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

Remi JEDWAB[a] Alexander MORADI[b]

[a] *Department of Economics, George Washington University, and STICERD, London School of Economics*
[b] *Department of Economics, University of Sussex*

This Version: October 14th, 2012

Abstract: What is the impact of colonial public investments on long-term development? We investigate this issue by looking at the impact of railway construction on economic development in Ghana. Two railway lines were built by the British 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 effect of rail connectivity on the production of cocoa, the country's main export commodity, and development, which we proxy by population and urban growth. First, we exploit various strategies to ensure our effects are causal: we show that pre-railway transport costs were prohibitively high, we provide evidence that line placement was exogenous, we find no effect for a set of placebo lines, and results are robust to instrumentation and nearest neighbor matching. Second, transportation infrastructure investments had large welfare effects for Ghanaians during the colonial period. Colonization meant both extraction and development in this context. Third, railway construction had a persistent impact: railway cells are more developed today despite a complete displacement of rail by other means of transport. We investigate the various channels of path dependence, including demographic growth, industrialization or infrastructure investments.

Keywords: Colonialism; Africa; Transportation Infrastructure; Trade
JEL classification: F54; O55; O18; R4; F1

*Remi Jedwab, George Washington University and STICERD, London School of Economics (e-mail: jedwab@gwu.edu). Alexander Moradi, University of Sussex (email: a.Moradi@sussex.ac.uk). We would like to thank Robert Allen, Gareth Austin, Jean-Marie Baland, Sonia Bhalotra, Francois Bourguignon, Miriam Bruhn, Denis Cogneau, Paul Collier, Robert Eastwood, Marcel Fafchamps, James Fenske, James Foster, Elise Huillery, Eliana La Ferrara, Molly Lipscomb, Stephan Litschig, Rocco Macchiavello, William Masters, Stelios Michalopoulos, Branko Milanovic, Nathan Nunn, Henry Overman, Thomas Piketty, Jean-Philippe Platteau, Ariell Reshef, Elizabeth Sadoulet, Sheetal Sekhri, Stephen Smith, Francis Teal and and seminar audiences at Bocconi (CEPR/AMID Development Conference), London School of Economics (SERC Conference, STICERD and Economic History seminars), Oxford (CSAE Conference, CSAE and Economic History seminars), Cambridge (RES Conference), George Washington University, University of Virginia, International Food Policy Research Institute, Sussex, Paris School of Economics, Carlos Tercero, Namur, Geneva (AEHW), University of Barcelona and Stellenbosch (WEHC) for helpful comments. The financial support of PSE, AMID, STICERD-LSE and ESRC First Grant (RES-061-25-0456) is gratefully acknowledged.

1. INTRODUCTION

What is the impact of colonial investments on long-term development? The existing literature has extensively 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, Naritomi, Soares & Assunção 2012; direct versus indirect rule: Banerjee & Iyer 2005, Iyer 2010) or the identity of the colonizer (Porta et al. 1998; Bertocchi & Canova 2002; Berkowitz & Clay 2003). Fewer studies investigated channels other than institutions. Glaeser et al. (2004) argued that human capital matters more than institutions and that white settlers brought with them their know-how rather than better institutions. Another channel could be investments directly realized by the colonizers, as argued by Huillery (2009) and Huillery (2011). Banerjee & Iyer (2005) and Dell (2010) argued that colonization has influenced public investments but they either look at current investments or they use proxies for public investments such as literacy and schooling as in Dell 2010. The effect of colonial investments in physical capital has been largely ignored.[1]

The case of railways in Sub-Saharan Africa is particularly strong. First, railways were often the single largest expenditure item in colonial budgets. In our case study, Ghana, for example, railway expenditures accounted for 31.4% of total public expenditure in 1898-1931.[2] Building the network represented a significant proportion of government investments - about two thirds of government development expenditures. Recurrent education expenditures in contrast amounted to a meagre 4%. Second, railways are an uttermost "colonial" transportation technology. About 88% of total railway mileage in sub-Saharan Africa was built before independence. These railway lines were built by Europeans for Europeans and served various purposes. Military domination, against natives or other colonial powers, was given as motivation in 35.5% of the cases. Mining was mentioned in 36.0% of the cases (with the support of European mining companies) and cash crop agriculture in 42.4% of the cases (with the support of White settlers and farmers).[3] Yet these lines also affected the economic lives of Africans by permitting spatial economic integration. Lastly, public investments in transportation infrastructure are highly localized. As such, they are likely to modify the economic geography of colonies, with long-term consequences on spatial development.

In this paper we investigate the impact of rail construction on development in colonial and post-colonial Ghana. Two railway 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 effect 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 identification strategies to ensure our effects are causal. We argue that the placement of the railway lines was exogenous to cocoa production and population growth. The first line _ the *Western line* _ connected the European owned gold mines to the coast and was then extended to the hinterland city of Kumasi for military domination. The second line _ the *Eastern line* _ connected the capital city of Accra to the network through Kumasi. First, we use GIS data on transportation networks in 1900 to show that pre-railroad transportation costs were prohibitively high limiting cocoa production to a narrow coastal strip. Railways were thus essential to the colonization of the hinterland. Second, the parallel trend assumption is verified using data before 1901. Third, having data at a very fine spatial level (11x11 km gridcells) allow to compare observations that are identical in terms of observables and unobservables. We confirm that cells in the 0-50 km range are all similar in terms of observables, but we find much stronger effects for cells just along the railway lines. Fourth, there are no effects for a set of placebo lines that were planned but not built for various exogenous reasons we discuss. Fifth, given cocoa trees take five years to bear produce, we also check that there are no effects for lines that were built but not in time to influence cocoa production in 1927. Sixth, cocoa production was mentioned as one of the motivations of

[1] Donaldson (2010), Burgess & Donaldson (2010) and Burgess et al. (2011) also analysed colonial investments in transportation infrastructure (railways and roads) but they did not investigate their effects on long-term development.

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

[3] For each railway line in Sub-Saharan Africa, we know when it was built and the main motivation behind the construction of the line. More details can be found in Data Appendix A.

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. In the event that the placement was not exogenous, we verify that our results are robust to employing various techniques used in the literature. We instrument rail connectivity by straight lines between the two coastal 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 to railroad. This strategy which echoes the works of Michaels (2008) and Banerjee, Duflo & Qian (2012). We also apply nearest neighbor matching and only use placebo grids - grids that would have been connected if the placebo lines had been built - as a control group.

Second, transportation infrastructure investments had large welfare effects for Ghanaians during the colonial period. We find a strong effect of railway connectivity on cocoa production, population and urban growth in 1901-1931. We argue that transportation costs decreased along the railway lines, which made cocoa production for export markets profitable. The railway effect on rural growth is explained by a *labor* effect, the fact that cocoa cultivation requires more labor in cocoa-producing villages. Second, the railway effect on urbanization is explained by a *trade* effect, the fact that cocoa-producing villages use towns as trading stations, whether these are large railway stations or smaller towns further away from the lines. Cocoa farmers export cocoa to the rest of the world and import foreign consumption goods in exchange. We also examine the welfare implications of rail construction. We follow the social savings approach of Fogel (1964) and find that railways account for 9% of GDP using cocoa only. An alternative approach is to show that railways have caused around 30% of cocoa production in 1927, which represents 4.5% of GDP implying a social return of around 20-30% on capital outlay. We find that half of the surplus went to Ghanaians while the other half was captured by the colonizer. Colonization meant both extraction and development in this context.

Third, we examine the long-term effects of railway construction. Railways have become obsolete in the 1970s, due to poor management, lack of maintenance and the competition of roads. Railway goods and passenger traffic collapsed after 1974, and railways now transport three times less than what they could do at independence. Besides, Jedwab (2011) describes how cocoa production has disappeared from the old cocoa-producing areas due to the shifting cultivation process characteristic of this crop. In other words, the cells along the railway lines have lost their initial advantage in terms of transportation infrastructure and cocoa cultivation. Yet we find that railway cells are more urbanized, have larger manufacturing and service sectors and better infrastructure in 2000. We investigate the dynamics of path dependence using panel data on population and urban growth in 1901-2000. We find that the railway cells did not lose their initial advantage in terms of urbanization. We investigate the factors of urban persistence and explain that demographic growth and capital accumulation (manufacturing, infrastructure investments) account for this result. Thus, colonial investments had long-term effects on development.

Our focus on the economic impact of railways also connects with the literature on transportation infrastructure, trade and development. Recent research has confirmed that trade has large positive effects on income (Acemoglu, Johnson & Robinson 2005, Feyrer 2009, Donaldson 2010, Donaldson & Hornbeck 2011). Similarly, research has suggested that the lack of intercontinental and intracontinental trade integration is a determining factor of African underdevelopment (Rodrik 1998, Johnson, Ostry & Subramanian 2007, Buys, Deichmann & Wheeler 2010). In this regard, the literature mentions the conjunction of bad geography and poor infrastructure as the main obstacle to trade expansion in Africa (Radelet & Sachs 1998, Buys, Deichmann & Wheeler 2010). African countries are underprovided with transport infrastructure and international organizations recommend massive investments in Africa (Fay & Yepes 2003, African Development Bank 2003, African Union 2006).[4] Despite this recent interest in transportation projects, little is known on their economic effects and more research is needed. A first strand of the literature looks at the macroeconomic impact of transport infrastructure on trade and development (Radelet & Sachs 1998, Limão & Venables 2001). Those studies find that better infrastructure diminish trade costs, especially for

[4]In 2005, Sub-Saharan Africa had 0.002 km of railroad track per 1,000 sq km and Ghana had around 4 km, while China had 7 km, India 21 km, the United States 25 km and Europe 51 km per 1,000 sq km (World Bank 2010). Turning to roads now, Sub-Saharan Africa had 85 km of roads per 1,000 sq km and Ghana had 239 km, while China had 201 km, India 1,115 km, the United States 702 km and European countries 1,377 km per 1,000 sq km. Magnitudes are similar if the lengths of both networks are standardized by population instead.

landlocked countries, with a positive impact on exports and growth. As infrastructure is endogenous to economic conditions, one cannot be sure such effects are causal. A second strand of the literature has focused on the impact of rural roads on neighboring communities (Jacoby 2000, Dercon et al. 2008, Jacoby & Minten 2009, Mu & van de Walle 2011). Those studies find that rural roads reduce poverty in connected villages, by integrating labor and goods markets, thus providing new economic opportunities to their inhabitants. For instance, farmers obtain higher profits thanks to cheaper agricultural inputs and higher farmgate prices for their crops. Yet, this literature also faces identification issues. A last 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 for connected areas.[5]

Our findings also advances the literature on the historical roots of African underdevelopment. Nunn (2007) proposes a model in which slave trade and colonial rule have caused African countries to switch from “one equilibrium with secure property rights and a high level of production" to another with “insecure property rights and low levels of production". Gennaioli & Rainer (2007) and Michalopoulos & Papaioannou (2012) emphasize the role of pre-colonial centralization in long-term development. Nunn (2008), Nunn & Wantchekon (2011) and Nunn & Puga (2012) investigate the economic impact of slave trade, while Acemoglu, Johnson & Robinson (2001), Huillery (2009), Huillery (2011) and Michalopoulos & Papaioannou (2011) study the effects of colonial rule, whether colonial institutions, White settlement, colonial public investments or the fact that borders were arbitrarily drawn. This paper shows how colonial infrastructure investments one century ago have shaped the economic geography of Ghana. This is also 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 railway cells, having lost their natural advantages, are in decline. We argue that railway cities have persisted as a result of demographic growth and capital accumulation.

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

2. BACKGROUND AND DATA

We discuss some essential features of the Ghanaian colonial economy and the data we have collected to analyze how railway construction has contributed to booming cocoa production, population and urban growth. Appendix A contains more details on how we construct the data.

2.1 New Data on Ghana, 1891-2000

In order to analyze the effect of railway construction on trade and development in Ghana, we have constructed a new data set on 2091 grid cells of 0.1x0.1 degrees (11x11 km) from 1891 to 2000. We choose a high resolution grid because we have very precise GIS data on railways, cocoa production, population and urbanization. We obtain the layout of railway lines in GIS from *Digital Chart of the World*. We then use various documents to recreate the history of railway construction. For each line, we know when it was started and finished, and when each station was reached and opened. From the same sources, we know lines that were built but not planned. For each real or placebo line, we create cell dummies equal to one if the Euclidean distance of the grid cell centroid to the line is 0-10, 10-20, 20-30, 30-40 or 40-50 km. Our main analysis focuses on railway lines in 1918. We also create a dummy equal to one if the cell contains a

[5]Those studies are more convincing in terms of identification. Michaels (2008), Banerjee, Duflo & Qian (2012) and Atack et al. (2010) 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 railway lines that were approved but never built.

railway station in 1918. We proceed similarly to construct a GIS database on transportation networks in 1901 (waterways and forest tracks) and motor roads in 1930.[6]

The data on cocoa land suitability was recreated 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 soils for cocoa cultivation. It is *very highly suitable* if more than 50% of its area consists of first class or second class ochrosols, the best two types of forest ochrosols. Production data was collected from two historical maps and we use GIS to calculate the amount of cocoa production (tons) for each cell in 1927 and 1950. Production was almost nil around 1900, and we proceed similarly to recreate cell production in 1901. We also have data on cocoa tonnages brought to each rail station in 1918.

To obtain urbanization figures, we construct a GIS database of localities with more than 1,000 inhabitants using census gazetteers. Using the 1,000 threshold, Ghana had respectively 144, 143, 353, 437, 627, 1110, 1262, 1895 and 2975 towns in 1891, 1901, 1921, 1931, 1948, 1960, 1970, 1984 and 2000. Since our analysis is at the grid cell level, we use GIS to recreate urban population for each cell-year observation. While we have exhaustive urbanization data for the whole country, we only have consistent population data for the South of Ghana in 1901 and the whole territory in 1931 and 1970-2000. The 1901 census was exhaustively conducted and geospatialized only in the South. We then have population data at a very fine spatial level from the 1931, 1970, 1984 and 2000 censuses. We reconstruct rural population by subtracting urban population from total population.

2.2 The Railway Age

2.2.1 Railroads Built

Improvements in transport infrastructure are typically endogenous, driven by the economic potential that would justify them. Hence, a simple comparison of connected and non-connected cells is misleading, and likely to overstate the output created by it. The railway age in Ghana provides us with a quasi natural experiment which we use to identify the effect of reduced transport costs on development. This summary draws on Gould (1960), Tsey (1986) and Luntinen (1996). Figure 1 and Figure 2 show the geographic location of the mentioned lines.

After the defeat of the Ashanti Kingdom in 1874, the British consolidated their power and established the Gold Coast colony in the South. They later extended their domination to the whole country in 1896. Improving transport infrastructure was on the agenda, mainly to permit military domination (against natives or other colonial powers) and boost trade historically constrained by high transport costs. Draft animals could not be used due to the tse-tse fly transmitting trypanosomiasis, which is deadly for livestock. Ghana also lacked navigable waterways, the Volta river being the only noteworthy exception. Headloading was the main means of transport, although palm oil and cocoa beans could also be rolled in barrels along the few forest tracks. Owing to the thick primary forest in Southern Ghana, there were only a few well-cleared tracks and villages along those paths had to maintain them. Railroads were the transport technology of the time, but the British had to choose between a western route, a central route or an eastern route.

The first line to be built followed the western route. Strong interest groups of British capitalists lobbied to connect the gold fields in the hinterland. Mines needed heavy machinery, large quantities of firewood (or coal) and workers from other regions. Headloading made gold production prohibitively costly. The colonial administration gave in to the pressure, turning down alternative lines, for which railroad surveys attested a greater potential for agricultural exports (mostly palm oil) and benefit to the country in general. Appendix B describes how the of Governorship of William E. Maxwell (1895-97) was instrumental in this evolution: Maxwell, who had previously worked in the Malay States where railways served the tin mines, thought that the same model should be applied to Ghana with gold and other lines could wait. There were

[6]There were two types of roads at that time: class II (*roads suitable for motor traffic but occasionally closed*) and class III (*roads suitable for motor traffic in dry season only*).

then military reasons to connect the Ashanti capital Kumasi. The British fought over four wars before they annexed the Ashanti Kingdom in 1896. An uprising in 1900 reinforced the perceived threat. The railroad was meant to allow to quickly dispatch troops. The construction of the railroad 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 indeed represents the shortest distance from the coast to the gold fields. 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 at least two thirds of the line's traffic (in volume) on average in 1904-1912. Cocoa did not play any role in the choice of location, but the line had a strong effect on cocoa cultivation, as argued by Tsey (1986, p.303-306). Cocoa production increased from 0 tons in 1904 to 1,934 tons in 1910 and 19,191 tons by 1915.

The second line to be built followed the eastern route. Colonial Governors had long favored the central route to build a second line (e.g., from Saltpond or Apam, see Fig. 2), but a series of unexpected events led to the Governorship of John P. Rodger (1904-1910) who thought that the capital Accra had to be the terminus of this second line to the interior (and ultimately 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 possibility to develop tourism on the Kwahu Plateau around Abetifi (see Fig. 2). Construction started early in 1909 and the line reached Mangoase in late 1909. However, serious floodings 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 then had to stop due to wartime shortages, and Kumasi was only connected in 1923. The issue with this line line is to what extent its placement was exogenous to cocoa cultivation. Cocoa originally spread out in the Eastern province from Aburi Botanical Gardens, where the British sold cocoa seedlings at a very low price (see section 2.2.3 below). Provincial production was already growing before the Mangoase line was built 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 bear produce, production before 1913 (ca. 30,000 tons) cannot be imputed to the line.[7] As soon as the line was opened, cocoa farmers decided to use it to transport cocoa, which immediately accounted for 62.7% of the line's traffic in 1912-1913 (this share decreased over time). That is why it will be important to show that: (i) transport costs were prohibitively high before, so that production would have remained constrained to pre-railroad levels, (ii) we find similar effects for the Western and Eastern lines, (iii) results are robust to controlling for the spatial diffusion of cocoa from Aburi, (iv) no positive effects are found for the extension Tafo-Kumasi (1923) since cocoa trees take five years to bear produce, as well as for other lines that were proposed for the central and eastern routes (see the description of placebo lines below), (v) results are robust to instrumentation (using the distance to the straight line Accra-Kumasi as an instrument) or nearest neighbor matching (using placebo cells as a control group, as discussed below).

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 railways. 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 to motor traffic throughout the year (Gould

[7] The line was officially opened to traffic in 1912, but it reached Mangoase three years before in 1909. There is some evidence that many farmers went there as soon as 1909 to grow cocoa, expecting the railway to be opened that year. Given cocoa trees take five years to bear produce, we should not see any effect on total production before 1913.

1960). Until the late 1920s, railways were by far the best transport technology for Southern Ghana.

Using a GIS map of transportation networks in 1900 (see Figure 3), we estimate for each cell the minimal transport cost of one ton of cocoa (£) to any coastal port. We account for topography when calculating the minimal transport cost as the topographic distance is always higher than Euclidean distance. From Cardinal (1931), we get an estimate of the production cost of one ton of cocoa, around £16 on average.[8] Given the coastal producer price was £31.3 on average in 1920-29, cell production is only profitable if 31.3 - 20.9 - transport costs > 0, or transport costs < £10.4.[9] As in Donaldson & Hornbeck (2011), the reduction in transport costs permitted by rail construction is likely to 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 3 displays the feasible and infeasible regions. Clearly, pre-railroad transportation costs were prohibitively high limiting cocoa production to a narrow coastal strip. Railways were thus essential to the colonization of the hinterland.

From 1912 on, the share of cocoa transported by rail was around 80% (see Figure 4). 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, so eagerly that foodstuffs had to be imported." The main effect of railroads was thus to make cocoa cultivation even more profitable. Roads were first complementary to the railways as they were feeders to them. The first lorry was imported in 1903, but there were only two lorries in the colony in 1914 (Luntinen 1996). Besides, roads were of poor quality until 1924 when the government started the "Tarmet Program" which made them suitable to motor traffic throughout the year (Gould 1960). Roads became serious competitor for the railway and opened new areas to cocoa cultivation. Even if no railway had been built, roads would have permitted the cocoa boom. But our goal is not to compare the respective impacts of railways and roads. We focus on the railway age in 1901- 1931 because it provides us with a natural experiment to identify the effect of reduced transport costs on development.

2.2.3 Placebo Lines

The British had to choose between a western route, a central route or an eastern route, and various private and government lines were proposed 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. Placebo control mimics the tested treatment in all ways except the treatment. Appendix B gives a brief background of each placebo line.

Five alternative railroad routes were proposed before the first line was actually built. These lines were driven by the same model. The aim was to ensure military domination and increase exports (mostly palm oil). Judged by observables, the proposed lines were influenced by population density, soil quality, altitude in a similar way as the actual lines built. But the proposed lines are not completely identical to the lines built. We argue in Appendix B that these 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 line Cape Coast-Kumasi (1873) was supposed to link the capital Cape Coast to Kumasi so as to send troops to fight the Ashanti. The project was dropped because the war broke out too rapidly. Second, Governor Griffith advocated the construction of central line from Saltpond to Oda and Kumasi (1893) so as to tap palm oil areas and link the coast to Kumasi. When he retired in 1895, he was replaced Governor Maxwell who favored the mining lobbies and the Western line. Third, even if the construction of the Western line was about to start, Governor Maxwell still thought that the colony needed a central line. There were two competing projects, Apam-Oda-Kumasi (1897) and Accra-

[8]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 produce, assuming a farm lifetime of 30 years.

[9]£10.4 is an upper bound as there are various costs that we cannot account for when a subsistence farmer decides to launch herself into cocoa production: risk premium due to risk aversion, losses during transport due to theft or rotting, etc.

Oda-Kumasi (1897), and the main issue was to know whether Apam or Accra should be the terminus of the new line. A conference was to be held in London in late 1897 to discuss the various proposals but unexpectedly Governor Maxwell died before reaching London. Fourth, Maxwell was replaced by Governor Hodgson who favored Accra, mostly because it was the capital and chief trading town. However, he thought that the Accra line should be built to Kpong on the Volta river (1898), so as to boost palm oil production and cotton cultivation there. The construction of the line was approved in 1903 but Governor Nathan retired in 1904 before works even begun. Fifth, Governor Rodger, who replaced Governor Nathan, did not see any interest in a line to Kpong and he proposed the Eastern line.

The rail network was subsequently expanded. Hence, we also consider lines that were actually built, but not in time to affect cocoa 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, peaks after 25 years, and declines thereafter. Hence, for observing an impact on cocoa production in 1927, farmers must have grown cocoa trees before 1923. The extension of the Eastern line from Tafo to Kumasi (1923) is a good counterfactual for the Accra-Tafo line (1918). Another line was also built from Huni Valley to Kade in 1927. It connected the diamonds mine at Kade and was supposed to encourage cocoa, kola, palm oil and timber exports. So we verify that we find 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" (Tosh 1980; Austin 2007). Yet, as cocoa is produced by consuming the forest, this success has led to deforestation. Cocoa farmers go to a patch of virgin forest and replace forest trees with cocoa trees. Pod production starts after 5 years, peaks after 25 years, and declines thereafter. When cocoa trees are too old, cocoa farmers start a new cycle in a new forest (Jedwab 2011).[10]

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 soon as 1911. Figure 4 shows aggregate cocoa production and the export share of cocoa from 1900 to 1927. Figure 5 shows grid cells that are suitable or highly suitable for cocoa cultivation and cocoa production in 1927. Cocoa production originally spread out in the Eastern province from Aburi Botanical Gardens, where the British sold cocoa seedlings at a very low price (Hill 1963, p.173-176).[11] As Ghanaian farmers realized how much profit they could make out of cocoa, more and more people specialized in the crop. The Eastern line was instrumental in opening new land to cocoa cultivation. Production also boomed around Kumasi when the Western line was opened in 1904. Why did production boom in Ashanti and not in the South-West, around Sekondi? Transportation costs were certainly lower there as it was closer to the coast. But the South-West of Ghana is characterized by poor cocoa soils and too much rainfall.[12] The railway line to Kumasi thus made the Ashanti more competitive than the South-West for the development of cocoa cultivation.

Ghana has experienced dramatic population growth after 1901, due to rising livings of standard, international migrations and large-scale health campaigns organized by the British colonizer. Its population increased from 1,9 millions in 1901 to 3,2 millions in 1931. 31.9% of this growth took place in the Gold Coast Colony, 10.6% in Ashanti, 21.6% in Northern Territories and 9.0% in British Togoland (see Figure 1). Population was 18.9 millions in 2000. Figure 6 displays total population for Southern cells in 1901, 1931 and 2000. The comparison with Figure 5 suggests that rural population increased in areas where cocoa production boomed. Austin (2007, p.8-14) describes how the labor-land ratio increased as a result of migration related to cocoa and how some areas in the forest zone were already approaching the "critical population density" for foodcrop growing.

[10]Removing forest trees alters the original environmental conditions and replanted cocoa trees die or are much less productive.

[11]The British established the Government Botanical Gardens in Aburi in 1890, because of its health climate and its proximity to the capital city, Accra.

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

While Ghana was almost unurbanized at the turn of the 20th century, it is now one of the most urbanized countries in Africa. It has started its urban transition earlier than most African countries, due to the cocoa boom (Jedwab 2011). Defining as urban any locality superior to 1,000 inhabitants, its urbanization rate increased from 23.5% in 1901 to 48.6% in 1931 and 68.5% in 2000. The two largest cities were Accra, the national capital, and Kumasi, the hinterland capital. Altogether, they have accounted for 9.5% of total urban growth in 1901-1931. 41.4% of it has come from the Gold Coast Colony, 21.5% from the Ashanti, 24.0% from Northern Territories and 3.6% from British Togoland. Thus, around 66.5% of urban growth has come from areas suitable for cocoa cultivation. 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. But there were also trading centres 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 mining, modern transportation and cocoa production (see Dickson 1968, p.246-261). Many towns grew because they were cocoa buying centres, the homes of wealthy cocoa farmers or market towns where they would spend their income.

3. EMPIRICAL STRATEGY

Having data on railway connectivity, cocoa production and population and urban growths at the grid cell level in 1901 and 1931, we test if connected cells experience a boom in cocoa production and population and urban growths. We explain the various strategies we implement to obtain causal effects.

3.1 Main Econometric Specification

The main hypothesis we test is whether railway connectivity drives cash crop production and population and urban growths before 1931. We follow a simple difference-in-difference strategy whereby we compare connected and non-connected grids over time. We run the following model for cells c and years $t = [1901, 1931]$:

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

where our dependent variable is cocoa production (tons) of cell c in year t. $Rail_{c,t}$ are cell dummies capturing railway connectivity: being 0-10, 10-20, 20-30, 30-40, 40-50 km away from a railway line in 1918 and having a railway station in 1918. These dummies are equal to zero in 1901. Besides, we know cocoa tonnages brought to each station in 1918. We then run a second model:

$$Pop_{c,t} = \alpha' + Rail_{c,t}\beta' + \gamma'_t + \delta'_c + \nu_{c,t} \tag{2}$$

where our dependent variable is population of cell c in year t. We expect railway 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 rail dummies on population. If that is the case, it means that railway connectivity has an effect on population growth through more cocoa production along the railway lines. There could still be an independent railway effect of population, so our goal is not to instrument production with railway connectivity. We just want to highlight one of the mechanisms at play. We can look at rural and urban growths to understand the nature of population growth.

We have a cross-section of 2091 cells. Our main analysis is performed on the restricted sample of suitable cells. We run the risk of just comparing the Southern and Northern parts of Ghana if we use the full sample. We also restrict our sample to those cells for which we have data on the cell population distribution in 1901 as this data is only available for a set of Southern cells that year.[13] We end up with 542 observations and we believe such restrictions give us very conservative estimates. We will show later that our results are robust to removing such restrictions. We privilege OLS regressions and the issue is whether railway

[13] As explained in data appendix A, localities with less than 1,000 inhabitants were exhaustively surveyed and georeferenced only in those parts of the country.

construction was endogenous to cocoa production and population and urban growths. We argue in Section 2.2.1 that this was not necessarily a concern, but we now describe the various tests we perform to ensure these effects are causal.

3.2 Exogeneity Assumptions and Controls

We could just assume that the placement was exogenous. First, even if the development of cocoa production could have been instrumental in the construction of the Eastern line, the Western line was clearly 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 only, production could have remained nil without modern transportation infrastructure because transport costs were prohibitively high before. The analysis in Section 2.2.2 provides evidence that production would have been constrained "spatially" without railroads. 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 population and rural populations. We add physical geography variables such as the shares of suitable, highly suitable and very suitable soils for cocoa cultivation, or the mean and standard deviation of altitude (m) and average annual rainfall (mm) in 1900-1960. We control for economic geography by having Euclidean distances (km) to Accra, Kumasi, Aburi, a port in 1901 and the coast and dummies for bordering another country or the sea.[15] Euclidean distance to Aburi is an important control since it ensures that our effects are not contaminated by the spatial diffusion of cocoa cultivation from Aburi Botanical Gardens, as discussed in Section 2.2.1. Fourth, we have data for 11x11 km grid cells. This means we compare locations that are very close to each other and are therefore very unlikely to differ in terms of observables and 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, which were used as spatial units in the literature, are 4,300 and 78,977 sq km respectively. Amongst the group of cells that are less than 50 km from the line, we verify that they are all not significantly different from each other for the set of variables described above.[16] Therefore, if placement is truly exogenous, we expect the effect to strongly decrease as we move away from the line, which is what we find in our analysis.

3.3 Parallel Trends

First, if the placement or the timing is exogenous, we should not observe any effect for the railway cells before the railway lines were indeed built. Having data on urbanization in 1891, we can directly test the parallel trends assumption.[17] We run the same model as model (2) except we consider urban population in 1891 and 1901. There are no significant positive effects of railway connectivity in 1891-1901, while we show later that we have strong positive effects of railway connectivity in 1901-1931.[18] Second, it is also informative to look at urban growth and cocoa production in 1931-1948, after the two lines are built. We run the same model as model (2) except we now consider urban population in 1948. We proceed similarly for cocoa production in 1950, for which we also have data at the grid cell level. Again, we do not find independent effect of railway connectivity when we control for the effect on production in 1927 (results not shown but available upon request).

[14] There were five mines in 1931: three gold mines, one diamond field and one manganese mine. Mineral exports amounted to 24.2% of exports in 1930 and the number of Africans engaged in mining was 12,048. Cocoa and mineral products accounted for 94.5% of exports in 1930.

[15] Those dummies also capture the fact that border cells are by construction not entirely contained in the territory of Ghana.

[16] We regress each variable on the railway dummies using the 40-50 km cells as the omitted group. There are no significant differences, except for rural population in 1901 and the probability of having a mine in 1901 for the 0-10 km cells, but these effects are small.

[17] Cocoa production was nil in 1891 and almost nil in 1901 (ca. 1,000 tons), so we do not test it for cocoa production. We have no data on rural population in 1891.

[18] We get the following coefficients (p-values) for 0-10, 10-20, 20-30, 30-40 and 40-50 km rail dummies in 1891-1901: 90 (0.756), 40 (0.880) and -155 (0.629), 70 (0.714) and 9 (0.959).

3.4 Placebo Regressions

As explained in Section 2.2.3 and shown in Figure 2, five major railway lines were planned but never built: Cape Coast-Kumasi (1873), Saltpond-Oda-Kumasi (1893), Apam-Oda-Kumasi (1897), Accra-Oda-Kumasi (1897) and Kpong-Kumasi (1898). Two lines were built after 1918: Tafo-Kumasi (1923) and Huni Valley-Kade (1927). The expansion of cocoa cultivation was mentioned as one of the objectives of the two latter lines, but they were unlikely to have any impact on cocoa production in 1927 and urbanization in 1931 given cocoa trees take five years to bear produce. For each planned but not built or not built yet line, we create a placebo treatment dummy equal to one if the cell is less than X km from the placebo line (counterfactually connected). We run the same model as model (2) except the treatment dummies are the placebo dummies. First, we expect no effect on both cocoa production and population growth for cells that would have gained access if the placebo had been built before 1918. Second, one issue here is that some of the placebo lines “intersect" with the area of influence (e.g., 0-20 km) of the two existing railway lines, so that there might be correlation between the treatment dummies and the placebo dummies. That is why we verify that there are no effects for the segments of these lines that do not intersect with the existing railway lines. Third, we also only compare the placebo cells with the other control cells, by dropping the railway cells.

3.5 Instrumentation

We can instrument the treated cells with the distance from the straight lines between the two main ports on the coast _ 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 railways with the distance from the straight line joining two major cities, exploiting the fact that transportation networks tend to connect large cities. This is a valid instrument as long as the distance from the straight line is not correlated with any variable that we do not control for but also affects the outcome. As argued in Section 2.2.1, the Western line was built to link Sekondi to the mining areas of Tarkwa and Obuasi and was later extended to Kumasi for military domination. It went through dense tropical forest and it can be argued that the random location of the mines explained why this interior line was indeed 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 of Ghana would be connected at one point.

3.6 Matching

We can compare connected cells to non-connected cells, or connected cells to cells that would have been connected if the placebo lines had been built before 1918. This guarantees that treatment and control cells are similar in terms of both observables and inobservables, in the spirit of nearest neighbor matching. The only difference between the two comes from the fact that connected cells were effectively treated. We can alternatively consider cells counterfactually connected to lines that were planned but not built or lines that were not built yet in 1918 or both. The lines that were planned but not built mostly followed historical trade routes and had more population in 1901 along their potential layout. We regress each available control on a dummy equal to one if the cell is less than 20 km from a 1918 railway 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 an downward bias. It is not obvious in which direction we should expect coefficients to be biased. Second, when compared to all placebo cells, treated cells have a larger rural population (upward bias) but are farther from main cities and have worse cocoa soil quality (downward bias). Third, we can compare treated cells with each placebo line individually, as some of them could prove a “better" counterfactual. For example, when compared to cells along the placebo line

Cape Coast-Kumasi (1873), treated cell are “worse" (soil quality, altitude and distance to Accra or the coast) or similar across all dimensions.[19] Using these cells as a control group should lead to a downward bias, which is not problematical at all as it only gives us a conservative estimate of the effect we wish to measure. We test the various control groups and confirm the results are not affected at all.

4. RESULTS

In this section, we display the main results, examine their robustness and investigate the mechanisms behind the effects of railway connectivity on cocoa production and population and urban growth.

4.1 Main Results

Table 2 contains our main results for cocoa production and population growth, distinguishing urban and rural growths. Column 1 reports the results for regression (1), while columns (2)-(13) display the results for regression (2). All regressions include gridcell fixed effects, year fixed effects and controls, but unconditional results are discussed in Table 3. We find a strong effect of railway connectivity on cocoa production, but this effect decreases as we move further away from the line and is nil after 40 km (see Col. (1)). There is a strong effect of railway connectivity on population growth but only for those cells less than 20 km away from the line (see Col. (2)). This indicates that 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 then has a strong effect on population growth (see Col. (3)). The remaining rail effect is picked up by the cell dummy for having a railway station in 1918 (see Col. (4)). But the latter effect also becomes lower and non-significant when we include the amount brought to the station in 1918 (see Col. (5)). This means the railway lines have a strong effect on population growth, and that this growth is coming from opportunities in the cocoa sector, and other sectors if there are intersectoral linkages.[20]

Railway connectivity has thus two positive effects on population growth, which we investigate by looking at rural growth (see Col. (6)-(9)) and urban growth (see Col. (10)-(13)). We call the first effect (1.53***) the “labor effect", as more cocoa being produced requires more labor. The comparison of columns (5), (9) and (13) indicates that most of the “labor effect" takes place in villages (1.15***) compared to towns (0.34*). This is logical as cocoa is produced on farms surrounding cocoa-producing villages (see Jedwab 2011). We call the second effect (0.86*) the “trade effect", as more cocoa being transported necessitates larger railway stations. The comparison of columns (5), (9) and (13) shows that the “trade effect" occurs in towns (0.91**) rather than villages (-0.06).

How do the labor and trade effects vary if we consider various size categories of towns: 1,000-2,000, 2,000-5,000 and 5,000-+? We run the same model as in Column (10) except we now consider the urban population for each size category of towns so as to decompose the urban effect (2,190**). We find no effect for 1,000-2,000 towns, a small effect for 2,000-5,000 towns and a strong effect for 5,000-10,000 towns. If we run the same model as in Column (13), we find that the labor effect is mostly explained by 2,000-5,000 towns and many of them did not exist (< 1,000 inh.) in 1901. The trade effect is mostly explained by 5,000-10,000 towns and many of them already existed (> 1,000 inh.) in 1901. To sump up, railway connectivity has induced a boom in cocoa production, which had various effects on population growth: more cocoa-producing villages till 30 km, small cocoa-producing towns till 20 km and larger trading stations (railway stations) till 10 km.

[19] Treated cells have less rainfall but the effect is small: 65 mm against an average of 1477 mm.

[20] Including urban population in regression (1) does not change the effect of railway connectivity on cocoa production. This confirms that the relationship is not from railway connectivity to population and then to cocoa production.

4.2 Alternative Identification Strategies and Robustness

Table 3 displays the results when we implement the various identification strategies and test for robustness. Column (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 cocoa production and the 0-20 km dummy for population as there are no effects beyond. Table 4 shows results for the placebo lines.

Western line vs. Eastern line. Column (2) of Table 3 shows that we have much stronger effects for cocoa production for highly suitable cells, so the railway effect comes from the interaction with suitability. When comparing the results for the Western and Eastern lines (see Col. (3) and (4)), we find that we have lower effects for the Western line, but this is explained by the fact that the Western line goes through many poorly suitable cells. If we restrict the analysis to highly suitable cells only, the effects are not significantly different for each line.

Placebo Regressions. Table 4 displays the results of the placebo regressions. Panel A, B and C show the results for cocoa 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 that we test whether there are any positive effects just along the placebo lines. The only concern is cocoa production for Accra-Oda-Kumasi 1897 (see Col.(4)). But this effect is explained by the fact that this placebo line "intersect" with the area of influence (0-20 or 0-40 km) of the Eastern line. That is why we verify that there are no effects for the segments of placebo lines that do not intersect with the existing railway lines, as we show in Panel B and E. In Panels C and F, we drop these observations that are less than 20 km from a railway line, so that we just compare the placebo cells and the other control cells. Except for Accra-Kpong 1898, coefficients are very small and always not significant.

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

Matching. Columns (6) and (7) of Table 3 show that coefficients are even higher when control cells are all the placebo cells or just the cells along the Cape Coast-Kumasi placebo line (< 20 km). Coefficients are even higher now. Results are also robust to using the other lines (results not shown but available upon request).

Instrumentation. In Column (8), we instrument the railway 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 B and 63.7 in Panel B) and results are similar to the main results.

Robustness. We now perform a few robustness checks. First, results are similar if we use the Euclidean distance to railway stations to create the railway dummies (see Col.(9)), if we drop the controls (see Col. (10)), if we drop the nodes of the network _ only Kumasi when we restrict our sample to suitable cells _ (see Col. (11)), if we use the full sample and not just suitable cells (see Col. 12) and if we use a log-linear functional form (see Col. (12)). We have no measure of total population for the whole country in 1901, but we verify the main results are the same for urban population at least (result not shown but available upon request).R Second, results are robust if we control for transport costs given transportation networks in 1900 and/or transportation costs by road in 1922 (result not shown but available upon request). We use roads in 1922 because they are more likely to predict production 5 years later, in 1927. 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 thus expect lower effects, unless there is land heterogeneity and farmer heterogeneity (e.g., in terms of risk aversion) and the decrease in transport costs also has an extensive margin effect in this region. We find a strong effect in the 0-20 km range, and there is no significant effect beyond 20 km. Third, there could be some spatial autocorrelation, and we would underestimate standard errors as a result. Results are robust to having Conley standard errors using a distance threshold of 50 km (around 5 cells) or more, or clustering standard errors at a more aggregate spatial level, such as provinces in 1931 or 2000 (result not shown but available upon request).

Results on infrastructure. For each gridcell, 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 model as equation (2) to test whether railway cells got better infrastructure over time. We find that railway connectivity has a strong positive effect 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. Then, these effects are strongly reduced when we control for population. This confirms that railway connectivity had a strong effect on population, and public goods had to be created as a result.

Results on height. We have data on Ghanaian male recruits in the British Army in 1888-1960. We know their year and place of birth, the year they were enlisted and their height. We restrict our analysis to those soldiers that have served during World War I or later on. In total, we have height data for 5,447 soldiers across 298 suitable grid cells. Using a regression similar to model (2), with the same controls as before, district-year fixed effects, and individual controls for age, farming, skills and ethnicity, we compare soldiers born in railway vs. non-railway cells (using 40 km as a cut-off) before and after 1923. Better standards of living in railway cells over time should lead to increasing height for native soldiers born just after the cocoa boom. We find that railway cells start with a disadvantage (-0.63 cm), which they more than cancel over time (+1.70* cm). The latter effect is large since it represents four times the change for the sample as a whole over the same period (+0.38 cm).[21] Interestingly, this effect is almost halved and becomes not significant, when we control for cocoa production in 1927, which becomes significant. This confirms that the railway effect on height goes through more cocoa production and better standards of living as a result.

General equilibrium effects. If rail connectivity only reallocates labor (population) across space, overall welfare does not change. First, historical evidence suggests that population growth in cocoa-producing areas was driven by migration, but migration from non-forest areas. The control cells were unlikely to have been negatively affected by the cocoa boom. We verify that non-railway cells that were historically populated before the boom gained population between 1901 and 1931 when compared to other control cells. Results in Panel (F) of Table 4 confirm that most placebo lines did not lose population. Besides, we find that locations that already had a city in 1891 gained population in 1901-1931, when compared to the other control cells. Second, population has boomed in cocoa-producing areas, but more than half of population growth happened in cities. Rail connectivity has thus caused a qualitative change in population, by increasing population density. If urbanization is a measure of economic development, this result indicates that there were large welfare gains. Third, cocoa production was almost nil in 1901, but amounted to 80% of exports in 1927. Using aggregate and export data, we verify that the production of other cash crops was not affected by the cocoa boom. Thus, rail connectivity had a major impact on the development of a new and highly profitable sector. By 1950, Ghana was one of the wealthiest African countries, and the main reason behind this success was the cocoa boom. Using data on production and transport costs for cocoa farmers around 1930, we find that cocoa farmers are 2 times wealthier on average than subsistence farmers. The employment share in the cocoa sector increased from 0 to one third in 1931. So one third of the subsistence farmers have become cocoa farmers and have been able to double their income. Lastly, we find positive railway effects on infrastructure and height, while we find no negative effects for historically developed areas.

[21] 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.

5. DISCUSSION ON RAILWAYS AND DEVELOPMENT DURING THE COLONIAL PERIOD

Although this paper does not make any theoretical contribution, here is the conceptual framework that we have in mind. Assume a country is divided into grid cells and there are no transportation networks. Each cell lives in autarky, and agricultural productivity is such that everyone lives at the subsistence level. Most cells are a tropical forest where human settlement is limited given disease conditions. Cash crops could be produced there but there are no means of transportation to permit their export. For exogenous reasons, some of those cells are directly connected to the rest of the world (the coast) via a new transportation network. Those cells experience a major decrease in the trade costs they face to export their products and import other products. Their comparative advantage in producing cash crops can be exploited and they export them 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 cocoa-producing areas and the coastal ports. Our results are in line with these theoretical intuitions. The export share of cocoa has increased from 0% in 1900 to almost 80% in 1927, and the increase in cocoa production has happened along the railway lines (or along the coast, where there are a few good-quality roads). The production of cocoa beans involves cocoa-producing villages, cocoa-producing towns and trading stations. We now use aggregate data on imports, railway traffic and the employment composition of Ghanaian towns to provide more evidence on channels.

5.1 Railways, Trade and Urbanization

Railway Traffic and Imports. First, we find that trade has accounted for 74% of railway 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 have accounted for 58%, mostly because the Western line was used to import building materials, machinery and coal for the gold mines. In the later period 1919-1931, exports have accounted for 55%, mostly because because cocoa production has boomed on both lines, as well as manganese and timber exports. Second, we can easily link the export structure and railway traffic for export. 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%. Third, 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, although it is not as clear as for exports: food, drinks and tobacco 31%, clothing 11%, construction 12%, fuels 16%, machinery 9% and transport equipment 6%. The analysis above confirms the role of railways in promoting trade integration with the rest of the world.

Railways and Urbanization Why did railway connectivity accelerate urbanization? First, many farmers settle in towns as they offer better conditions of living that villages in the forest. Using data from the 1931 census, we find that 48.5% of the urban male workforce works in agriculture. The census does not distinguish cocoa farmers and other farmers, but this figure indicates that many farmers use towns as their main residence to work on farms outside the city limits.[22] Second, towns serve as trading stations for cocoa exports (transportation to the coast) and food and non-food imports (transportation from the coast). Trade and transport account 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. Third, cocoa farmers are relatively wealthy compared to non-cocoa subsistence farmers. The production of cocoa allowed them to generate a surplus, and the Engel curve implied that they spent more on "urban" goods and services (Jedwab 2011). The analysis above shows that food, drinks and tobacco, clothing and other consumption goods altogether amount to around two thirds of imports. But some of these goods were also produced locally, as we

[22]The criterion to define cities differs from the 1,000 threshold we use. The total urban population we find for Ghana is three times larger than the one reported in the census. This implies that the urban sample in the census includes larger towns, and given larger towns are probably less "agricultural", the analysis above underestimates the employment share of farmers in our urban sample.

find that light manufacturing and services account for 30.9% of urban male employment. This includes 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%).

5.2 Welfare Effects

The railway proofed to be profitable. The private rate of return was 1% in 1903, steadily increased to 9% in 1909, and varied in the period 1910-1928 thereafter around 7%.[23] Chaves, Engerman & Robinson (2010) have estimated the social savings of railroads in the Gold Coast and arrived at a value of 9% of GDP in 1925. The social savings approach is based on Fogel (1964) and is defined as the cost difference to the next-best transportation alternative in the absence of railroads:

$$\text{Social savings} = (c_a - c_r)R \tag{3}$$

where c_r and c_a represent marginal costs of the railroad and the next-best alternative transport respectively; R is the total volume transported by rail. Methodologically, it is best to think of a simple supply demand framework of transportation where cocoa producers represent transport consumers: Under perfect competition (where price p equals c) and completely inelastic demand for transport (a horizontal supply and vertical demand curve), social savings are identical to the change in consumer surplus brought about by the railroad. Where those conditions do not hold, the social savings estimate will exceed the actual increase in consumer surplus (Fogel 1979, Leunig 2010). In particular, the bias increases with price elasticity and $\frac{pa}{pr}$. Hence, for Ghana, the two measures may disagree quite considerably. Demand was highly elastic: our results suggest that railways triggered significant cocoa production inland. Secondly, the railroad was much more efficient than available alternatives in 1900; $\frac{pa}{pr}$ was around 10 when head porterage is considered. In sum, assuming R under c_a implies a deadweight loss that the social savings approach counts as a gain. In the following, we calculate the social rate of return based on cocoa income alone using: (i) the social savings approach, and (ii) a more direct approach estimating producer rents.

5.2.1 Social Savings Approach

As described in section 2.2.1, these were the transport technologies in 1901: headloading, cask rolling, steam launch, canoe and lorries. However, farmers privileged headloading given the lack of good quality roads and navigable rivers. That is why we consider head porterage as the alternative to railroads.[24] Chaves, Engerman & Robinson (2010) assumed costs of 3s per ton mile for head porterage, which is derived from a daily wage of 9d that was paid to forced labor and in poorer regions of West Africa. Our preferred estimate of costs is 5s per ton mile, which is more in line with what contemporaries reported and the relatively high wage for free labor in the Gold Coast colony.[25] The railroad transported a total of ca. 15 million ton miles of cocoa in 1927. Under zero profits freight revenues equal costs. Available evidence suggests that supernormal profits are indeed largely negligible.Lorry transport was a strong competitor, to which the Gold Coast railway responded with price reductions (Gould 1960). Moreover, the private rate of return was only slightly higher than the interest on capital. Nevertheless, relaxing this assumption of perfect competition, allowing for positive profits lowering railroad costs, would not change results, because in our case social savings are overwhelmingly determined by the very high costs of the alternative transport method. We then calculate the hypothetical costs, if the same volume of cocoa was moved by these other transport methods. The cost difference are the social savings, which we find to amount to 8.7%

[23]The interest rate on the loan amounted to 3.5% till the early 1920s, before increasing to 6% (Luntinen 1996).

[24]With the expansion of the road network from the 1920s on, the next-best alternative and biggest competitor to railroads was lorry transport, with costs of 2s3d per ton mile in 1922 and falling. In this exercise, however, we do not consider motor transport as counterfactual, but rather as part of modern transport technology introduced in the Gold Coast colony.

[25]This 5s per ton mile is still a lower bound, as it only concerns headloading along well-cleared forest tracks. It is certainly far more costly if one has to go through the dense tropical forest.

of GDP. Cocoa accounted for 30% of revenues. Applying the costs of 5s per ton mile to all goods would give social savings of about 27% of GDP. This estimate is considerably larger than the 9% from Chaves, Engerman & Robinson (2010).[26] But we also believe that this estimate overestimates the welfare effects, as the deadweight loss is counted as a gain.

5.2.2 Distributional Analysis of the Social Rent

We also estimate the rent generated by rail-induced cocoa production for the society as a whole, and for the Whites and the Ghanaians separately. Using estimates from Table 2 column (1), we find that cocoa production would have been ca. 30% lower without railroads. Given the average export price in 1920-1930, this gives 7.9% of GDP in 1927. Yet this does not take into account the costs of producing and transporting cocoa and the distribution of the gains across the two groups.

Ghanaians were paid the producer price in Accra, which was around 65% of the export price on average in 1920-1930. This export tax was then captured by the colonizer. This implies that Ghanaians received around 5.2% of GDP. When removing the transport costs (railway tariffs) and the production costs (including the opportunity costs), we find that they received 3.1% of GDP. The Whites received the export tax, around 2.7% of GDP. They also received railway tariffs but had to incur the costs of building and operating the railway lines. Our calculations indicate that the rent generated for the Whites was around 3.2% of GDP. In the end, the total rent amounted to 3.1 + 3.2 = 6.3% of GDP and Ghanaians received 49.4% of it while the Whites 50.6% of it. This indicates an almost equal sharing of the rents, with positive gains for Ghanaians. These gains are in line with an increase in per capita GDP and an acceleration of urbanization.[27]

6. COLONIAL RAILWAYS AND LONG-TERM DEVELOPMENT IN GHANA

In this section we document the decline of railways in Ghana and investigate their long-term effects on economic development.

6.1 Quantitative Evidence on the Decline of Railways

By 1930, 500 miles of railroad track had been laid in Ghana, and the rail network transported 759,000 tons (or 1,518 tons per mile) of goods and 1,336,000 passengers (or 2,673 passengers per mile). The network reached its maximum size, 750 miles, in 1957 the year Ghana became independent. From 1944 to 1974, the rail network transported on average 2,400 tons of good and 8,000 passengers per mile of railroad track. Railway traffic collapsed in the late 1970s and early 1980s. In 1984, railways only transported 500 tons of goods and 2,900 passengers per mile of railroad track. Railway traffic never fully recovered and only transported 970 tons of goods and 2,900 passengers in 2000, almost three times less than before the mid-1970s. Similarly, while railways accounted for more than 70% of cocoa transport until the early 1970s, this share decreased to one third in the 1980s and was 7% in 2000. Railways are mostly used to transport manganese and bauxite now, as these commodities are too bulky to be transported by rail.

What did cause the obsolescence of railways in the late 1970s and early 1980s? Luntinen (1996) describes how considerable investments in road infrastructure and underinvestments and management issues in the railway sector caused a significant decline of the latter. First, the fist government of independent Ghana had massively invested in roads, which were deemed as more “modern" and three times cheaper

[26] Besides the low wage rate for carriers, Chaves, Engerman & Robinson (2010) used an inflated figure for GDP. This demonstrates how sensitive conclusions are to measurement issues.

[27] The issue here is why the Whites did not capture the whole rent. First, the Whites could not produce cocoa themselves as the mortality rate of settlers was high in the tropical forests of West Africa. Second, they were taken aback by the magnitude of the cash crop revolution in Ghana and did not have a clear view of the total gains that accrued to the Ghanaians. Third, the colonial strategy in Ghana was both extractive and non-extractive. The Whites did exploit the gold mines and did tax the cash crop sector, but they also thought that they were there to bring civilization, whether christianity or capitalism (through trade).

to build than railways. However it was argued that maintenance costs were much lower for railways, which made them more competitive in the long-run. Second, political and economic instability - the overthrow of Nkrumah in 1966, the succession of military coups and fragile political regimes after 1966, the economic downturn in 1966-1969, the economic crisis in 1974-1983 - had a damaging effect on long-term infrastructure investments and public companies like the Ghana Railway Corporation (GRC). By 1980, track, motive power and rolling stock were in desperate physical condition. Third, there were increasing management issues at the GRC. In 1974, it employed a staff of 15,000, twice more than in 1958 although traffic did not increase. The payroll absorbed 70% of railway expenditure and the GRC had been in deficit since 1966. Luntinen (1996) also discusses how the railway had alienated the Cocoa Marketing Board with poor wagon turn-over. The Board had thus bought lorries of their own to carry the crop.

To conclude, locations along the railway lines have lost their initial advantage in their late 1960s and railway traffic collapsed in the following decade. In addition to that, 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 starts after 5 years, peaks after 25 years, and declines thereafter. When cocoa trees are too old, cocoa farmers start a new cycle in a new forest. Indeed, removing forest trees alters the original environmental conditions and replanted cocoa trees die or are much less productive. (Jedwab, 2011) uses panel data on cocoa production at the district level in 1901-2000 to describe how cocoa production has disappeared from the original regions of cultivation. Production density in the Eastern province of Ghana, 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 of Ghana, along the Western line, peaked in 1964 (12.1) before continuously decreasing afterwards (3.9 on average in 1980-2000). Not only have the railway cells lost their initial advantage in terms of transportation infrastructure, but they have also lost their initial advantage in terms of cocoa production. In these conditions, should we also expect these locations to lose their initial advantage in terms of economic development (e.g., population density and urbanization)?

6.2 Evidence on Path Dependence

We use various sources _ in particular the 2000 *Population* and Facility Censuses _ to reconstruct several measures of development at the gridcell level in 2000. We then test whether railway cells are relatively more developed today. In other words, have colonial investments durably transformed the economic geography of Ghana? We run this cross-sectional regression for suitable cells c (N = 554) in year 2000:

$$Development_{c,2000} = \alpha'' + Rail_{c,1918}\beta'' + X_{c,1918}\zeta + w_c \tag{4}$$

with $Development_{c,2000}$ being the development outcome in 2000, $Rail_{c,1918}$ the set of railway dummies (using the lines in 1918) and $X_{c,1918}$ the same controls as before. Regression results for model (4) are reported in Table 5. In Columns (1)-(2), we show that cells along the railway lines are more urbanized, whether we drop railway nodes or not. In Column (3), we use satellite data on night lights as an alternative measure of economic development, as in Storeygard (2011) and Henderson, Storeygard & Weil (2012).[28] In Column (3), we show that the share of inhabitants living in residences with solid walls increases with railway connectivity. In Columns (5)-(7), we show that railway cells have less agricultural employment and more manufacturing and service employment. Columns (8)-(17) confirm railway cells have better access to communication, education and health infrastructure, and have higher levels of human capital as a a result. Lastly, locations along the railway lines have better roads today, as they are more likely to be crossed by a paved road. Interestingly, these effects decrease as we move further away from the line, they seem stronger for more exclusive public goods (i.e., those with a lower mean) and they only slightly decrease when we also control for the population and infrastructure effects in 1931.

Even if the railway cells are more developed today, there might have been some convergence between the railway and the non-railway cells between 1931 and 2000 (or 1970 and 2000). So the main issue is

[28] We would like to thank Adam Storeygard for giving us access to their data.

whether the gap due to the level effect in 1931 has been widening or narrowing over time. We investigate this issue by using panel data on cell population and urban growth every ten years over one century. As in Bleakley & Lin 2012, we study whether how the railway effects varies over time. We run the following panel regression for 554 suitable cells c and years t:

$$LogPop_{c,t} = Rail_c \beta_t + \gamma'_t + \delta'_c + X_c \zeta'_t + \nu'_{c,t} \tag{5}$$

with $LogPop_{c,t}$ being the log of population and $Rail_{c,t}$ the railway dummies. We use a log-linear model because population in Ghana has boomed in the last century, and taking logs allows to remove any trend. For each year t, β_t are the estimates of the effects of rail connectivity relative to 1901. We include gridcell fixed effect, year fixed effects and the same controls as before interacted with year dummies. As we have no data for total population in some years, we only use t = [1901, 1931, 1970, 2000] for total population, rural and urban growth. Figure 7 displays the effects β_t (relative to 1901). While total and rural populations have boomed in 1901-1931, the level effect is much lower in 2000, which implies there has been relative convergence between 1931 and 2000. Locations along the railway lines are still more developed today as the railway effects are still significantly positive in 2000, but they are not as high as what they used to be. Urbanization has followed a different pattern, as the strong level effect in 1901-1931 was halved in 1931-1970 before increasing even further in 1970-2000. Since we have data on urban population every ten years more or less, we can thoroughly study the evolution of these effects. The level effect was high in 1931, decreased in 1948, increased in 1960, was halved in 1970 and 1984, but increased even further in 2000. Yet the level effect always remains positive and significant.

A look at the history of Ghana helps us understand these patterns. In 1938-1948, the cocoa-producing areas have been severely hit by the cocoa swollen-shoot virus disease (Hill 1963). Aggregated production was halved between 1938 and 1944. Because no attempt could be made to control the disease until after the Second World War, millions of trees were killed and more millions had to be removed to try to control it. (Hill, 1963) describes how the towns of the cocoa-producing areas were mechanically affected by the shocks. In 1948-1960, economic growth was high and cocoa production boomed in the Ashanti province, along the Western line. These two factors probably explain why the railway cells recovered their initial advantage. In 1960-1970 and 1970-1984, the political and economic crisis, the decline of railways and the end of the cocoa boom could explain why the towns of the old cocoa-producing regions were "losing" relatively to the rest of the country. Between 1966 and 1984, per capita GDP decreased by almost 30% and the urban formal sector was hit hardest, with both reductions in productivity and employment (Jedwab & Osei 2012). Growth resumed after 1984, and productivity and employment simultaneously rose in the urban formal sector. This probably explains the patterns found in our analysis. The railway cells have an initial advantage but this initial advantage only persists if the economic situation makes it possible.

6.3 Discussion on Path Dependence

To investigate the factors of path dependence, we examine urban growth between 1984 and 2000. During this period, the level effect strongly increased, although the railway cells had lost their initial advantages in transportation infrastructure and cocoa production. This provides with a natural experiment for opening the black box of path dependence. We first regress the log-difference in urban population on initial urban population, the railway dummies and the same controls as before:

$$LogUPop_{c,2000} - LogUPop_{c,1984} = Rail_c \beta + \theta LogUPop_{c,1984} + X_c + \mu_c \tag{6}$$

We find strong effects of railway connectivity of urban growth in 1984-2000: +0.92*** and +0.96** for 0-10 and 10-20 km respectively, and no significant effects beyond. We add additional controls to see how they capture the railway effects. First, including rural population in 1984 halves the coefficients. Rail connectivity had a strong effect on rural growth in 1931 and beyond, providing cities with a large pool of rural workers and creating villages that eventually “urbanize" (i.e., pass the 1,000 inhabitants threshold). Second, adding the employment share of manufacturing further decreases the two effects.

Adding measures of service employment, human capital and infrastructure does not modify at all the remaining railway effects but the manufacturing effect becomes twice smaller. We use the same variables as in Table 5. This indicates that there has been capital accumulation along the railway lines, with capital actually including various types of capital that are all correlated to each other: physical capital (a la Solow), human capital (a la Mankiw, Romer and Weil) and public capital (a la Barro and Romer). It is impossible to distinguish the effects of the various types of capital, but the analysis above points to demographic growth and capital accumulation as the main sources of path dependence.

Yet should we conclude that railways have initiated modern economic growth in Ghana? Per capita income in Ghana has remained very low, around $1,350 (PPP, 2010 USD). Although railway cells are more industrialized today (see Col. (6) of Table 5, e.g. +3.5*** for 0-10 km cells), the national employment share of manufacturing was only 6.4% in 2000. Manufacturing is relatively unproductive in Ghana, as it consists mostly of labor-intensive subsectors, such as African processed foods, African furniture and African textiles and clothing (Jedwab 2011, Jedwab & Osei 2012). That is why we characterize this effect as proto-industrialization. In other words, the railway lines have made some locations more attractive within Ghana but they did not industrialize the country as a whole.

7. CONCLUSION

Using railroad construction in colonial Ghana as a natural experiment, we have analyzed the impact of colonial investments on long-term development in Africa. Using panel data at a fine spatial level over one century (11x11 km grid cells in 1891-2000), we find a strong effect of rail connectivity on the production of cocoa, the country's main export commodity, and development, which we proxy by population and urban growth. First, the exogeneity assumption is supported by the data and our results are robust to using various identification strategies. Second, transportation infrastructure investments had large welfare effects for Ghanaians during the colonial period. We discuss how colonization meant both extraction and development in this context. Third, railway construction had a persistent impact: railway cells are more developed today despite a complete displacement of rail by other means of transport in the late 1970s. We investigate the factors of path persistence and explain that demographic growth and capital accumulation account for this result. We believe that the channels and patterns that we discussed for Ghana appear to be paralleled in other African countries, e.g. coffee and tea in Kenya or groundnuts in Senegal. Colonial investments in transportation infrastructure have shaped the economic geography of African countries, allowing them to specialize in the export of natural resources.

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.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review*, 95(3): 546–579.

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.

African Union. 2006. *Rail Development in Africa: Stakes and Prospects, Objectives and Mis- sions of The African Rail Union (ARU).* Document prepared by the ARU General Secretariat for the Conference of African Ministers In Charge of Rail Transport 10-14 April 2006, Brazzaville (Congo).

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 A. Margo. 2010. "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60." *Social Science History*, 34(2): 171–197.

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

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.

Berkowitz, Daniel, and Karen Clay. 2003. "Initial Conditions, Institutional Dynamics and Economic Performance: Evidence from the American States." William Davidson Institute at the University of Michigan William Davidson Institute Working Papers Series 2003-615.

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.

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

Chaves, Isaías, Stanley L. Engerman, and James A. Robinson. 2010. "Reinventing the Wheel: The Economic Benefits of Wheeled Transportation in Early Colonial British West Africa." Unpublished manuscript, Department of Political Science, Stanford University.

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.

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehan. 2008. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." International Food Policy Research Institute. IFPRI discussion papers 840.

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

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 and Harvard University.

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.

Fay, Marianne, and Tito Yepes. 2003. "Investing in Infrastructure : What is Needed From 2000 to 2010?" The World Bank. Policy Research Working Paper Series 3102.

Feyrer, James. 2009. "Trade and Income – Exploiting Time Series in Geography." National Bureau of Economic Research, Inc. NBER Working Papers 14910.

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.

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

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 Capitalism.* 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 Economics*, 1(2): 176–215.

Huillery, Elise. 2011. "The Impact of European Settlement within French West Africa: Did Pre-colonial Prosperous Areas Fall Behind?" *Journal of African Economies*, 20(2): 263–311.

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.

Johnson, Simon, Jonathan David Ostry, and Arvind Subramanian. 2007. "The Prospects for Sustained Growth in Africa: Benchmarking the Constraints." International Monetary Fund. IMF Working Papers 07/52.

Leunig, Tim. 2010. "Social Savings." *Journal of Economic Surveys*, 24(5): 775–800.

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.

Naritomi, Joana, Rodrigo R. Soares, and Juliano J. Assunção. 2012. "Institutional Development and Colonial Heritage within Brazil." *The Journal of Economic History*, 72(02): 393–422.

Nunn, Nathan. 2007. "Historical Legacies: A Model Linking Africa's Past to its Current Underdevelopment." *Journal of Development Economics*, 83(1): 157–175.

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.

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

Rodrik, Dani. 1998. "Trade Policy and Economic Performance in Sub-Saharan Africa." National Bureau of Economic Research, Inc. NBER Working Papers 6562.

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.

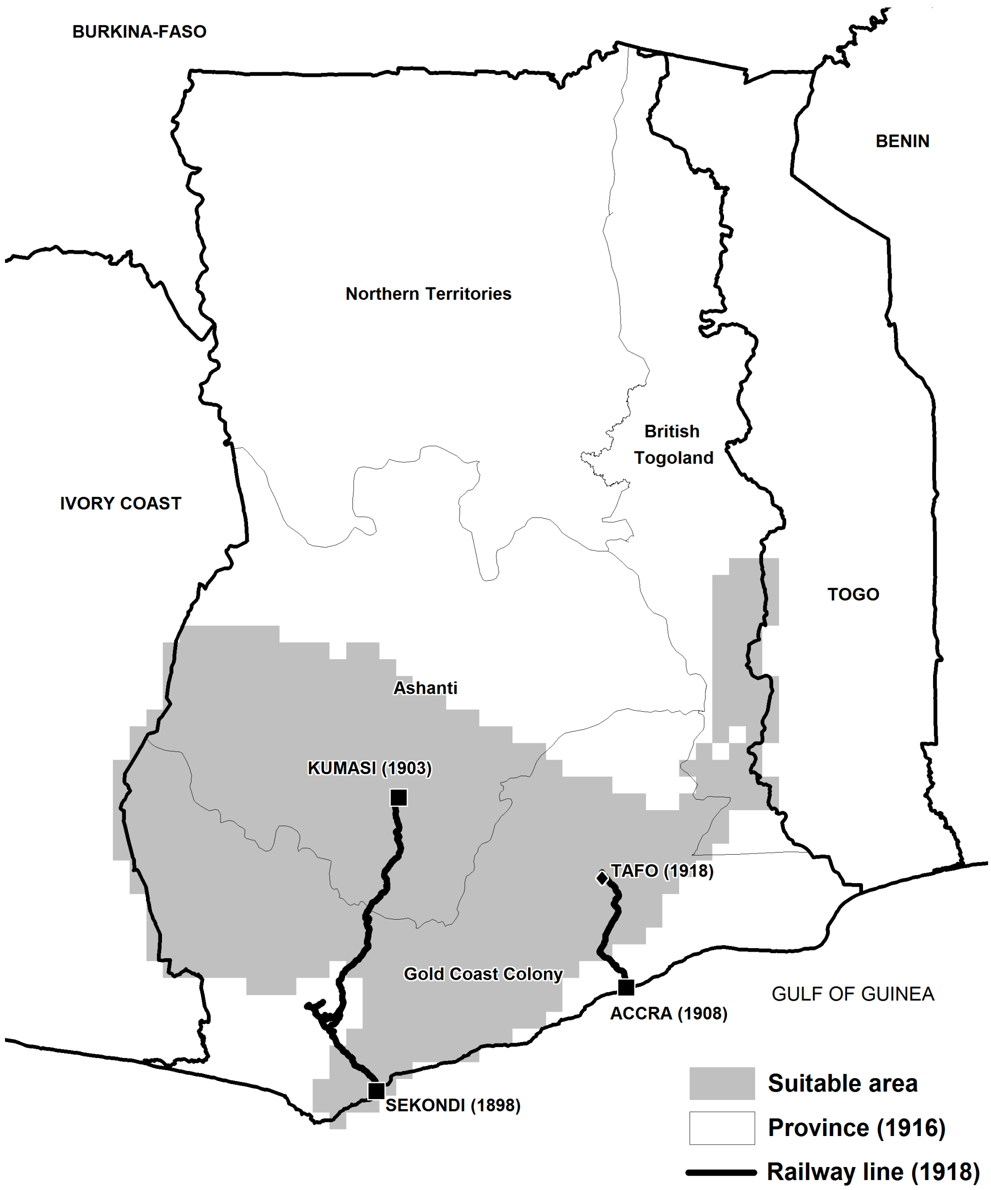
Tosh, John. 1980. "The Cash Crop Revolution in Tropical Africa: An Agricultural Reappraisal." *African Affairs*, 79(314): 79–94.

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.

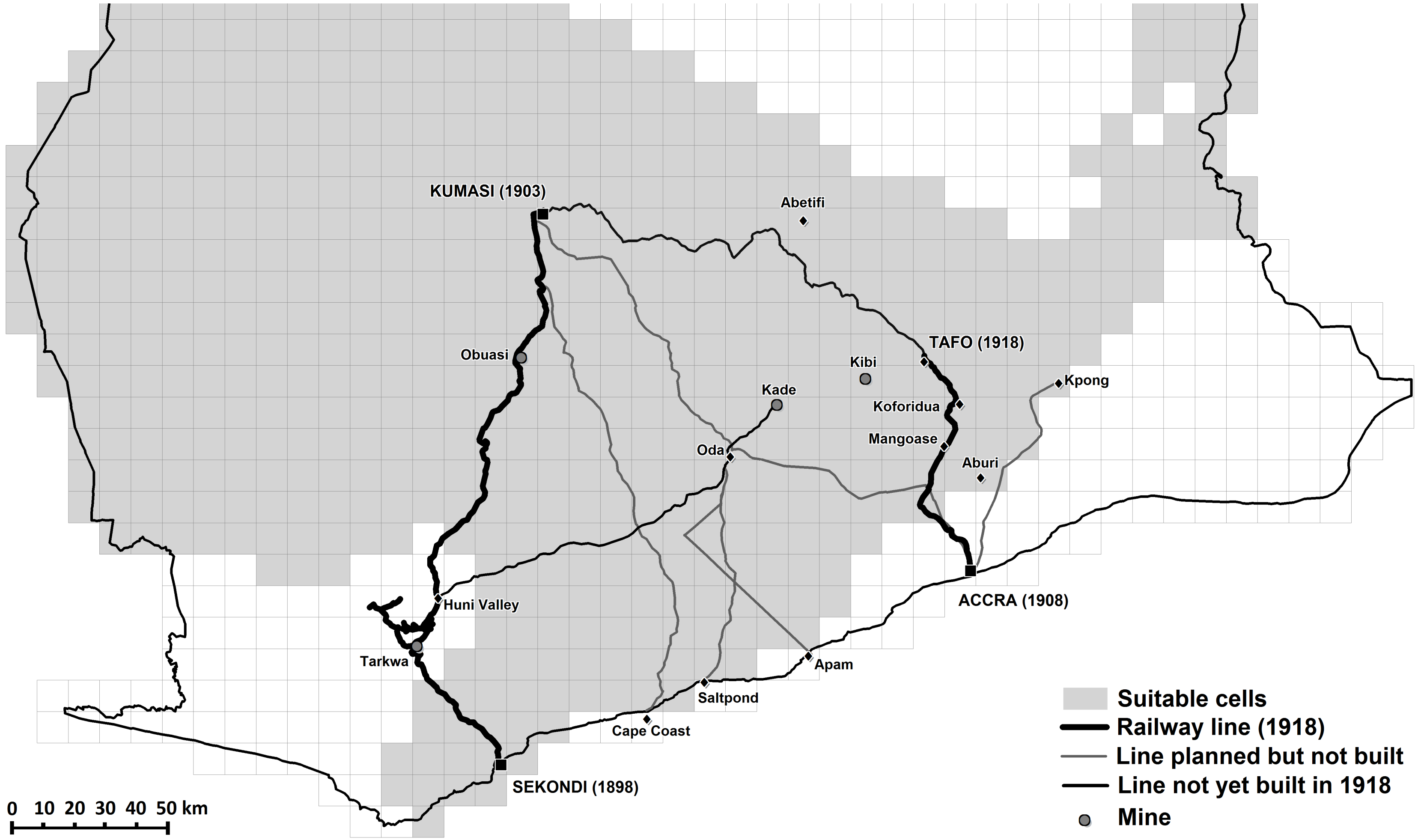
FIGURES

Figure 1: Cocoa Suitability, Provinces and Railway Lines in 1918.



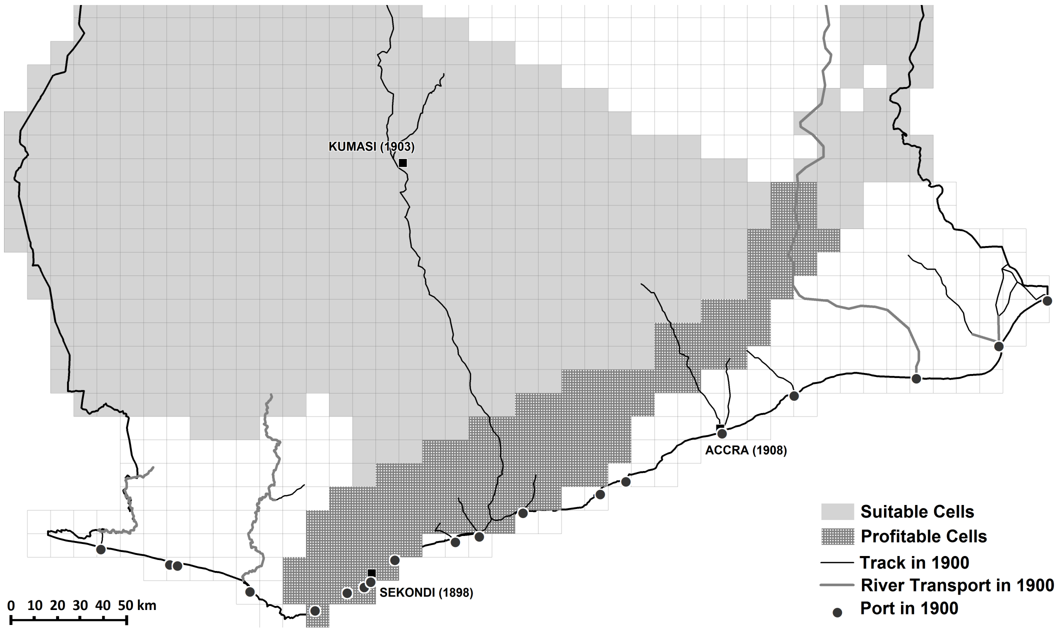
Note: The suitable area corresponds to the tropical forest around 1900. Province boundaries date from 1916. See Data Appendix A for sources.

Figure 2: Railway Lines in 1918 and Placebo Lines.



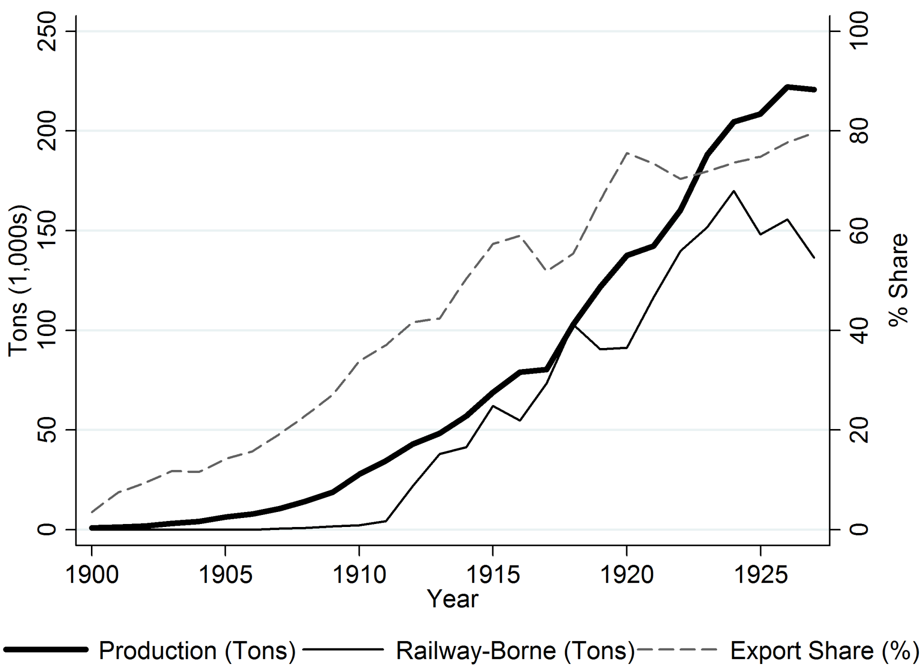
Note: The map only shows Southern Ghana. A 10x10 km cell is defined as suitable if it contains cocoa soils. The map displays the railway lines in 1918 and the placebo lines: lines that were planned but not built (Cape Coast-Kumasi 1873, Saltpond-Kumasi 1893, Apam-Kumasi 1897, Accra-Kumasi 1897, Accra-Kpong 1898) and lines that were not built yet to affect cocoa production in 1927 (Tafo-Kumasi 1923, Huni Valley-Kade 1926). It also shows gold mines and diamond fields.

Figure 3: Transportation Networks in 1900 and Area of Profitable Production.



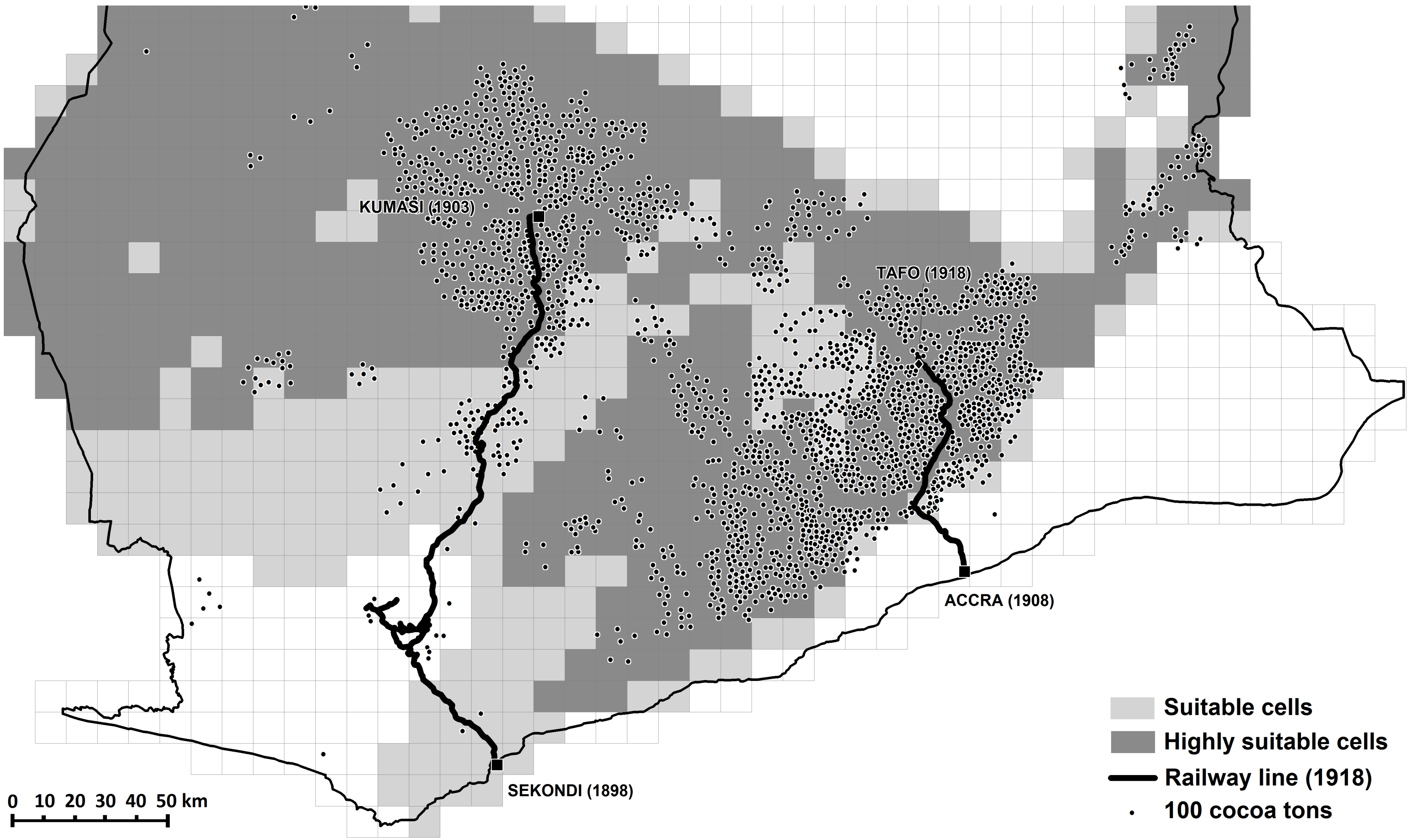
Note: The map shows transportation networks in 1900, the cells suitable for cocoa cultivation and the cells for which cocoa cultivation is profitable given transport costs in 1900 and the export price offered at the port in 1908-1931. See Data Appendix A for sources.

Figure 4: Cocoa Production, Exports and Transportation, 1900-1927.



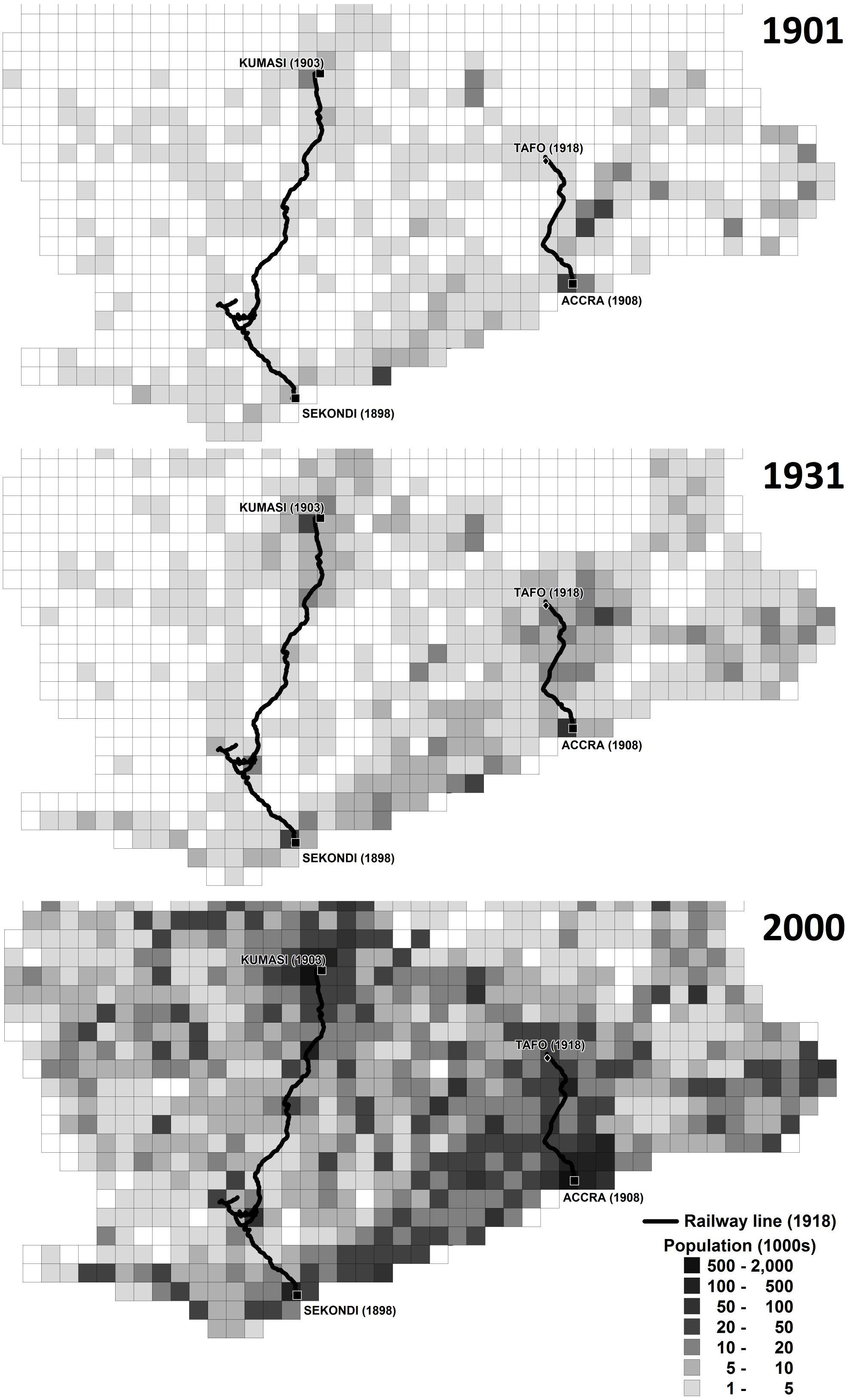
Note: The figure displays three-year moving averages for cocoa production, cocoa tonnages transported by rail to a coastal port, and the share of cocoa exports in total exports from 1901 to 1927. See Data Appendix A for sources.

Figure 5: Railway Lines in 1918, Cocoa Suitability and Production in 1927.



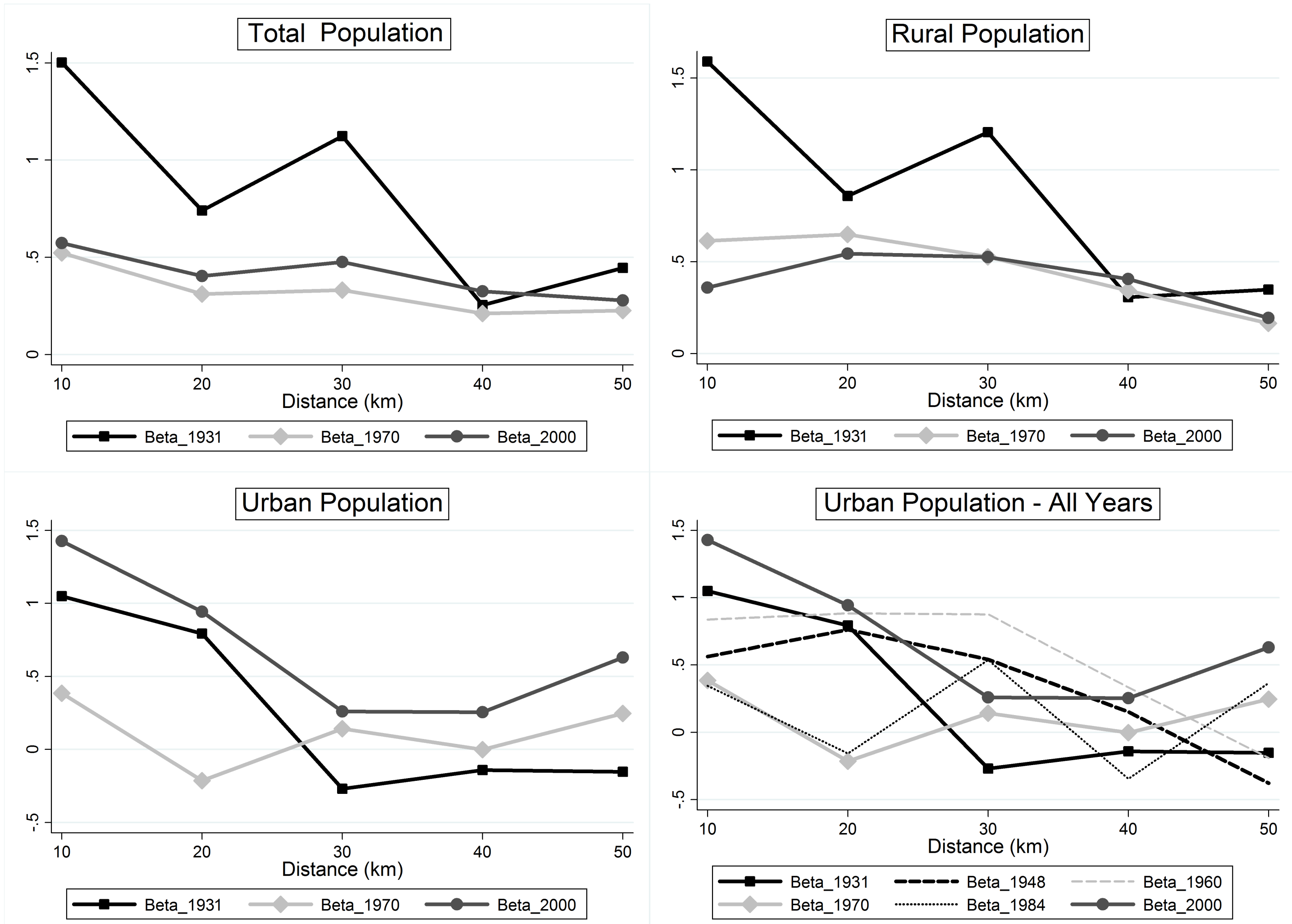
Note: The map only shows Southern Ghana. A cell is defined as suitable if it contains cocoa soils and highly suitable if more than 50% of its area consists of forest ochrosols, the best cocoa soils. The map displays the railway lines in 1918, suitable cells, highly suitable cells and cocoa production in 1927.

Figure 6: Railway Lines in 1918 and Cell Population in 1901, 1931 and 2000.



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

Figure 7: Long-Term Effects (Relative to 1901) of Railway Connectivity



Note: This graph displays estimates of Equations (5) and (6) for each distance threshold and each year = [1931, 1970, 2000], using 1901 as the reference year. We show the railway 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		
Dummy “the cell contains a mine”	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 for cocoa (%)	-0.21***	-0.26***	-0.32***
Share soils very highly suitable for cocoa (%)	-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 the 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: * 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 railway line. 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-Prasu-Kumasi placebo line.

TABLE 2: RAILROADS, COCOA PRODUCTION AND POPULATION GROWTH

Dependent Variable:	**Cocoa 1927** (Tons)	**Population 1931** (Number of Inhabitants)				**Rural Population 1931** (in Localities < 1000)				**Urban Population 1931** (in Localities > 1000)			
	(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]
Gridcell 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: Standard errors clustered at the gridcell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. Additional controls interacted with year dummies: dummy "the cell contains a mine in 1931", cocoa production in 1901, urban population in 1901, rural population 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), Euclidean distances (km) to Accra, Kumasi, Aburi, a port in 1901 and the coast.

TABLE 3: RAILROADS, COCOA PRODUCTION AND POPULATION GROWTH, ROBUSTNESS

Regression:	*Main*	*Highly Suitable Cells*	*West vs. East*	Highly Suit.	*District-Year FE*	*Matching* Placebo Cells	C.Coast Kumasi	*IV Straight Lines*	*Distance To Rail Station*	*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]									
Gridcell 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: Standard errors clustered at the gridcell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. Additional controls interacted with year dummies: dummy “the cell contains a mine in 1931”, cocoa production in 1901, urban population in 1901, rural population 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), Euclidean distances (km) to Accra, Kumasi, Aburi, a port in 1901 and the coast. Only cocoa production, urban population and rural population in 1901 are included in Column (10). Rural population in 1901 is omitted in Column (12) as we have no such data for the rest of the country.

TABLE 4: RAILROADS, COCOA PRODUCTION AND POPULATION GROWTH, 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]
Controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the gridcell level are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. Additional controls interacted with year dummies: dummy "the cell contains a mine in 1931", cocoa production in 1901, urban population in 1901, rural population 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), Euclidean distances (km) to Accra, Kumasi, Aburi, a port in 1901 and the coast. In Panel C and F, we drop those cells that are less than 20 km from the railway lines in 1918.

TABLE 5: RAILROADS AND ECONOMIC DEVELOPMENT IN 2000

Dependent Variable:	Urban Population (Inh.)		Lights	Solid Walls	Employment Share (%)			Post Office	Telephone
		No nodes	(%)	(%)	Agri.	Manuf.	Serv.	(%)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dummy Rail 1918, 0-10 km	33,791*	7,825**	29.8***	14.1***	-16.0***	3.5***	9.5***	8.7*	15.9***
	[17,862]	[3,238]	[4.7]	[3.0]	[2.8]	[0.8]	[1.9]	[4.7]	[4.9]
Dummy Rail 1918, 10-20 km	3,785	2,962**	16.0***	4.6*	-7.7***	2.1***	4.2***	6.2	9.5**
	[2,767]	[1,344]	[4.3]	[2.4]	[2.3]	[0.7]	[1.6]	[4.4]	[4.2]
Dummy Rail 1918, 20-30 km	1,441	1,234	7.4**	1.0	-5.4***	1.7***	3.1**	-2.4	3.9
	[2,719]	[1,488]	[3.6]	[2.1]	[2.0]	[0.6]	[1.4]	[3.9]	[3.9]
Dummy Rail 1918, 30-40 km	1,046	830	3.7	2.7	-2.7	0.9	0.9	1.1	0.5
	[2,222]	[1,279]	[3.2]	[2.2]	[2.0]	[0.6]	[1.3]	[4.1]	[3.6]
Dummy Rail 1918, 40-50 km	4,633	1,513	-0.0	4.8*	-4.3**	1.6***	1.8	4.6	2.7
	[3,143]	[1,403]	[3.0]	[2.7]	[2.0]	[0.6]	[1.3]	[4.3]	[4.1]
Mean	10,551	8,067	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: Standard errors are reported in parentheses; * p<0.10, ** p<0.05, *** p<0.01. We use various sources to reconstruct 18 outcomes for 554 gridcells around 2000. All regressions include the same controls as before. 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): share of adults ($\succeq$18y.o.) being literate, having 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. See Data Appendix for Data Sources.

APPENDICES

A. DATA DESCRIPTION

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

Spatial Units:
We assemble data for 2091 grid cells of 0.1x0.1 degrees (11x11 km) in Ghana from 1891 to 2000. We choose a high resolution grid because we have very precise GIS data on railways, cocoa production, population and urban growth in 1891-2000, and economic development in 2000. Each grid cell has the same size, except those cells that are coastal or crossed by a border. We create two dummies equal to one if the grid cell is coastal or bordering another country to control for this issue. Grid cells belong to 110 districts and 10 provinces using 2000 boundaries, or 4 regions using 1931 boundaries.

Railway Data:
We obtain the layout of railway lines in GIS from *Digital Chart of the World*. We then use Gould (1960), Dickson (1968) and Luntinen (1996) to recreate the history of railway construction. For each line, we know when it was surveyed, planned, started and finished, and when each station was reached and opened. From the same sources, we know lines that were built but not planned. Most of those placebo lines follow historical trade routes and have become roads later. Using the GIS road network also available from *Digital Chart of the World*, we recreate in GIS those placebo lines. We calculate for each grid cell the Euclidean distance (km) from the cell centroid to each real or placebo line. Lastly, we create a set of dummies equal to one if the grid cell is less than X km away from the railway line: 0-10, 10-20, 20-30, 30-40 and 40-50 km. We create a dummy equal to one if the grid cell contains a rail station in 1918. We also know how many tons of cocoa were brought to each station in 1918.[1] Data on railway traffic for both lines was obtained from various sources.[2]

Cash Crop Production and Price Data:
A very precise map of cash crop production in 1927 was obtained from the *1927 Yearbook of the Gold Coast* and digitized. This map displays dots for each 100 tons of cocoa production.[3] We then use GIS to reconstruct total cocoa production (tons) for each grid cell using 1927 as an approximation for 1931. We proceed similarly to create cocoa production in 1950 using a production map published in the *Report on the Cocoa Industry in Sierra Leone and Notes on the Cocoa Industry of the Gold Coast* by Cadbury Brothers LTD in 1955. We then use Bateman (1965) to obtain the international and national producer price.

Population and Urban Data, 1891, 1901, 1931, 1948, 1960, 1970, 1984 and 2000:
We collect population data from the gazetteers of the *Population and Housing Censuses* 1891, 1901, 1931, 1948, 1960, 1970, 1984 and 2000. They list localities and their population size. Defining as a city any locality with more than 1,000 inhabitants, we obtain a geospatialized sample of 1,373 different cities in Ghana for all these years. Using GIS, we recalculate total urban population for each grid cell. We are then able to recreate rural population for each gridcell in 1901, 1931, 1970, 1984 and 2000. From the census gazetteer, we know the population size of each village (locality with less than 1,000 inhabitants). But it was impossible to find the geographical coordinates of all of them. Yet, the 1901 census was exhaustively conducted and geospatialized in the South of Ghana (756 cells). We know for each cell the number of *large towns*, *towns* (more than 500 inhabitants), *head chief towns*, *large villages* (100-500 inhabitants) and *villages* (less than 100 inhabitants). Using GIS, we can deduce for each cell the number of villages that are less than 100 inhabitants, the number of villages that have between 100 and 500 inhabitants and the number of villages that have between 500 and 1,000 inhabitants. From the census, we know the average settlement size for each category and we can reconstruct total rural population for each cell in 1901. For 1931, we have a map of the distribution of population for the whole country.[4] This map displays at a very fine spatial level settlements that have less than 500 inhabitants and settlements that have between 500 and 1,000 inhabitants. From the census, we know the average settlement size for each category and we can reconstruct total rural population for each cell in 1931. We use the 2000 Facility Census which has population data on all settlements in 1970, 1984 and 2000. We first retrieved the geographic coordinates of localities by matching place names with several geographic

[1] This information was retrieved for the main railway stations from the *Administration Report of Gold Coast Railways for the Year 1921*.

[2] These sources are: *Government Gazette Supplement of Accra, Gold Coast, West Africa* 1914 and *Administration Report of Gold Coast Railways for the Year* 1920, 1921, 1929-30 and 1931-32.

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

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

data bases. Second, when no match could be found, we use paper maps of Enumeration Areas and localities to get an approximation of the geographical coordinates. To conclude, we have cell urban population in 1891, 1901, 1931, 1948, 1960, 1970, 1984 and 2000. We have cell rural and total populations for each Southern cell in 1901, and each cell in 1931 and 1970-2000.

Height Data:
Data on Ghanaian soldiers enlisted in the British Army was collected by Alexander Moradi using official soldier files. The data set is described in Moradi (2009).

Transportation Networks in 1901 and Road Data:
Transportation networks in 1901 are obtained from Dickson (1968). The layout of historical trade routes is obtained from Dickson (1968) (p.215), and digitized in GIS. We use various sources to reconstruct a GIS database of roads in 1922 and 1930: Gould 1960 and *Map of The Gold Coast with Togoland Under British Mandate*, published in 1930 by the Survey Headquarters. Those road map have a consistent legend showing class 1 roads ("*roads suitable for motor traffic throughout the year*"), class 2 roads ("*roads suitable for motor traffic but occasionally closed*") and class 3 roads ("*roads suitable for motor traffic in dry season only*") and other roads. Other roads are not suitable to motor traffic and are not consider here. We use a Michelin paper map to recreate the 2000 road network in GIS, distinguishing paved (bitumenized), improved (laterite) and earthen roads.

Mining Production and Price Data:
We use annual production data for Ghanaian mines in 1901-1931 for four commodities: gold, manganese and diamond.[5] As we have the geographical coordinates of each mine, we create a dummy equal to one if the cell contains a mine.

Geographical Data:
Forest data comes from land cover GIS data compiled by Globcover (2009). The data displays those areas with virgin forest or mixed virgin forest/croplands, which were areas with virgin forest before it was cleared for cash crop production. Soil data comes from the *1958 Survey of Ghana Classification Map of Cocoa Soils for Southern Ghana*. This map was digitized in GIS and we calculated for each cell the share of land which is suitable to cocoa cultivation. We also know the respective shares of land which consists of ochrosols (first class, second class, third class, unsuitable), oxysols and intergrades. A cell is defined as suitable if it contains cocoa soils. It is then highly suitable if more than 50% of its area consists of forest ochrosols, the best soils for cocoa cultivation. It is very highly suitable if more than 50% of its area consists of class 1 and class 2 ochrosols. Climate data comes from *Terrestrial Air Temperature and Precipitation: 1900-2007 Gridded Monthly Time Series, Version 1.01*, 2007, University of Delaware. We estimate for each grid cell average annual precipitations (mms) in 1900-1960. Topography comes from SRTM3 data. We estimate for each grid cell the mean and standard deviation of altitude (meters). The standard deviation captures the slope and ruggedness of the terrain. Lastly, we use GIS to obtain the Euclidean distance (km) from each cell centroid to the coast.

Economic Geography Data:
For each grid cell, we use GIS to get the Euclidean distances (km) to Accra, the capital city, Kumasi, the largest hinterland city, Aburi, the city where Ghanaian cocoa production originated, a port in 1901, and the coast.

Employment, Human Capital and Infrastructure Data in 2000:
We use the 2000 *Facility Census* and the 2000 *Population and Housing Census* to recreate data on economic development at the gridcell level in 2000: sectoral employment shares, share of inhabitants 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, share of inhabitants with access to clean water, share of adults ($\succeq$18y.o.) being literate, having attended a SSS.

B. PLACEBO LINES

This appendix presents background information on the placebo lines. It draws heavily from Gould (1960), Tsey (1986) and Luntinen (1996). Figure 2 shows their location.

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

B.1 Lines Proposed But Never Built

Several private initiatives submitted proposals to convince the Colonial Office of the profitability of their railway schemes. All asked for a government guarantee of interest on capital outlay. The government opposed this, with reference to the obvious incentive problems of such a guarantee, subsidizing British investors and leading, as in India, to over-capitalization. These initiatives did not enter a phase of concrete planning, nor were they able to raise the capital necessary for their schemes. Yet pressure mounted to build railways in Ghana, as the French and the Germans were already building their own networks in West Africa. Tsey (1986, p.17) writes: "Indeed this had become all the more urgent in view of the fact that the Brussels Conference of 1889-1890 had called on all the European powers to back up their claims to colonial territories in Africa by the establishment of effective administrations particularly through the construction of railways and roads." Eventually, the state favored public ownership. The government proposals described below can be considered as counterfactual alternatives to the two routes built. We argue that these lines all had the same probability of being built, and only random events (e.g., a change in the colonial Governor) explain why construction did not happen.

Cape Coast-Kumasi 1873

The first proposal to build a railway was made in 1873 to connect Cape Coast to Kumasi via Prasu. The British planned to use the line to send troops to fight the Ashanti. Some railway materials were landed at Cape Coast, but the project was dropped, since it was not possible to build the line in time for the military operation. The line was also proposed in 1891 and 1892 by private consortia (Luntinen 1996, p.18). Cape Coast was the capital of the Gold Coast colony before it was moved to Accra in 1877. The latter had a drier climate and was believed to be a healthier place for Europeans. In 1901 Cape Coast was similar in size to Accra (respectively 28,948 and 30,144). Moreover, Cape Coast was the starting point of an important historical trade route to Kumasi. Villages clustered along this road like pearls on a string (see Figures ...). This permitted Cape Coast to achieve the largest trade volume of the coastal towns in 1900 (Gould 1960, p.17). Hence, in terms of existing traffic, the line also had some potential.

Saltpond-Oda-Kumasi 1893

In 1893, the colonial government commissioned a survey for a railroad network that would benefit the whole country, not only the mining industry. Government and engineers favored a trunk route. Saltpond was chosen as starting point for its central location and because construction materials could be landed easily. The line was to reach Kumasi, crossing densely populated and rich palm kernel and palm oil areas. The line had the support of Governor Griffith. Tsey (1986, p.19) writes: "A major step towards a full railway proposal came when Governor Griffith suggested a 300-400 mile network, the first phase of which was to form a trunk route from the most central port of the colony into the interior. Such a central line through the agricultural districts, the Governor argued, not only offered some hope of financial return, but in addition, it would form a suitable basis for future railway development." Yet Governor Griffith retired in April 1895, and the new Governor Maxwell changed course and again favored the mining industry. Tsey (1986, p.32) explains that Governor Maxwell, who previously worked in the Malay States where railways served the tin mines, thought that the same model should be applied to Ghana with gold.

Apam-Oda-Kumasi 1897, Accra-Oda-Kumasi 1897

Although the construction of the Western line to mining areas was about to start, it was still widely accepted that the colony needed a central line to be built to tap the palm oil areas. Yet the Saltpond-Oda-Kumasi project was eventually dropped because of the relatively higher capital outlay compared to Apam as starting point. The Apam-Oda-Kumasi line was slightly shorter, thus cheaper, and had all the advantages of a central railway route. Calculations of the consulting engineer indicated its profitability. However, discontent grew amongst Accra merchants who thought that the capital and chief trading town should be connected first. The Accra-Oda-Kumasi was also surveyed, and "the railway surveys led to the conclusion that Apam was the better terminus for a Kumasi railway" (Luntinen 1996, p.33). A conference was to be held in London before the end of 1897 to discuss the various proposals but unexpectedly Governor Maxwell died before reaching London. He was replaced by Governor Hodgson who "was of the opinion that the principal railways of the colony should converge upon Accra. This town had been selected as the seat of the Government, and it was the most healthy site on the coast for Europeans. Much money had been invested there and he thought that it would be a mistake to raise a rival port." (Luntinen 1996, p.34). Besides, he thought that a central line from Accra would be useless as it would not directly traverse the palm oil areas below Apam, so he favored another line to Kpong, on the Volta River. The possible extension of the Sekondi-Tarkwa line - though originally intended as a short, local mining railway - to Kumasi also undermined the central route strategy.

Accra-Kpong 1898

The main objective of this line was to tap palm oil areas to the north-east of Accra and boost cotton cultivation in the districts bordering Togo. It would also link Accra to the Volta river, to facilitate the transport of government stores to the North, and to the "Government Sanatorium and Botanical Station at Aburi, 1400 feet above sea level, [...] so that Europeans will be able to reside in this delightful spot, coming daily to their offices in Accra" (Tsey 1986, p.55). The line was approved for construction, but it was decided that works would not begin before the extension to Kumasi was completed in late 1903. However, Governor Nathan decided to retire in February 1904. He was replaced by Governor Rodger, who thought that a railway line from Accra to Kumasi was more important and that the Volta river was already being used from the coastal port of Apam. The mercantile representative on the Legislative Council also strongly opposed the line, given "he was operating a boat service on the Volta" between Kpong and Apam and did not want the competition of the rail (Tsey 1986, p.57).

B.2 Lines Not Yet Built

The rail network was later expanded. We also consider lines that were actually built, but not in time to affect production in 1927. Cocoa is a perennial crop. Pod production starts after 5 years and peaks after 25 years. Hence, for observing an impact on cocoa production in 1927, farmers must have grown cocoa trees before 1923. There is no qualitative evidence that this happened to a significant degree. If the prospect of railway connectivity did indeed induce much production in advance, it would add to the positive correlation expected from reverse causality (cocoa production attracted the railroad). However, we do not find any positive effect of the two placebo lines below.

<u>Tafo-Kumasi 1923</u>
The Eastern line, with Accra as terminus, reached Tafo in 1916, when war time restrictions on construction materials suspended all further railway projects. Tafo station was opened in July 1917. Bauxite discoveries, midway between Tafo and Kumasi in 1917, led to the decision to extend the railway to Kumasi (Tsey (1986), p.64). Actual construction, however, only started in 1920. The line was completed in 1923.

<u>Huni Valley-Kade 1927</u>
The line ran parallel to the coast, about 80 km inland. It connected the diamond mines at Kade and was supposed to encourage cocoa, kola, palm oil and timber exports. By conveying more traffic to the newly developed harbor at Takoradi (Sekondi), it was hoped to make the port viable. Construction begun in 1923. Several roads already connected the area to the coast, but they were of poor quality. Railroad surveyors believed that lorry traffic could not operate profitably beyond 50 km from the coast, but this turned out to be wrong. The short distance to the coast made lorry transport very competitive reviving the old ports of Cape Coast, Saltpond and Winneba directly in the South.

References

BATEMAN, M. (1965): *Cocoa in the Ghanaian Economy*. Unpublished Ph.D. dissertation, MIT.

DICKSON, K. (1968): *A Historical Geography of Ghana*. Cambridge: Cambridge University Press.

GLOBCOVER (2009): *Globcover Regional: Africa*. Louvain: European Space Agency Ionia GlobCover.

GOULD, P. R. (1960): *The Development of the Transportation Pattern in Ghana*. Chicago: Northwestern University Press.

GUNNARSSON, C. (1978): *The Gold Coast Cocoa Industry 1900-1939: Production, Prices and Structural Change*. Lund: Published Thesis.

LUNTINEN, P. (1996): *Railway on the Gold Coast: A meeting of two cultures : a colonial history*. Helsinki: Suomalainen Tiedeakatemia.

MORADI, A. (2009): *Men under Arms in Colonial Africa: The Gold Coast Regiment*. Unpublished manuscript, Department of Economics, University of Sussex.

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