.3% in 1901-1921. Population was 18.9 million in 2000. Figure 5 displays total population for southern cells in 1901, 1931, and 2000. The comparison with Figure 4 suggests that population increased in cocoa-producing areas. While Ghana was almost unurbanized at the turn of the 20th century, it is now one of the most urbanized countries in Africa. It

[11]The Botanical Gardens were established in Aburi in 1890, because of its health climate and its proximity to Accra.

[12]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.

started its urban transition earlier than most African countries, due to the cocoa boom (Jedwab 2011; Gollin, Jedwab & Vollrath 2012). 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 and 68.5% in 2000. The two largest cities are Accra, the national capital, and Kumasi, the hinterland capital. Before the 20th century, towns were state capitals or trading centres (see Dickson 1968, p.70-71). Most of the latter were on the coast, where European merchants would meet local merchants from the interior. There were also trading centers in the north, which benefitted from their location on historical trade routes. In the early 20th century, most of urban growth took place in the forest zone, with the development of cocoa production, modern transportation and mining (see Dickson 1968, p.246-261). Towns grew because they were cocoa buying centers, the homes of wealthy cocoa farmers, or market towns where they would spend their income.

3. EMPIRICAL STRATEGY IN 1901-1931

We first test if connected cells experience a boom in cocoa production, population growth, and urban growth in 1901-1931. We explain the various strategies we implement to obtain causal effects.

3.1 Main Econometric Specification

The main hypothesis we test is whether rail connectivity drives cocoa production and population growth. 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 $t = [1901, 1931]$:

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

where our dependent variable is the production (tons) 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 run a second model:

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

where our dependent variable is population of cell c in year t. We expect rail connectivity to have a positive and significant effect on cocoa 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 cocoa production along the lines. There could be an independent railroad effect on population, so our goal is not to instrument production with railroads. Our goal is just to highlight one of the mechanisms at play.

We have a panel of 2,091 cells. Our main 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. We also restrict our sample to those cells for which we have total population in 1901.[13] We end up with 554 observations, and we believe these restrictions give more conservative estimates. We argue in Section 2.2 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.

3.2 Exogeneity Assumptions and Controls

First, even if the Eastern Line was endogenous, the Western Line was built for mining and military domination. Endogeneity is not a concern if we find similar effects for both lines. Second, even if the Eastern Line had been built for cocoa, production would have remained small before the line was built because trade costs were prohibitively high. This is similar to arguing that the timing of line construction was exogenous. Third, we 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.[14] 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, a port in 1901, and the coast. Fourth, since we have data for 11x11 km grid cells, we compare neighboring locations that are unlikely to differ in terms of unobservables. Cell area is 122 sq km, only 40% more than Manhattan's area. By comparison, the mean areas of Indian districts and American regions, used as spatial units in the literature, are 4,300 and 78,977 sq km respectively. Cells less than 50 km from the lines are all similar in terms of observables.[15] If the placement is truly exogenous, the effect should strongly decrease as we move away from the line, which is exactly what we find. We can also include district-year fixed effects to control for time-variant heterogeneity. We have only 9 cells per district.

[13]We only have georeferenced population data for Southern Ghana in 1901.

[14]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 12,048. Cocoa and mining accounted for 94.5% of exports.

[15]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 having a mine for the 0-10 km cells, but these effects are small.

3.3 Parallel Trends

First, if the placement or timing is 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. There are no significant effects in 1891-1901, while we will show there are strong effects in 1901-1931.[16]

3.4 Placebo Regressions

Five lines were planned but never built. Two lines were built after 1923. For each line, we create a placebo treatment dummy equal to one if the cell is less than X km from the line. First, we expect no effect for the placebo cells. Second, 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 effects for the segments of these lines that do not intersect with existing lines. Third, we also compare the placebo cells with the other control cells, by dropping the railroad cells.

3.5 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) and Banerjee, Duflo & Qian (2012), who respectively instrument U.S. highways and Chinese railroads 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 not correlated 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 domination. It went through dense tropical forest and the random location of the mines explained why this interior line was built from Sekondi to Kumasi. Regarding the Eastern Line, Accra was the administrative and economic capital of Southern Ghana while Kumasi performed the same role for the hinterland. It was obvious that the two largest cities would be connected at one point. We also use as an instrument the distance from the straight lines Sekondi-Tarkwa-Obuasi and Accra-Kibi (see Fig. 3) to exploit the fact that mining was a major motivation behind rail building.

[16]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.

3.6 Matching

We 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 unobservables. We regress each control on a dummy equal to one if the cell is less than 20 km from a 1918 rail line. We alternatively consider as a control group all the suitable cells (see Col. (1) of Table 1), placebo cells (see Col. (2)) and placebo cells using Cape Coast-Kumasi only (see Col. (3)). Placebo cells are similarly defined as cells that are less than 20 km from a placebo line. First, when compared to all suitable cells, treated cells have a larger rural population and are closer to main cities, which could lead to an upward bias, and have worse cocoa soil quality, which could lead to a downward bias. It is not obvious in which direction coefficients would be biased. Second, the same biases exist when comparing treated cells and all placebo cells. Third, we can compare treated cells with each placebo line, as some of them could prove a better counterfactual. For example, 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.[17] Using these cells as a control group should lead to a downward bias and give more conservative estimates. Results are the same whichever control group is selected.

4. RESULTS

In this section, we display the main results and examine their robustness.

4.1 Main Results

Table 2 shows the main results for cocoa production and population growth. Column 1 reports the results for model (1), while columns (2)-(13) 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 (see Col. (1)). There is a strong effect on population growth up to 20 km (see Col. (2)). People tend to live in the vicinity of the line, although there is some production beyond 20 km. Interestingly, the rail effect is lower when we include cocoa production, which has a strong effect on population (see Col. (3)). The remaining rail effect is picked up by the cell dummy for having a rail station in 1918 (see Col. (4)), but that effect also becomes lower and non-significant when we include the amount brought to the station (see Col. (5)). This means the railroads have a strong effect on population growth,

[17] Treated cells have less rainfall but the difference is small: only 4% of the average.

and that this growth is coming from opportunities in the cocoa sector and other sectors if there are intersectoral linkages.[18]

The railroads have two effects on population growth, which we study by looking at rural growth and urban growth. We call the first effect, the number of additional inhabitants per ton of cocoa produced, the *labor effect* (1.53***, see Col. (5)), as more cocoa production requires more labor. The comparison of Columns (5), (9) and (13) indicates that most of the labor effect takes place in villages (1.15***, see Col. (9)). This is logical as cocoa is produced on farms surrounding producing villages (Jedwab 2011). We call the second effect, the number of inhabitants per ton of cocoa transported, the *trade effect* (0.86*, see Col. (5)), as more cocoa being transported requires larger rail stations. The comparison of Columns (5), (9) and (13) shows that the trade effect occurs in towns (0.91**, see Col. (13)). When using the urbanization rate as the main outcome, we find positive effects but only significant at 15% (not shown, but available upon request). Indeed, rural population increases almost as much as urban population.

Do the effects vary for various size categories of towns: 1,000-2,000 (small), 2,000-5,000 (medium) and 5,000-10,000 (large)? Running the same model as in Col. (10), we find no effect for small towns, a small effect for medium towns and a strong effect for large towns. If we run the same model as in Col. (13), the labor effect is mostly explained by medium towns, and many of them did not exist in 1901. The trade effect is mostly explained by large towns, and many of them already existed in 1901. To sum up, the railroads induced a cocoa boom, which had various effects on population growth: more producing villages up to 30 km away, more small producing towns up to 20 km and larger rail stations up to 10 km away.

4.2 Alternative Identification Strategies

Table 3 displays the results when we implement the various identification strategies. Col. (1) replicates our main results from Table 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. Column (2) of Table 3 shows there are stronger effects for highly suitable cells. When comparing the results for

[18]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 (2011) discusses how the causality is unlikely to run from population to cocoa in 1931. First, settlement was limited in tropical forests due to thick vegetation, high humidity and disease incidence. Farmers overcame these constraints when they could grow cocoa. Second, cocoa cultivation did not depend on cities for the provision of inputs, as it only required forested land, axes, machetes, hoes, cocoa beans and labor. This traditional mode of production was not conducive to a role for cities in the diffusion of innovations.

the Western and Eastern Lines (see Col. (3) and (4)), we find lower effects for the Western Line, but this is explained by the fact that it goes through poorly suitable cells. If we restrict the analysis to highly suitable cells only, the effects are not significantly different for both lines.

District-Year Fixed Effects. Column (5) of Table 3 shows results are robust when we include district-year fixed effects. Using the district boundaries in 2000, we have 554 cells in 62 districts, or 9 cells per district.

Placebo Regressions. Table 4 displays the results of the placebo regressions. Panel A, B and C show the results for production, while Panel D, E and F show the results for population. 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. The only concern is cocoa production for Accra-Oda-Kumasi 1897 (see Col.(4)). However, this effect exists because the line intersects with the area of influence of the Eastern Line. We find no effects for the segments of the placebo lines that do not intersect with the existing rail lines, as shown in Panel B and E. The issue here is that the railroad cells belong to the control group, which leads to significant negative placebo effects. In Panels C and F, we drop the railroad cells to conduct our preferred placebo test, which only compares the placebo cells and the other control cells. Except for Accra-Kpong 1898, coefficients are very small and never significant.

Matching. Columns (6) and (7) of Table 3 show that coefficients are higher when all the placebo cells or just the cells along the Cape Coast-Kumasi placebo line (20 km) are used as the control cells. Results are robust to using the other placebo lines (not shown, but available upon request). Results are also robust if we restrict the control group to the cells in the 40-50 km range.

Instrumentation. In Column (8), we instrument the rail dummies by dummies for being 0-20 and/or 20-40 km from the straight lines Sekondi-Kumasi and Accra-Kumasi. In both cases, the IV F-statistic is strong enough (5.7 in Panel A and 63.7 in Panel B), and results are unchanged. Results are the same if we use the distance from the straight lines Sekondi-Tarkwa-Obuasi and Accra-Kibi as the instrument instead.[19]

4.3 Robustness and General Equilibrium Effects

We now perform robustness checks and discuss general equilibrium effects.

[19] 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 and was underestimated beyond 40 km. 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 use exhaustive census data for population.

Robustness. Results are similar if we use the Euclidean distance to rail stations instead (see Col.(9)), if we drop the controls (see Col. (10)), if we drop the nodes of the network (see Col. (11)), if we use the full sample (see Col. 12) and if we use a log-linear functional form (see Col. (12)). We have no measure of total population for the full sample in 1901, but the results are the same for the urban population (not shown, but available upon request). Second, results are robust if we control for trade costs given transportation networks in 1901 and/or the road network in 1922 (not shown, but available upon request). We use roads in 1922 because they could affect production 5 years later. Estimates are slightly higher if we restrict our sample to the infeasible region. By definition, production is already profitable *on average* in the feasible region. We expect lower effects, unless there is land and farmer heterogeneity and the decrease in trade costs has an extensive margin effect in this region as well. Accordingly, we find a strong effect in the 0-10 km range (+595.5**), and no significant effect beyond. Third, we may underestimate standard errors if there is spatial autocorrelation. Results are robust to having Conley standard errors using a distance threshold of 50 km or more, or clustering them at an aggregate spatial level, such as provinces in 1931 or 2000 (not shown, but available upon request).

Magnitude and the Reversal of Fortune. We estimate that the railroad effects explain 44.5% of the change in cocoa production in 1901-1927 and 46.5% of population growth in 1901-1931. Another way to assess their magnitude is to test if they were large enough to produce a reversal of fortune in Southern Ghana. We run the same model as before except we add a second treatment variable, a dummy equal to one if the year is 1931 and the cell already had a town in 1891. We thus compare the railroad effects with the effect of being historically developed, using other cells as a control group. The effect of pre-railroad development is zero for production and slightly lower than the railroad effects for total population (not shown, but available upon request). In 1901, railroad cells were half as populated as the already developed cells. In 1931, there is no significant difference anymore. No reversal of fortune is observed, but railroads allowed some cells to entirely catch up with the most developed cells.

Results on Infrastructure. For each cell, we know the number of government and non-government schools and European and African hospitals, and whether the cell was crossed by a class 1, class 2 or class 3 road in 1901 and 1931. We use the same regression as model (2) to test whether railroad cells got better infrastructure over time, although no significant difference is observed in 1901. There are strong positive effects on the number non-government schools (+0.70* 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 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 public goods, whether government schools or European hospitals. These effects strongly decrease when we control for population. This confirms that railroads increased population density, and public goods had to be created as a result.

Results on Height. We have data on all Ghanaian male recruits in the British Army in 1888-1960. We know their year and place of birth, the year they enlisted, and their height. We restrict our analysis to soldiers that served during World War I or later. In total, we have height data for 5,447 soldiers across 298 cells. Using a regression similar to model (2), with the same controls plus district-year fixed effects, and individual controls for age, farming, skills, and ethnicity, we compare soldiers born in railroad vs. non-railroad cells (using 40 km as a cut-off) before and after 1923. Better standards of living should increase height for native soldiers born just after the cocoa boom. Railroad cells start with a disadvantage (-0.63 cm), which they more than overcome over time (+1.70* cm). This height improvement is four times the improvement for the sample as a whole over the same period (+0.38 cm).[20] This effect is halved and becomes not significant when we control for cocoa production in 1927, which becomes significant. This confirms that the railroad effect goes through more production and better standards of living.

Migration vs. Natural Increase. The total population of the 554 cells doubled from 1901-1931. This growth was mostly due to in-migration from the non-forested areas. 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." Here is a simple model of population growth for cell c at time t: $\triangle P_{c,t} = (CRB_{c,t-1} - CRD_{c,t-1})P_{c,t-1} + (I_{c,t-1} - E_{c,t-1})$. Population growth $\triangle P_{c,t}$ can either be explained by natural increase (when the crude rate of birth CRB is higher than the crude rate of death CRD), or by migration (when there are more immigrants I than emigrants E). First, if growth is driven by natural increase, we expect a strong effect of $P_{c,t-1}$ on $\triangle P_{c,t}$. We control for 1901 population in all regressions. Since the results are the same when initial population is not included as a control, natural increase did not play a major role here. Second, until 1921, the birth and death rates were similar, around 52 per 1,000 people, and natural increase was small. In 1921, the death rate decreased to 25 in the main towns and 44 elsewhere. Natural increase contributed to population growth from 1921, but this effect was larger for towns. Thus, migration was the main source of population growth, and this effect was reinforced by natural increase.

[20] If we compare soldiers born before and after 1914, thus also including the potential effects of better standards of living on these soldiers who were already 1-9 year-old when cocoa production was booming, the effect increases to +2.12* cm.

Migration and General Equilibrium Effects. If railroads reallocate labor across space, does overall welfare increase? First, are control cells negatively affected by migration? Non-railroad cells that were historically populated gained population between 1901 and 1931 when compared to other control cells. Locations that already had a city in 1891 grew further after 1901, and most placebo cells did not lose population when compared to other control cells (see Panel F of Table 4). Yet it is likely that some cells of the non-forest areas, which are not included in the sample of suitable cells, lost population and welfare may have decreased as a result. Second, has overall welfare increased despite the loss of population in some areas? People migrate because they obtain a higher utility at the destination location (Harris & Todaro 1970). Migration is rational and it leads to a more efficient spatial allocation of resources. In Ghana, the railroads gave access to a new factor of production that made people more productive, forested land that 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 *on average* than subsistence farmers (see Web Appendix 4). Subsistence farmers accounted for 95% of employment in 1901. The employment share in the cocoa sector increased from almost 0% in 1901 to one third in 1931.[21] Thus, one third of the population became more productive because they gained access to a new factor of production. Our framework is different from other studies of infrastructure projects, for which selective migration may be an issue (see Dinkelman 2011 for an interesting discussion in the case of electrification). This allowed Ghana to become one of the wealthiest African countries (Austin 2008; Gollin, Jedwab & Vollrath 2012). Ghana was the fourth most urbanized African country in 1931 and had the fifth highest per capita GDP in 1950. More than half of population growth happened in cities. Railroads caused a qualitative change in population, by increasing overall density. We find positive effects on infrastructure and height.

5. RAILROADS AND DEVELOPMENT DURING COLONIZATION

In this section we offer a general discussion on the role of railroads in development before independence. We study the channels through which railroads transformed the colonial economy and discuss welfare implications.

5.1 Colonial Railroads, Trade, and Urbanization

Here is the conceptual framework we have in mind. Assume a poor country is divided into grid cells. Each cell lives in autarky, and productivity is such

[21]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.

that everyone lives at the subsistence level. Most cells are a tropical forest with limited settlement. Cash crops could be produced there but there are no means of transportation. For exogenous reasons, some cells are connected to the rest of the world via a new transportation network. They specialize in their comparative advantage and export cash crops against food and non-food imports. Non-connected cells remain in autarky and close to the subsistence level. In this framework, towns grow as intermediary trading stations between producing areas and the ports. We use data on trade and urban employment to provide more evidence on these channels.

Colonial Railroads and Trade Integration. First, we find that trade (exports plus imports) accounted for 74% of railroad traffic (in volume) on average in 1904-1931. More precisely, it was 35% for exports, 39% for imports and 26% for internal traffic. In the period 1904-1918, imports accounted for 58%, because the Western Line was used to import building materials, machinery and coal for the gold mines. In the later period 1919-1931, exports accounted for 55%, because cocoa boomed along both lines, as well as manganese and timber exports. Second, we can easily link the export and import structures and railroad traffic. The export structure was the following in 1930: cocoa 76%, mining 23% and timber 1%. The traffic structure for export was: cocoa 72%, mining 24% and timber 1%. The import structure was the following in 1930: food, drinks, and tobacco 30%, clothing 20%, other consumption goods 15%, construction 9%, fuels 4%, machinery 6%, and transport equipment 10%. We find a rather similar traffic structure for import: food, drinks, and tobacco 31%, clothing 11%, construction 12%, fuels 16%, machinery 9%, and transport equipment 6%. This analysis confirms the role of railroads in promoting trade.

Colonial Railroads and Urbanization Why do railroads increase urbanization? First, many farmers settle in towns as they offer better conditions of living. Using data from the 1931 census, we find that 48.5% of the urban male workforce works in agriculture. The data does not distinguish cocoa and other farmers, but this figure shows these farmers probably work on farms outside the city limits. Second, towns serve as trading stations for cocoa exports (transportation to the coast) and imports (from the coast). Trade accounts for 20.6% of urban male employment, with 15.2% being traders, 1.2% being cocoa brokers, and 4.2% working in the transport sector. Railroad workers contributed to less than 0.75% of urban change in the south between 1901 and 1931. Third, cocoa farmers are wealthier than subsistence farmers. Cocoa generated a surplus for these farmers, and the Engel curve implies they spent more on "urban" goods and services, i.e. goods that were produced in the cities or distributed through the cities (Jedwab 2011; Gollin, Jedwab & Vollrath 2012). The analysis above shows that food, drinks, and tobacco, clothing, and other consumption goods amount to two thirds of imports. Some of these goods were also produced locally, as we find that light manufacturing and services account for 30.9% of urban male employment: domestic servants (6.6%), carpenters (4.5%), tailors (2.4%), government civil servants (2.3%), masons (2.0%), goldsmiths (1.3%), policemen (1.3%), teachers (1.1%), and washermen (1.0%).

Colonial Railroads, Trade and the Central Place Theory Our results are in line with the central place theory (Christaller 1933): Settlements are central places that provide services to surrounding areas and urban systems hierarchically organize these settlements. We provide evidence for the emergence of a complex urban system as a result of colonial investments. First, villages are where cocoa beans are produced. Second, small-sized towns serve as relay stations between the producing villages and the railroad stations. Third, medium-sized towns are railroad stations from which cocoa beans are brought to the coastal ports. Fourth, large cities are ports through which cocoa beans are exported or administrative centers.

5.2 Welfare Effects

In this section we contrast the welfare gains from railroad construction based on cocoa production alone using: (i) the social savings approach, and (ii) a more direct approach estimating producer rents.

5.2.1 Social Savings Approach

The social savings are calculated as the cost difference between railroads r and the next-best transportation alternative a (Fogel 1964): 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, lower-cost technology. Details of the social savings calculation are provided in Web Appendix 5. We consider head porterage as the main alternative to railroads.[22] The railroad transported a total of 13.5 million ton miles of cocoa at costs of about £350,000 in 1927. If the same volume had been moved by headloading instead, the hypothetical costs would have been £3.7 million. The cost difference are the social savings, i.e. 8.7% of GDP.[23] We can also relate the benefits to the cost of railroad construction. Cocoa represented only 30% of freight revenues, though rail building required the same fixed capital. In 1927, the railroad's capital outlay was £8.5 million, implying a social rate of return of 39.6% for cocoa alone. For comparison, the social rate of return to primary education is below 20% (Psacharopoulos & Patrinos 2004).

[22]With the expansion of roads from the 1920s on, the next-best alternative was lorry transport. In this exercise, however, we do not consider motor transport as counterfactual, but rather as part of modern transport technology introduced in Ghana.

[23]Applying the costs of 5s per ton mile to all goods would give social savings of 27% of GDP. Fogel (1964) found social savings of 5% for the U.S. in 1890. Donaldson (2010) found social savings of 9.7% for India.

Under perfect competition and inelastic demand for transport, social savings are identical to the gain in consumer surplus brought about by rail (see Web Appendix 5). If these conditions do not hold, social savings are larger than the increase in consumer surplus (Fogel 1979). First, demand was highly elastic in Ghana, as railroad construction triggered significant cocoa production inland. Without rail, production would have been 44.5% lower. Second, railroads were much cheaper than available alternatives in 1901. Hence, for Ghana, social savings and actual welfare gains may differ considerably.

5.2.2 Producer Rents

Our GIS analysis allows us to estimate the producer rents created by the railroad as follows: $\Delta\text{Producer rent} = \sum_i \left[\pi_{1918,i} - \max(\pi_{1900,i}, 0)\right]$, where π represents the profit of growing cocoa at observed location i in 1927. π_{1900} are profits based on the 1900 transportation costs, whereas π_{1918} refers to 1900 transportation plus the rail network of 1918 (see Web Appendix 5). We assumed costs of a small "native" cocoa farm, and we took the 1920-29 average producer price as a basis for revenues. For 88,900 tons or 40.7% of cocoa production in 1927, π_{1900} is negative. In these cases, cocoa production is unprofitable, cocoa would not have been grown and therefore the rent is zero. With the 1922 rail network only 3.5% of cocoa production is calculated as unprofitable - which is within the error margin of heterogeneity in production costs. Overall, we find that railroads increased producer rents by £1,264,895 or 3.3% of GDP. The real contribution to GDP is much higher as production costs and transportation also create added value.

Both approaches ignore externalities. Chaves, Engerman & Robinson (2012) pointed to the ending of slavery as one positive externality. Austin (2005) attributed the ending of slavery in Ashante to the adoption of cocoa cultivation. Our evidence suggests this connection may have existed between railroads and slavery. Without railroads, there would have been less cocoa, and a slower decline in slavery. In our paper, we tested for the existence of other externalities, increasing population densities and urbanization.

5.2.3 Rent Extraction

The colonial administration and European cocoa buying firms captured parts of the rail-induced cocoa rent. In 1920-1929, cocoa farmers received on average 79% of the export price. The price wedge results in a sum of £1,635,226, of which only 44.5% can be attributed to the railroad. In sum, while Ghanaians received rents of 3.3% of GDP, Europeans received 1.9% of GDP. Railroads generated rents on the order of 5.2% and two thirds of the surplus went to Ghanaians. Colonization meant both extraction and development in this context. The issue here is why Europeans did not capture the whole rent. First, Europeans could not produce cocoa themselves as the mortality rate of settlers was high in the tropical forests of West Africa. Second, the colonial strategy in Ghana was both extractive and non-extractive. Europeans thought that they were bringing civilization, whether christianity or capitalism (through trade). Luntinen (1996, p.57) writes: "The proposed railway was believed to be beneficial for the natives of the Gold Coast. Of course, the main objective of the several schemes was to advance trade and industry on the coast, to develop the colony, and to bring profit and revenue, i.e. to make the British colonialists rich and British imperialists mighty. But this did not exclude a benevolent attitude towards the natives."

6. COLONIAL RAILROADS AND LONG-TERM DEVELOPMENT

In this section we document the decline of colonial railroads and study their effects on long-run development. Railroad cells are more developed today, although they have lost their initial advantage in terms of transportation and cocoa cultivation. We investigate the channels of path dependence.

6.1 Evidence on the Decline of Colonial Railroads

By 1931, 500 miles of track had been laid, and rail transported 759,000 tons of goods (1,518 tons per mile) and 1,336,000 passengers (2,673 passengers per mile). The network reached its maximum size, 750 miles, in 1957 the year of independence. From 1944 to 1974, rail transported on average 2,400 tons of goods and 8,000 passengers per mile. Traffic collapsed after 1974. In 1984, rail only transported 500 tons of goods and 2,900 passengers per mile. Traffic never recovered, and rail only transported 970 tons of goods and 2,900 passengers per mile in 2000. Similarly, while railroads accounted for more than 70% of cocoa transport until 1970, this share decreased to one third in the 1980s and 7% in 2000. Rail only transports manganese and bauxite now, these commodities being too bulky for road transport.

What caused the obsolescence of rail? Luntinen (1996) describes how considerable road investments and underinvestments and management issues in the rail sector produced a significant decline of the latter. The first government of Ghana massively invested in roads, which were deemed more modern and three times cheaper to build. However it was argued that maintenance costs were much lower for railroads, which were more competitive in the long-run. Second, political and economic instability had a damaging effect on public investments.[24] 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.

[24]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

Moreover, 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 cocoa 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 (2011) uses district panel data from 1901-2010 to describe how cocoa production has disappeared from the original regions 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 continuously 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 continuously decreasing afterwards (3.9 on average in 1980-2000). Railroad cells have lost their initial advantage in both transportation and cocoa cultivation. In these conditions, should we also expect these locations to be in decline?

6.2 Evidence on Colonial Railroads and Path Dependence

We use various development outcomes to test whether railroad cells remain more developed today. We run this cross-sectional regression for suitable cells c (N = 554) in year 2000:

$$Y_{c,2000} = \alpha'' + Rail_{c,1918}\beta'' + X_{c,}\zeta + w_c \tag{3}$$

with $Y_{c,2000}$ being a development outcome in 2000, $Rail_{c,1918}$ the set of rail dummies (using the lines in 1918) and X_c the same controls as before. Regression results for model (3) are reported in Table 5. In Columns (1)-(2), we show that railroad cells are more urbanized, whether we drop railroad nodes or not. When adding a dummy for having being historically developed (having a town in 1891), the railroad effect is twice as large as the pre-railroad development effect, which is not significant anymore, indicating a reversal of fortune (not shown, but available upon request). In Column (3), we use satellite data on night lights as an alternative measure of development, as in Storeygard (2011) and Henderson, Storeygard & Weil (2012).[25] Railroad cells have a higher share of inhabitants living in residences with solid walls (see Col. (3)), higher employment shares in manufacturing and services (see Col. (5)-(7)), and better access to communication, education and health infrastructure, as well as higher levels of human capital. These effects decrease as we move further away from the line, they seem stronger for more exclusive public goods (those with a lower mean) and are reduced when we control for population density and infrastructure provision in 1931. Even if these cells are more developed today, there may have been convergence after 1931. Has the level effect narrowed over time? As in Bleakley

[25] We thank Adam Storeygard for giving us access to their data.

& Lin (2012), we use population panel data over one century to study the dynamics of path persistence. We run the following model for 554 suitable cells c and years $t = [1901, 1931, 1970, 2000]$:

$$LPop_{c,t} = Rail_c \beta_t + \gamma'_t + \delta'_c + X_c \zeta'_t + \nu'_{c,t} \tag{4}$$

with $LPop_{c,t}$ being the log of population and $Rail_c$ the rail dummies. We use a log-linear model as Ghana's population boomed after 1931, and taking logs removes any linear trend. 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. While total and rural populations boomed in 1901-1931, the level effect is lower (but still significant) in 2000. Urbanization followed a different pattern, as the level effect in 1931 was halved in 1970 before increasing even further in 2000. Since we have urban data every decade, we can thoroughly study urban patterns. The level effect was high in 1931, decreased in 1948, increased in 1960, was halved in 1970 and 1984, and increased in 2000. These patterns are verified whether we raise the urban threshold to 5,000 or use urbanization rates as the main outcome (not shown, but available upon request). While the rail effects are significant at 15% in 1931, they are highly significant in 2000. This confirms that these locations became increasingly urban over time. A look at the history of Ghana helps us understand these patterns. Total production was halved between 1938 and 1944, because of the cocoa swollen-shoot virus disease. Hill (1963) describes how the cocoa towns were affected by the shock. In 1960-1984, the political and economic crisis, the movement of cocoa cultivation to other areas, and the decline of railroads could explain why the towns of the old producing areas were relatively losing. Between 1966 and 1984, per capita GDP decreased by 30% and the urban sector was hit hardest (Jedwab & Osei 2012). Growth resumed after 1984, and urban productivity and employment rose. It seems railroad cells had an initial advantage but this advantage only persists if the economic situation makes it possible.

6.3 The Channels of Path Dependence

We study how the rail effects on urban population today vary as we control for the various potential channels of path dependence. We run the following model for 554 suitable cells c and year 2000:

$$LUpop_{c,2000} = Rail_{c,1918}\beta + \theta LUpop_{c,1901} + \zeta LRpop_{c,1901} + X_c\lambda + \mu_c \tag{5}$$

with $LUpop_{c,2000}$ being the log of urban population and $Rail_c$ the rail dummies. $LUpop_{c,1901}$ and $LRpop_{c,1901}$ control for urban and rural population in 1901. X_c is the same set of controls as before. We drop the cells including the nodes. Results are reported in Table 6. Column (1) shows that railroad cells are more urbanized today. These effects disappear when we control for urban and rural population in 1931 (see Col. (2)). The long-run effects are thus explained by the level effect in 1931.

What could explain the effect of population density in 1931 on population density in 2000? Bleakley & Lin (2012) contrast the respective roles of historical and contemporary factors. First, *sunk investments* in 1931 could induce people to stay at these locations. If schools and hospitals are expensive to build, people are less mobile and original advantages have long-run effects. 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. Third, most papers on multiple spatial equilibria have focused on developed countries, while this study looks at an African country that is experiencing a *demographic transition*. The population of Ghana was five times larger in 2000 than in 1931, for the mere reason that the death rate dropped after 1945. In this context, we should expect any original advantage to be mechanically multiplied, although this does not tell us whether rising densities will result in higher productivity. We use data on investments in 1931 and 2000 and population panel data from 1931-2000 to examine the respective roles of sunk investments, coordination failures, and the demographic transition in explaining path persistence.

First, the 1931 population effects are unchanged when controlling for sunk investments (see Col. (3)). Public goods in 1931 have positive effects on urban density in 2000 but they do not explain path persistence. Second, the 1931 population effects decrease when we control for the cell population history, i.e. urban and rural population in 1970 and 1984. The population effects are much stronger in 1984. In Columns (6), we examine how these effects change when we control for contemporary factors: *physical capital*, which we capture with manufacturing and service employment (i.e., structural change), *human capital*, which we capture with the literacy rate and the share of adults with secondary education, and *public capital*, which we capture with various measures of infrastructure provision. These variables are the same as in Table 5. Human capital does not modify the population effects (not shown, but available upon request). Controlling for structural change and current infrastructure reduce them (see Col. (5)). Yet the 1984 population effects are still large and significant. These effects cannot be explained by coordination failures. We attribute them to demographic growth. Given our results, if urban (rural) population is 100% higher in one cell in 1984, it is 142% (133%) higher in 2000. We then split the urban sample into 1,000-5,000 localities and 5,000-+ localities. We find that coordination failures matter more for 5,000-+ localities as the effects of contemporary factors (structural change and infrastructure provision) strongly increase (see Col. (6) vs. Col. (5)). Demographic growth also matters as localities above 5,000 inhabitants grow further after 1984 (see Col. (7)). The coordination problem matters less for smaller towns (see Col. (8)), whose growth is a function of rural population and urban population in 1,000-5,000 localities in 1984 (see Col. (9)). These demographic effects seem large, but they match results using aggregate data. In 1984, the crude rates of natural increase were 31 and 23 per 1,000 people in rural and urban Ghana. They imply that if urban (rural) population is 100% higher in one cell in 1984, it is 144% (163%) higher in 2000. Rural migration to the cells in which contemporary factors are co-located could explain why we find higher urban growth rates in localities above 5,000 inhabitants (153%, see Col. (7)).

In line with Bleakley & Lin (2012), we do not find that sunk investments matter and we argue that railroads helped solve the coordination problem of capital accumulation. We also provide evidence that demographic growth reinforced these effects. Yet should we conclude that railroads have initiated modern economic growth in Ghana? Per capita income has remained very low, around $1,350 (PPP, 2010 USD). Although railroad cells are more industrialized today (see Col. (6) of Table 5, e.g. +3.5*** for 0-10 km cells), Ghana's employment share in manufacturing was only 6.4% in 2000. Manufacturing is relatively unproductive, as it consists of labor-intensive subsectors, such as African processed foods, furniture and clothing (Jedwab 2011, Jedwab & Osei 2012). Thus, colonial railroads made some locations more attractive *within* Ghana but did not industrialize the country as a whole.

7. CONCLUSION

What is the impact of colonial infrastructure investments on long-term development? We study this issue by looking at the effects of rail construction on development in Ghana. Using panel data at a fine spatial level over one century, we find strong effects of rail connectivity on the production of cocoa, the country's main export commodity, and development, which we proxy by population and urban growth. We exploit various strategies to ensure our effects are causal. Second, these investments had large welfare effects for Ghanaians before independence. Colonization meant both extraction and development in this context. Third, rail construction had a persistent impact: railroad cells are more developed today despite a complete displacement of rail by other transportation means. Physical capital accumulation and the demographic transition account for path dependence.

We believe the paper makes the following contributions to the literature. First, we study colonial investments in physical capital instead of colonial institutions. Second, we look at transportation infrastructure investments, since they represent a very high share of colonial budgets. Third, we find strong evidence for path dependence and discuss how these colonial investments shaped the economic geography of Ghana. Fourth, we contribute to the literature on multiple spatial equilibria, by examining how the channels of path dependence could be different in a poor developing economy. Fifth, to our knowledge, we are the first paper to use population panel data at a fine spatial level over one century for an African country. This data set will help study why Africa is still underdeveloped.

Lastly, we believe that the patterns found for Ghana are paralleled in other countries, e.g. coffee and tea in Kenya or groundnuts in Senegal. Colonial investments in transportation infrastructure shaped the economic geography of African countries, with consequences for long-run development.

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.

African Development Bank. 2003. *Review of the Implementation Status of the Trans-African Highways and the Missing Links.* African Development Bank: Tunis, and UN Economic Commission for Africa: Addis Ababa.

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.

Austin, Gareth. 2005. *Labour, Land and Capital in Ghana: From Slavery to Free Labour in Asante, 1807-1956.* Rotchester, NY: University of Rochester Press.

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

Banerjee, Abhijit, 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.

Bertocchi, Graziella, and Fabio Canova. 2002. "Did Colonization Matter for Growth?: An Empirical Exploration into the Historical Causes of Africa's Underdevelopment." *European Economic Review*, 46(10): 1851–1871.

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

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 Padro-i Miguel. 2011. “Ethnic Favoritism.” Unpublished manuscript, Department of Economics, London School of Economics.

Buys, Piet, Uwe Deichmann, and David Wheeler. 2010. “Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa.” *Journal of African Economies*, 19(3): 399–432.

Cardinall, A. W. 1931. *The Gold Coast.* Government Printer.

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

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

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

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

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

Dinkelman, Taryn. 2011. “The Effects of Rural Electrification on Employment: New Evidence from South Africa.” *American Economic Review*, 101(7): 3078–3108.

Donaldson, Dave. 2010. “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” National Bureau of Economic Research, Inc. NBER Working Papers 16487.

Donaldson, Dave, and Richard Hornbeck. 2011. “Railroads and American Economic Growth: A "Market Access" Approach.” Unpublished manuscript, Department of Economics, Massachussets Institute of Technology.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2000. “Factor Endowments: Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States.” *Journal of Economic Perspectives*, 14(3): 217–232.

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.

Fogel, Robert William. 1979. “Notes on the Social Saving Controversy.” *The Journal of Economic History*, 39(01): 1–54.

Gennaioli, Nicola, and Ilia Rainer. 2007. “The Modern Impact of Precolonial Centralization in Africa.” *Journal of Economic Growth*, 12(3): 185–234.

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. 2012. “Urbanization without Structural Transformation.” Unpublished manuscript, Department of Economics, George Washington University.

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

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

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

Hill, Polly. 1963. *The Migrant Cocoa-Farmers of Southern Ghana: A Study in Rural Capital-*

ism. Cambridge: Cambridge University Press.

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

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

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

Jacoby, Hanan C. 2000. "Access to Markets and the Benefits of Rural Roads." *Economic Journal*, 110(465): 713–37.

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

Jedwab, Remi. 2011. "Why is Africa So Urbanized? Evidence from Resource Exports in Ghana and Ivory Coast." Unpublished manuscript, Department of Economics, Paris School of Economics and London School of Economics.

Jedwab, Remi, and Robert Darko Osei. 2012. "Structural Change in Ghana 1960-2010." Case study prepared for the project "Structural Change in Developing Countries" under the supervision of Margaret McMillan and Dani Rodrik.

Limão, Nuno, and Anthony J. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade." *The World Bank Economic Review*, 15(3): 451–479.

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*.

Mu, Ren, and Dominique van de Walle. 2011. "Rural Roads and Local Market Development in Vietnam." *Journal of Development Studies*, 47(5): 709–734.

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.

Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. "Law and Finance." *Journal of Political Economy*, 106(6): 1113–1155.

Psacharopoulos, George, and Harry Anthony Patrinos. 2004. "Returns to investment in education: a further update." *Education Economics*, 12(2): 111–134.

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

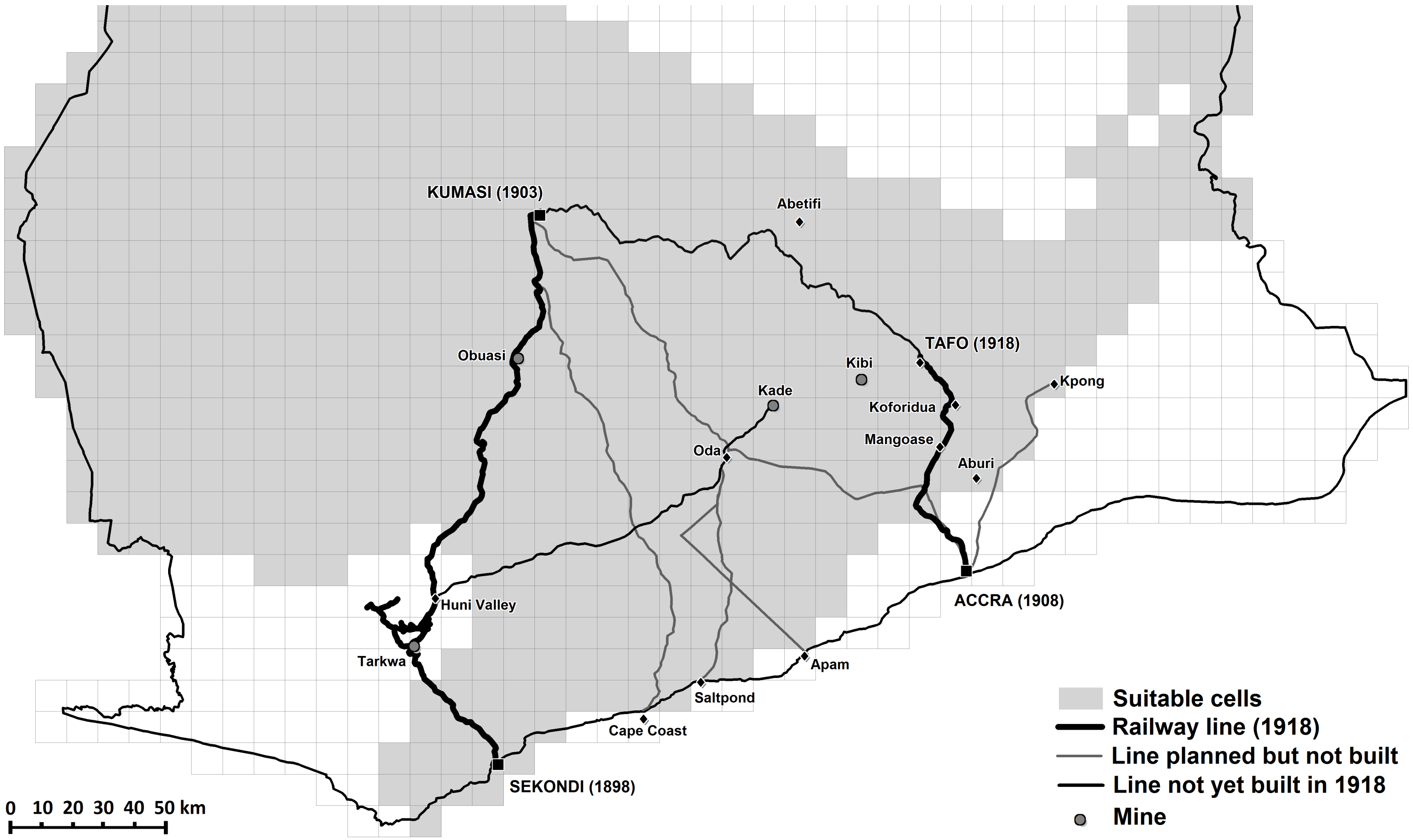
Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf. 2011. "History and Industry Location: Evidence from German Airports." *Review of Economics and Statistics*, 93(3): 814–831.

Storeygard, Adam. 2011. "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.

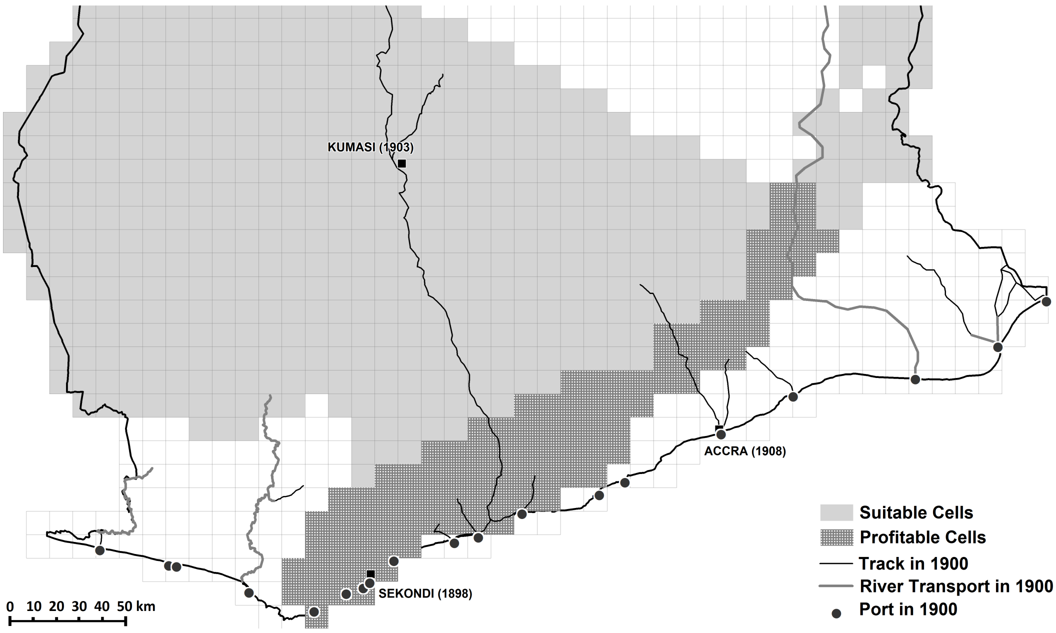
World Bank. 2010. *World Development Indicators.* Washington DC: World Bank Data.

Figure 1: Cocoa Suitability, Railroads and Placebo Lines in Southern Ghana.



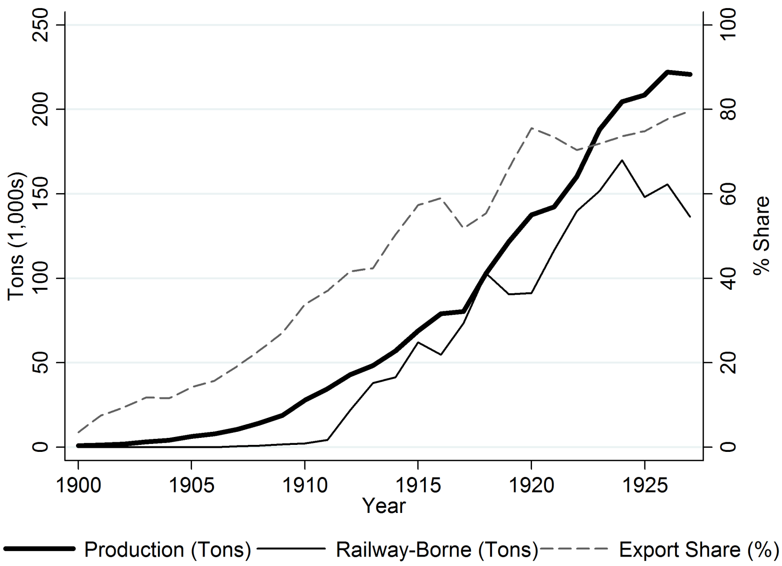
Note: An 11x11 km cell is defined as suitable if it contains cocoa soils. The map displays the rail lines in 1918, the years line construction begun and finished, the main rail stations and the placebo lines: lines that were planned but not built (Cape Coast-Kumasi 1873, Saltpond-Kumasi 1893, Apam-Kumasi 1897, Accra-Kumasi 1897, Accra-Kpong 1898) and lines that were not built early enough to affect cocoa production in 1927 (Tafo-Kumasi 1923, Huni Valley-Kade 1926). It also shows gold mines (Tarkwa, Obuasi and Kibi) and diamond fields (Kade). See Web Appendix 1 for data sources.

Figure 2: Transportation Networks in 1900 and Area of Profitable Production.



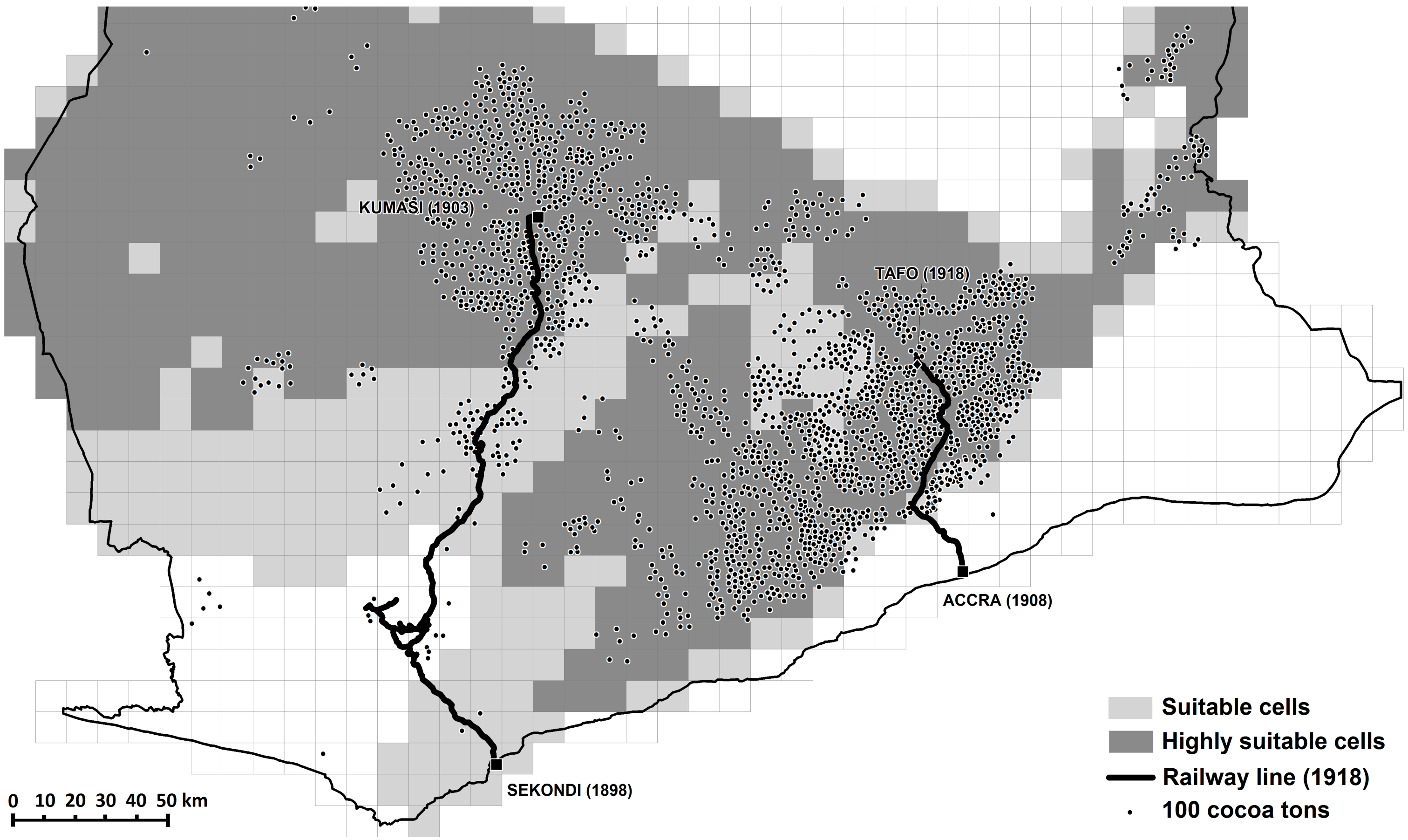
Note: The map shows transportation networks in 1900, the cells suitable for cocoa cultivation and the cells for which cocoa cultivation is profitable given transportation costs in 1900 and the producer price offered at the port in 1920-1929. Web Appendix 3 describes how this map was created. See Web Appendix 1 for data sources.

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



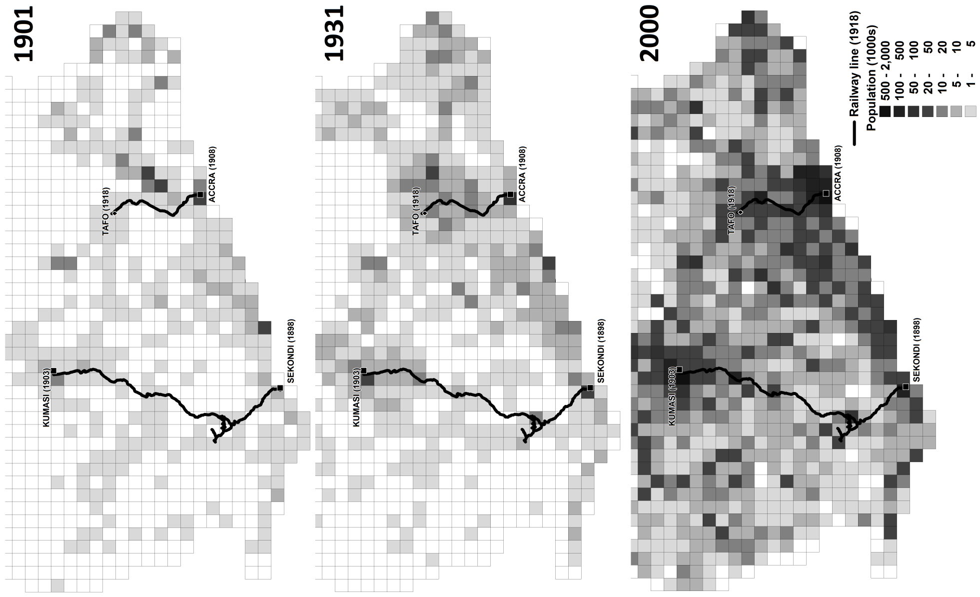
Note: The figure displays three-year moving averages for cocoa production, cocoa tonnages transported by rail to a coastal port, and the share of cocoa exports of total exports from 1900 to 1927. See Web Appendix 1 for data sources.

Figure 4: Railroad Lines in 1918, Cocoa Suitability, and Production in 1927.



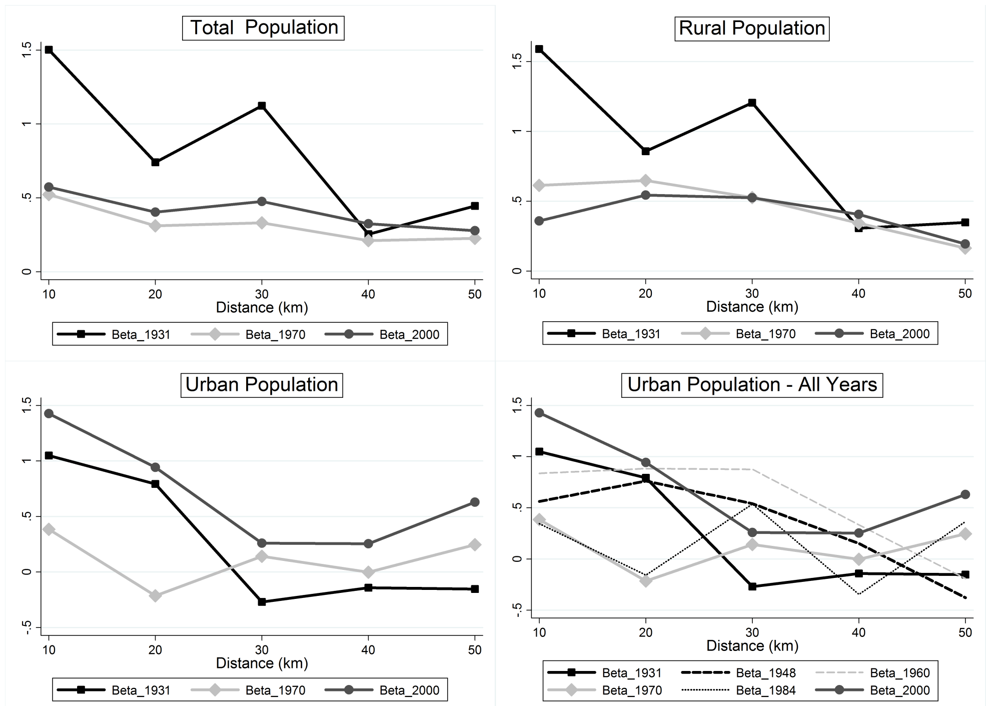
Note: The map only shows Southern Ghana. A cell is defined as suitable if it contains cocoa soils and highly suitable if more than 50% of its area consists of forest ochrosols, the best cocoa soils. The map displays the railroad lines in 1918, suitable cells, highly suitable cells, and cocoa production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

Figure 5: Railroad Lines in 1918 and Cell Population in 1901, 1931 and 2000.



Note: The maps displays cell population in 1901, 1931 and 2000. See Web Appendix 1 for data sources.

Figure 6: The Long-Term Effects (Relative to 1901) of Railroad Connectivity



Note: This graph displays estimates of Equation (4) for each distance threshold and each year = [1931, 1970, 2000], using 1901 as the reference year. We show the rail effects on urban population for additional years 1948, 1960 and 1984.

TABLE 1: OBSERVABLES FOR TREATED CELLS VERSUS CONTROL CELLS

RHS Variable:	**Dummy Rail 1918, 0-20 km**		
Control Cells:	All	Placebo	C.Coast-Kumasi
LHS 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 the coast (km)	-32.1***	5.9	28.6***
N Treated Cells:	104	104	104
N Control Cells:	450	152	44

Notes: Robust standard errors (not reported): * $p<0.10$, ** $p<0.05$, *** $p<0.01$. We regress each control variable on a dummy equal to one if the cell is less than 20 km from a 1918 railroad line. There are 15 different regressions for each column. In Column (1), all control cells are included. In Column (2), the control cells are the cells less than 20 km from a placebo line. In Column (3), the control cells are the cells less than 20 km from the Cape Coast-Kumasi placebo line (1873). See Web Appendix 1 for data sources.

TABLE 2: RAILROADS, COCOA PRODUCTION, AND POPULATION GROWTH

Dependent Variable:	Cocoa 1927 (Tons)	Population 1931 (Number of Inhabitants)				Rural Population 1931 (in Localities < 1,000 Inh.)				Urban Population 1931 (in Localities > 1,000 Inh.)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dummy Rail 1918, 0-10 km	666***	3,568***	2,405**	368	560	1,378***	625*	98	330	2,190**	1,781*	26	230
	(114)	(1,112)	(1,056)	(650)	(623)	(378)	(354)	(426)	(480)	(986)	(944)	(524)	(431)
Dummy Rail 1918, 10-20 km	506***	1,385***	502	442	577	916**	344	159	326	469	158	106	250
	(110)	(513)	(507)	(510)	(481)	(377)	(328)	(327)	(328)	(309)	(361)	(363)	(311)
Dummy Rail 1918, 20-30 km	341***	782	187	139	228	758**	372	165	360	24	-186	-226	-132
	(100)	(613)	(573)	(576)	(555)	(357)	(346)	(333)	(347)	(383)	(370)	(372)	(343)
Dummy Rail 1918, 30-40 km	246**	464	34	-4	87	206	-73	-90	-84	258	106	74	171
	(111)	(438)	(426)	(429)	(431)	(301)	(281)	(270)	(282)	(293)	(299)	(303)	(302)
Dummy Rail 1918, 40-50 km	55	203	107	67	89	-21	-83	-82	-90	224	190	156	179
	(93)	(383)	(391)	(383)	(376)	(239)	(251)	(241)	(251)	(248)	(246)	(236)	(230)
Cocoa (Tons Produced) 1927			1.75***	1.77***	1.53***		1.13***	1.13***	1.15***		0.62**	0.64**	0.38
			(0.40)	(0.41)	(0.39)		(0.24)	(0.20)	(0.24)		(0.30)	(0.31)	(0.27)
Dummy Rail Station 1918				4,106*	1,551			485	734			3,538*	817
				(2,102)	(1,746)			(561)	(591)			(2,013)	(1,519)
Cocoa at Rail St. (Tons) 1918					0.86*				-0.06				0.91**
					(0.44)				(0.06)				(0.44)
Cell FE and Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	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	1,108	1,108	1,108	1,108
R-squared	0.61	0.73	0.75	0.77	0.77	0.70	0.74	0.76	0.74	0.67	0.67	0.70	0.80

Notes: 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 and year fixed effects, and 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, and the coast. See Web Appendix 1 for data sources.

TABLE 3: ROBUSTNESS CHECKS

Regression:	*Main*	*Highly Suitable Cells*	*West vs. East*	Highly Suit.	*District-Year FE*	*Matching* Placebo Cells	C.Coast Kumasi	*IV Straight Lines*	*Distance Rail Stat.*	*No Ctrls*	*No Nodes*	*Full Sample*	*Logs*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Panel A: Dependent Variable = Cocoa (Tons Produced) in 1927													
Dummy Rail 1918, 0-20 km	556***	929***	316***	747***	419***	775***	731***	622***	571***	544***	552***	424***	2.06***
	(85)	(168)	(79)	(224)	(104)	(121)	(155)	(170)	(88)	(113)	(87)	(77)	(0.40)
Dummy Rail 1918, 20-40 km	275***	391***	115	326***	221***	407***	421***	35	260***	373***	275***	194***	0.96**
	(78)	(111)	(71)	(122)	(83)	(106)	(149)	(313)	(80)	(94)	(78)	(66)	(0.41)
Rail 1918 x East. Line 0-20 km			781***	334									
			(227)	(308)									
Rail 1918 x East. Line 20-40 km			509**	171									
			(200)	(272)									
Panel B: Dependent Variable = Population (Number of Inhabitants) in 1931													
Dummy Rail 1918, 0-20 km	2,034***	2,387***	1491**	1950*	1,496***	2,052***	1,754**	2,944**	2,027***	1,939***	1,486***	–	0.71**
	(605)	(858)	(624)	(1059)	(561)	(714)	(720)	(1,328)	(622)	(563)	(427)		(0.35)
Rail 1918 x East. Line 0-20 km			1,943	815									
			(1,434)	(1,749)									
Cell FE and Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,108	708	1,108	708	1,108	628	490	1,108	1,108	1,108	1,108	4,182	1,108

Notes: 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 and year fixed effects and cell controls interacted with year dummies. These are the same as in Table 2. In Col. (2), the sample is restricted to highly suitable cells. In Col. (3) and (4), we compare the effects for the Western and Eastern Lines. In Col. (4), we restrict the sample to highly suitable cells when comparing both lines. In Col. (5), we include district-year fixed effects. In Col. (6) and (7), the sample of control cells is restricted to all placebo cells and cells along the Cape-Coast placebo line (1873) respectively. In Col. (8), we instrument the rail dummies by the distance to the straight lines Sekondi-Kumasi and Accra-Kumasi. In Col. (9), the rail dummies are generated using the distance to rail stations. In Col. (10), we drop the controls. In Col. (11), we drop the railroad nodes. In Col. (12), we use the full sample, but the data set is not exhaustive for total population. In Col. (13), we use a log-linear functional form. See Web Appendix 1 for data sources.

TABLE 4: PLACEBO LINES

Type of Placebo Line:	**Planned But Never Built (From West to East)**					**Not Built Yet**		**All Lines**
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 1927							
Dummy Placebo 1918, 0-20 km	-169	-51	-86	299**	541	-13	-206*	13
	(104)	(104)	(102)	(121)	(357)	(126)	(112)	(99)
	Panel B: Dependent Variable = Cocoa (Tons Produced) in 1927							
Dummy Placebo 1918, 0-20 km	-519***	-231**	-273***	-9.2	71	-277**	-240*	-326***
x Dummy Rail 1918, 0-20 km = 0	(98)	(106)	(105)	(134)	(421)	(129)	(138)	(89)
	Panel C: Dependent Variable = Cocoa (Tons Produced) in 1927							
Dummy Placebo 1918, 0-20 km	-323***	-78	-132	130	523	-68	-181	-57
Drop if Dummy Rail 1918, 0-20 km = 0	(97)	(112)	(111)	(138)	(383)	(144)	(142)	(110)
	Panel D: Dependent Variable = Population (Number of Inhabitants) in 1931							
Dummy Placebo 1918, 0-20 km	-84	-83	-128	660	892	794	-1,203***	77
	(578)	(549)	(561)	(786)	(2,325)	(751)	(440)	(477)
	Panel E: Dependent Variable = Population (Number of Inhabitants) in 1931							
Dummy Placebo 1918, 0-20 km	-1,531***	-1,035*	-1,089**	-1,034*	489	-400	-1,611***	-1,209**
x Dummy Rail 1918, 0-20 km = 0	(487)	(538)	(524)	(574)	(3,594)	(606)	(499)	(539)
	Panel F: Dependent Variable = Population (Number of Inhabitants) in 1931							
Dummy Placebo 1918, 0-20 km	-684	-176	-341	-523	3,617	392	-1,213***	161
Drop if Dummy Rail 1918, 0-20 km = 0	(454)	(478)	(474)	(529)	(2,944)	(681)	(454)	(497)

Notes: 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 and year fixed effects and controls interacted with year dummies. In Panels A and D, we compare the placebo cells with the non-placebo cells (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 line with the other cells. In Panels C and F, we drop the cells less than 20 km from the railway lines in 1918, in order to compare the placebo cells with the other control cells.

TABLE 5: RAILROADS AND ECONOMIC DEVELOPMENT IN 2000

Dependent Variable:	Urban Population (Inh.)		Lights	Solid	Employment Share (%)			Post Office	Telephone
		No nodes	(%)	Walls (%)	Agri.	Manuf.	Serv.	(%)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dummy Rail 1918, 0-10 km	33,791*	10,059***	29.8***	14.1***	-16.0***	3.5***	9.5***	8.7*	15.9***
	(17,862)	(3,844)	(4.7)	(3.0)	(2.8)	(0.8)	(1.9)	(4.7)	(4.9)
Dummy Rail 1918, 10-20 km	3,785	3,231**	16.0***	4.6*	-7.7***	2.1***	4.2***	6.2	9.5**
	(2,767)	(1,373)	(4.3)	(2.4)	(2.3)	(0.7)	(1.6)	(4.4)	(4.2)
Dummy Rail 1918, 20-30 km	1,441	1,361	7.4**	1	-5.4***	1.7***	3.1**	-2.4	3.9
	(2,719)	(1,501)	(3.6)	(2.1)	(2.0)	(0.6)	(1.4)	(3.9)	(3.9)
Dummy Rail 1918, 30-40 km	1,046	858	3.7	2.7	-2.7	0.9	0.9	1.1	0.5
	(2,222)	(1,281)	(3.2)	(2.2)	(2.0)	(0.6)	(1.3)	(4.1)	(3.6)
Dummy Rail 1918, 40-50 km	4,633	1,569	0	4.8*	-4.3**	1.6***	1.8	4.6	2.7
	(3,143)	(1,406)	(3.0)	(2.7)	(2.0)	(0.6)	(1.3)	(4.3)	(4.1)
Mean	10,551	8,235	29.4	19.6	74.8	6.4	15.8	29.9	23.6
Dependent Variable:	Literate (%)	SSS Educ. (%)	Primary School (%)	JSS (%)	SSS (%)	Clean Water (%)	Clinic (%)	Hospital (%)	Paved Road
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Dummy Rail 1918, 0-10 km	7.6***	2.9**	7.1*	10.8***	12.3**	7.8**	9.1**	19.9***	0.31***
	(1.9)	(1.2)	(3.6)	(3.7)	(5.0)	(3.6)	(4.6)	(4.8)	(0.08)
Dummy Rail 1918, 10-20 km	2.4	-0.1	7.9**	9.6***	8.9**	0.2	5.4	10.7***	0.23***
	(1.6)	(1.0)	(3.1)	(3.2)	(4.2)	(2.5)	(4.0)	(3.5)	(0.07)
Dummy Rail 1918, 20-30 km	1.3	0.1	4.7	6.6**	1.7	1.1	0.4	7.2**	0.21***
	(1.6)	(0.9)	(2.9)	(3.2)	(3.7)	(2.7)	(4.0)	(3.3)	(0.07)
Dummy Rail 1918, 30-40 km	0.5	-0.6	5.1*	5.8*	-0.5	2.1	-0.9	5.3*	0.14**
	(1.5)	(0.8)	(2.7)	(3.1)	(3.5)	(2.3)	(4.0)	(3.1)	(0.06)
Dummy Rail 1918, 40-50 km	1.2	-0.4	5.1*	6.0*	-0.6	4.0	1.0	4.7	0.16**
	(1.5)	(0.9)	(2.8)	(3.2)	(3.7)	(2.5)	(4.1)	(3.3)	(0.06)
Mean	52.0	13.4	84.0	74.5	22.5	14.2	43.4	14.6	0.29

Notes: Robust standard errors are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. We have 18 outcomes for 554 cells in 2000. All regressions include the same controls as in Table 2. Col. (1) and (2): urban population (number of inh.), with and without the railroad nodes. Col. (3): share of cell area (%) for which a light is observed by satellite. Col. (4): share of inh. (%) in a residence with solid walls. Col. (5)-(7): sectoral employment shares (%). Col. (8), (9), (12), (13), (14), (16) and (17): share of inh. (%) living less than 5 km from a: post office, telephone line, primary school, junior secondary school (JSS), senior secondary school (SSS), health clinic or hospital. Col. (10)-(11): shares of adults ($\geq$ 18 y.o.) that are literate, have attended a SSS. Col. (15): share of inh. (%) with access to clean water. Col. (16): dummy if the cell is crossed by a paved road.

TABLE 6: RAILROADS AND THE CHANNELS OF PATH DEPENDENCE

Dependent Variable:	**Log Urban Population (Number of Inh. in Loc. > 1,000 Inh.)**								
Localities:						5,000-+	5,000-+	1,000-5,000	1,000-5,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dummy Rail 1918, 0-10 km	1.21***	0.13	0.22	0.51	0.07	-0.93**	-0.32	-0.08	-0.12
	(0.45)	(0.41)	(0.44)	(0.34)	(0.33)	(0.45)	(0.41)	(0.45)	(0.41)
Dummy Rail 1918, 10-20 km	0.91*	0.25	0.15	0.45	0.30	0.05	0.35	0.39	0.28
	(0.49)	(0.44)	(0.44)	(0.37)	(0.35)	(0.36)	(0.34)	(0.38)	(0.37)
Dummy Rail 1918, 20-30 km	0.24	-0.40	-0.49	-0.45	-0.54	-0.01	0.15	-0.45	-0.47
	(0.54)	(0.47)	(0.48)	(0.33)	(0.33)	(0.39)	(0.35)	(0.37)	(0.36)
Dummy Rail 1918, 30-40 km	0.26	0.11	0.30	0.24	0.18	-0.35	-0.16	0.39	0.31
	(0.46)	(0.44)	(0.45)	(0.33)	(0.31)	(0.32)	(0.29)	(0.35)	(0.33)
Dummy Rail 1918, 40-50 km	0.62	0.44	0.57	0.31	0.13	0.49	0.45	-0.01	0.04
	(0.45)	(0.39)	(0.39)	(0.30)	(0.28)	(0.41)	(0.35)	(0.35)	(0.34)
Log Urban Population, 1901	0.21***	0.08*	0.03	0.03	0.00	0.06	-0.04	0.04	0.04
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.08)	(0.06)	(0.05)	(0.04)
Log Rural Population, 1901	0.62***	0.19***	0.16**	0.06	0.01	0.07	0.08*	-0.01	-0.01
	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Log Urban Population, 1931		0.22***	0.19***	-0.00	-0.01	0.17***	0.10*	-0.06	-0.04
		(0.04)	(0.04)	(0.02)	(0.02)	(0.05)	(0.05)	(0.03)	(0.03)
Log Rural Population, 1931		0.60***	0.58***	0.21***	0.16***	-0.05	-0.03	0.24***	0.21***
		(0.07)	(0.06)	(0.06)	(0.05)	(0.03)	(0.03)	(0.06)	(0.06)
Log Urban Population, 1970				0.08**	0.04	0.06	0.06*	0.03	0.02
				(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Log Rural Population, 1970				0.26**	0.10	-0.06	-0.02	0.08	0.05
				(0.11)	(0.12)	(0.09)	(0.06)	(0.14)	(0.14)
Log Urban Population, 1984				0.51***	0.42***	0.03		0.39***	
				(0.05)	(0.05)	(0.03)		(0.05)	
Log Rural Population, 1984				0.35**	0.33**	0.13	0.09	0.46**	0.48**
				(0.14)	(0.14)	(0.11)	(0.10)	(0.19)	(0.19)
Mfg. Employment Share (%)					0.05*	0.12***	0.08**	-0.01	0.02
					(0.03)	(0.03)	(0.03)	(0.05)	(0.04)
Serv. Employment Share (%)					0.03*	0.08***	0.07***	0.04**	0.03*
					(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Log Urb.Pop. 1000-5000 1984							0.03		0.42***
							(0.02)		(0.05)
Log Urb.Pop. 5000-+ 1984							0.53***		-0.00
							(0.05)		(0.04)
Sunk investments 1901-1931	_	_	Y	Y	Y	Y	Y	Y	Y
Sunk investments 2000	_	_	_	_	Y	Y	Y	Y	Y
Human capital 2000	_	_	_	_	Y	Y	Y	Y	Y
Observations	552	552	552	552	552	552	552	552	552
R-squared	0.34	0.50	0.50	0.71	0.74	0.67	0.74	0.64	0.67

Notes: Robust standard errors are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. Sunk investments 1901-1931: number of government and non-government schools, European and African hospitals and dummy for being crossed by a class 1 or class 2 road in 1901 and 1931. Sunk investments 2000: share of inh. (%) in a residence with solid walls, share of inh. (%) living less than 5 km from a post office, telephone line, primary school, junior secondary school, senior secondary school (SSS), health clinic or hospital, share of inh. (%) with access to clean water, and dummy if the cell is crossed by a paved road. Human capital 2000: shares of adults that are literate, have attended a SSS. See Web Appendix 1 for data sources.