# On the Dispensability of New Transportation Technologies: Evidence from Colonial Railroads in Nigeria

Okoye, Dozie and Pongou, Roland and Yokossi, Tite

Dalhousie University, University of Ottawa, Massachusetts Institute of Technology

25 November 2016

# On the Dispensability of New Transportation Technologies: Evidence from Colonial Railroads in Nigeria*

Dozie Okoye
Dalhousie University

Roland Pongou
University of Ottawa

Tite Yokossi
MIT

November 25, 2016

**Abstract**

We examine Fogel's influential hypothesis that new transportation technologies may be dispensable if pre-existing technologies are viable or can simply be improved. Exploiting the construction of colonial railroads in Nigeria, we find that the railway has large long-lasting impacts on individual and local development in the North, but virtually no impact in the South neither in the short run nor in the long run. This heterogeneous impact of the railway can be accounted for by the level of pre-railway access to ports of export. Consistent with Fogel's argument, the railway did not transform areas that had viable transportation alternatives for exporting purposes. Using information on changes in shipping costs and quantities, we highlight the importance of opportunity costs to the adoption and impact of new transportation investments.

**Keywords:** Fogel's Hypothesis, Colonial Investments, Railway, Africa, Development, Nigeria
**JEL Classification:** I20, N30, N37, N47, O15, Z12.

*We would like to acknowledge suggestions arising from discussions with David Atkin, Abhijit Banerjee, Esther Duflo, Jason Garred, Talan Iscan, Benjamin Olken, Lars Osberg, and Frank Schilbach, that greatly improved the paper. We are also grateful for valuable comments from participants at the Development Economics Lunch at MIT, SIER conference at the African School of Economics, the 2016 Canadian Development Study Group Meetings, the Macro and Development Group Lunch at Dalhousie University, University of Western Ontario's 50th Anniversary conference, and St. Francis Xavier University. Special thanks to Remi Jedwab and Alexander Moradi for making their dataset on city growth in Africa publicly available and accessible. Address correspondence to: Dozie Okoye at cokoye@dal.ca; Roland pongou at rpongou@uottawa.ca; or Tite Yokossi at titey@mit.edu.

# 1 Introduction

A number of recent empirical studies document the importance of transportation technologies for economic development in a variety of settings.[1] From a policy perspective, however, the conditions under which "new" transportation technologies lead to sustained economic growth and development are still not well understood. Identifying such conditions is important for the design of optimal policies on new transportation infrastructure in societies that lack such infrastructure. For example, given the positive "average" impact of colonial railroads on African urbanization, is it the case that the adoption of new transportation technologies always generates a net benefit everywhere in Africa compared to pre-exisiting alternatives? Specifically, what is the opportunity cost of the colonial railway in Africa, and how is it related to the net impact of the railway on economic growth and development? We address these questions in this paper.

Fogel (1964) offers an important answer to the questions above in the context of the American railway and its presumed indispensability to late-nineteenth century American economic performance. Fogel argues that new technological innovations may be dispensable, and of significantly lower net benefit, if pre-existing innovations can simply be improved. In the context of the American railway, the pre-existing transportation technology and primary opportunity costs of the railway were the extensive waterways and canals that were halted with the advent of the railway. Fogel constructs counterfactual canal networks that would have existed if the railways were not developed. His results indicate that the contribution of the railway on economic performance was substantially smaller than believed (2.5% of GDP). The economics behind the argument was intuitive and the empirical argument thorough, but a major point of contention in Fogel's analyses has been the assumptions needed to construct the counterfactual scenarios and the opportunity cost of the railway.[2]

We revisit Fogel's hypothesis in the context of the colonial railway in Nigeria. We focus on Nigeria because it gives us the benefit of analyzing the net impact of the railway without having to construct "artificial" counterfactual transportation networks. The railroads were primarily constructed to enhance export trade with Europe, but various parts of the country differed with respect to the availability of alternative transportation technologies and initial market access to Europe, as revealed by pre-railway trade volumes.[3] Hence,

[1]See Banerjee et al. (2012), Faber (2014), and Baum-Snow et al. (2012) on railways and roads in China; Donaldson (2016) on railways in India; and Donaldson and Hornbeck (2016) and Bleakley and Lin (2012) on the impact of the railway and portage sites in the United States. Jedwab et al. (2015), Jedwab and Moradi (2015), Fourie and Herranz-Loncan (2015), and Storeygard (2016) provide evidence on the positive economic impacts of roads and rail networks in Africa.

[2]Contemporary critiques of the analyses in Fogel (1964) may be found in Taylor (1965) and Rothstein (1965). Donaldson and Hornbeck (2016) recalculate the benefit of the railway using changes in market access and land values to find that the railway had larger impacts than implied by Fogel's analyses.

[3]The sharpest differences in market access are those between the North and the South as shown in Figure 1. The colonial railroads connected the interior of the country to coastal ports, but the South, by virtue of

we test the hypothesis, advanced by Fogel (1964), that the railway would have smaller impacts, or higher opportunity costs, in areas with viable pre-existing alternatives by comparing the estimated long-term and short-term impacts of the railway in areas with and without pre-existing alternatives. In doing so, we uncover new facts about the economic impact of the colonial railway in Nigeria, and document some conditions under which this new transportation technology had sustained impacts on local development.

We proceed in a number of steps. First, we present a framework that enables us to causally estimate the long-run impacts of the railway on local economic development. Based on individual and household data from the 2008 Nigerian Demographic and Health Survey (DHS) and railway data from the Digital Chart of the World (DMA, 1992), the framework involves the use of spatial discontinuities in railway access within otherwise homogeneous areas, and an instrumental variable approach involving the distance to straight lines between nodes[4] as an instrument for connection to the railway line. We use the framework to investigate the differential impacts of the railway in the North and the South of Nigeria, areas with significant differences in alternative transportation technologies. Our main finding is that the colonial railway has neither a short-run nor a long-run economic impact in Southern Nigeria, but it has large positive impacts on local development in the North. This is true for broad indicators of economic development measured at the individual and household level. These measures include human capital, occupational characteristics, media access, household wealth, and urbanization.

Second, we analyze historical urbanization and city growth data from Jedwab and Moradi (2015) in order to establish that the heterogeneous impacts of the railway we document were also present (or absent) in colonial times. The non-impact of colonial railways in the South and its large positive impact in the North are stable over time and have persisted long after the railways became dysfunctional. These empirical findings are consistent with a theoretical model in which, in the North, railroads improved market access to Europe and encouraged the concentration of production factors in connected localities, inducing a spatial equilibrium that persisted in the long run even after the demise of railroads. This was possible because of the lack of viable alternative transportation technologies, and low opportunity costs of railways, in the North. In the South, however, railroads did not significantly change the initial spatial equilibrium because of viable pre-existing alternatives, and larger opportunity costs. Consequently, they hardly had any short- or long-run local development impacts.

Third, we present evidence from historical sources in support of the mechanism described above. We document the fact that prior to the construction of the railway, Southern

its proximity to the sea, already had viable alternatives, such as roads and waterways, which enabled trade with Europeans, and these alternatives co-existed with the railway (unlike the canals in the United States that largely declined). Such alternatives did not exist in the North, where the railroads were essential in opening this region to the European trade and shifting trade from the Sahara to the coast.

[4]Hypothetically, this is the shortest path between railway nodes.

Nigeria already had a large and significant export trade with Europe, and earned a reputation as the "oil palm coast" well before the twentieth century (Law, 2002). Following the introduction of the railway, this trade did not grow significantly, compared to double digit growth rates in the North. In the early twentieth century when the construction of the railways started, exports of the main Northern crops (groundnuts and cotton) were close to zero. They grew exponentially during the period of railway expansion. Over the same period, from 1900 to 1949, exports of palm produce grew at an annual rate of only 1.9%.

Furthermore, railway adoption rates were significantly lower in the South, with less than 30% of Southern crops being shipped by rail compared to over 80% for most Northern crops. We find that the proportion of the main Southern crops that were railed to the coast significantly declined during the period of railway expansion, which indicates that railways did very little to stimulate economic activities in these areas. We show evidence on shipping volumes and costs that indicate that the low adoption rates in the South are due to higher opportunity costs. While the railway decreased transportation costs in the North by more than 65% compared to roads, our calculations reveal that they were significantly more expensive than alternatives (roads and rivers) in the South. This is true when we use information on prices, shipping volumes, and distances in our calculations of shipping costs. It is also true when we use information on how the adoption of the railway would have changed shipping distances to make simple inferences on reductions in transportation costs. For example, average distance to a river is 24 $km$ in the South compared to an average distance of 65 $km$ to a rail station.

Lastly, we show that railroads have a persistent effect only in areas where they increased access to the European market. Using proximity to the coastal ports as a proxy for market access, we find that the effects of railroads are larger for individuals living in localities with initially lower market access. For individuals who live in localities with higher pre-existing market access (localities closer to ports), these effects are small and statistically insignificant. Hence, the non-impact of the colonial railway in the South is consistent with the fact that it did not significantly improve market access in this region which, because of its geographical proximity to the sea, already had routes and waterways connecting it to various ports. This is direct evidence for the view in Fogel (1964).

To uncover the long-term impacts of colonial railroads on local economic development, we exploit a number of empirical stragies, falsification exercises and robustness checks. Our first empirical strategy is to compare areas connected to the railway to areas unconnected to the railway within localities, or states, that were targeted. The assumption is not that states were exogenously connected, but that locations within a state were connected exogenously. In other words, the precise location of the railway, within a targeted state, was exogenous to local characteristics related to current development. For example, while the railway was intended to connect the areas surrounding Kano known for being suitable for groundnut production, the precise location of the line within the Kano area was plausibly exogenous to contemporary or future development. We claim that within a state, the railway was not systematically placed in the most developed localities or in localities that had the most potential for growth.[5]

To provide evidence that localities connected to the railway and non-connected localities were similar, we compare railroad and non-railroad localities with respect to broad geographic and climatic determinants of development: presence (and size) of early cities and Christian mission stations, temperature, rainfall, elevation, soil nutrient retention, soil workability, elevation, oil palm suitability, cocoa suitability, cotton suitability, and groundnut suitability, taken from the FAO GAEZ database (Fischer et al., 2008). We do not find railroad and non-railroad locations to differ on these time-invariant geographic characteristics that would have been important for prosperity in an agricultural society. Furthermore, we do not find that connected localities are more likely to have a Christian mission station, or to be connected to the road network or to have a river running through their local area.[6] Our estimates, which show large positive impacts of the colonial railway in the North and its non-impact in the South, are robust to the control of the aforementioned factors in addition to other individual and household variables, including ethnicity and state fixed effects.

Nevertheless, in the absence of information on all the factors that contribute to local development, we are unable to completely rule out the claim that railroads were endogenously placed within states. In order to address further endogeneity concerns, we use an instrumental variable approach. We compute the distance to straight lines joining major nodes and use it as an instrument, for being connected to the railway line. This identification approach has been implemented in Banerjee et al. (2012), Jedwab et al. (2015), and Jedwab and Moradi (2015), among several others. Once again, within a state, the straight line connecting nodes is the hypothetical line that, in theory, would have minimized construction costs, all else equal. Deviation from this line therefore might reveal the extent to which the actual rail trajectory might not have been chosen randomly by the colonial government.[7] The identification assumption in this empirical strategy is that straight lines between nodes affect economic development only through their correlation with actual lines. In implementing this method, we exclude observations in nodes, as these connected

[5]In fact, there are examples where the railway was located in less prosperous areas for a variety of geographic and other local idiosyncrasies. For example, the railway in "Lagos" began in neighboring Iddo because Lagos itself was an island. A second example is the line that terminated in the state of "Oil Rivers", which in fact ended in Port Harcourt, a city built from scratch, as opposed to more prosperous pre-colonial ports such as Bonny, Calabar, New Calabar, and Opobo.

[6]The local government is the smallest administrative unit in Nigeria with an average area of 1020 $km^2$, and median area of 705 $km^2$, and serves as the primary measure of the "local area" in which the individual lives. Individuals are identified by their DHS clusters which we refer to as localities, and we match localities to the local government area that they belong to.

[7]This argument ignores other geographic and climatic conditions that might call for deviations from straight lines between nodes. Deviations from the straight lines might not be endogenous to the local economic development of connected areas if they are motivated by technical or geo-climatic characteristics of the localities the railway passes through.

locations might have been endogenously chosen. We do not find the instrumental variable estimates to be very different from the estimates based on spatial discontinuities. We continue to find that the railway has a positive effect in the North and no effect in the South. Interestingly, the first stage of the instrumental variable results reveals that most of the geo-climatic and other local area variables are not significant determinants of connectedness to the railway lines.

We perform several identification checks. We show that our results are not driven by alternative transportation means. We also estimate the effect of placebo railway lines on our outcome variables. These lines are segments that were surveyed and proposed for railway construction, but were not constructed. They were abandoned for a variety of reasons unlikely to be related to short- or long-term economic development, such as the turnover of officials in charge of colonial railways and the conflicting interests of the colonial government (Jaekel, 1997). If the effects of the railway we identify using our instrument do not reflect causality but the connection of areas that would have continued to develop independently of the railway, we would expect placebo lines to also have significant impacts on development as they were intended to connect developed nodes. However, if our instrument reflects the developmental impact of the railway on localities closer to a straight line connecting nodes, we would expect the placebo lines to have no impact. Indeed, we do not find the various placebo lines to have any economic effect, whether the effects are estimated for the whole country or separately for the North where the railway had a significant impact.

Additionally, we analyze the impact of the railway by comparing railway localities to localities close to placebo lines. Precisely, we estimate the effect of being within 20 km of the railway line relative to being within 20 km of the placebo line. We find a large economic effect of the railway in the North. In the South, the effect is close to zero and is not statistically significant. These findings provide further evidence that the impact of the railway in the North and its non-impact in the South are indeed causal, and not merely driven by being close to a "straight line" connecting early urban areas. This is especially true if the placebo lines were not constructed for idiosyncratic reasons, as Jedwab et al. (2015) and Jedwab and Moradi (2015) argue.

We also estimate the impact of railways differentially by the timing and purpose of their construction. If the railway lines were constructed for endogenous reasons, in places where colonial officials knew would develop faster, then we would expect the estimated impacts to vary based on the reason and timing of construction. For example, if the railway was endogenous, then one would expect the railway to be constructed earlier in places of greatest benefit, or to have larger benefits in places where it was necessary for economic reasons (versus political reasons). We do not find that railways that were constructed earlier had greater impacts than those that were constructed later. The results also show that there are neither large nor significant differences in the estimated local effect of the railway by reason for construction, be they political/administrative, agricultural, or for the transportation of minerals.

Our results are also robust to a variety of other confounders. In addition to the geo-climatic variables discussed above, our estimates are robust to controlling for the presence of mission stations within the local area, the presence of rivers, and also distance to a major road (possibly endogenous to railways). The estimated impacts of the railway lines are also found for men and women, migrants and non-migrants, and areas with and without mission stations. We detect a positive impact of the railway when we exclude local areas that are run through by railway tracks, nodes or stations. In addition, our results are robust to the exclusion of oil producing areas of the South that might have altered the post-railway spatial equilibrium. Using a continuous measure of connectedness to the railway line (distance to a railway line) instead of the discrete measure in our main specification (being within 20 $km$ of a railway line) yields quantitatively similar results. The estimated impacts of the railway lines are robust to various definitions of the control group: individuals outside 20 $km$, between 20 to 40 $km$, between 40 to 60 $km$, etc, and robust to using proximity to lines that were planned but not constructed for a variety of reasons (placebo) as the control group. We interpret these findings as showing that there are no significant negative spillovers to adjoining localities, and that being close to a line joining major nodes is by itself not associated with development today, respectively.

To the best of our knowledge, this is the first study to evaluate the impact of opportunity costs on the net benefit of the railway - which we call the Fogelian hypothesis - by estimating the impacts of new transportation infrastructure in areas with different levels of viable pre-existing alternatives. A few recent studies, however, have examined the long-run economic impact of infrastructure investments. Huillery (2009) finds that colonial investments in Francophone West Africa have had a long-run economic impact. Banerjee et al. (2012) examine the effect of transportation networks on economic development in China, and show that proximity to transportation networks affects positively but modestly GDP per capita. Baum-Snow et al. (2012) find that transportation infrastructure decentralizes economic activities to adjoining areas. Donaldson and Hornbeck (2016) analyze the historical impact of railroads on the total value of the United States' agricultural land and finds a positive effect, and Donaldson (2016) shows that the colonial railway had large impacts on colonial Indian trade and development. Bleakley and Lin (2012) document path dependence in the economic effect of historical portage sites in the United States.

Other papers have looked at the effect of colonial railroads in Africa, an important form of investment that accounted for about one-third of the colonial budgets (Chaléard et al., 2006). Closer to our paper are Jedwab et al. (2015) and Jedwab and Moradi (2015) who show that colonial railroads have had a sustained impact on African urbanization. Similar to the argument we provide to shed light on the long-run economic impact of colonial railroads in Northern Nigeria, Jedwab and Moradi (2015) explain that, in Ghana, colonial railways lowered trade costs and boosted the cultivation of cocoa in railroad locations, fostering the emergence of cities in these locations. This initial spatial equilibrium persisted because railroad locations facilitated the coordination of subsequent investments.[8]

Our paper differs from these studies in an important respect. First, we are interested in the analysis of conditions under which colonial investments transform local economies in the short and long run. More precisely, we test the Fogelian view that the railway (and other transportation infrastructure) would have smaller impacts in areas with viable pre-existing alternatives. In particular, we do not find that colonial railways have had any local economic impact in Southern Nigeria, and areas closer to the coast, in contrast to their positive effect in the Northern regions of the country. In further contrast to the general conclusions in Jedwab et al. (2015) and Jedwab and Moradi (2015), and the average impacts of the railway we also estimate in Nigeria, we find no evidence that colonial railways were the engine of urbanization in Southern Nigeria. In fact, most cities in the South do not lie along the railway line, while almost all large cities in the North are connected to the railway. This is important for understanding the policy implications of recent studies of the impact of transportation investments. Our results suggest that these investments are most worthwhile in areas where it would improve market access and stimulate new trade.

The results on the impacts of railways on individual-level developmental outcomes are also of independent interest to studies of African urbanization. Fay and Opal (2000) and Jedwab and Vollrath (2015) document the poor economic performance of several urban areas in developing countries, compared to historical examples from other regions. We find that, in areas without pre-existing viable transportation technologies, connection to the colonial railway increased urbanization, and that individuals living in these urban areas are more educated, more literate, more likely to work in professional occupations, less likely to work in agriculture, more likely to engage with mass media (TV, radio, newspapers), and live in wealthier households. This suggests that, while urban areas are not industrializing or growing as fast as one would expect from historical examples, urban areas connected to the railway are still generally better off than surrounding countrysides.

Our study is also related to the literature on the long-run impacts of European contact on Africa. Many studies have focused on the legacy of colonial institutions and the scramble for Africa. Acemoglu et al. (2001) and Michalopoulos and Papaioannou (2014) respectively look at the effect of colonization and pre-colonial institutions on the long-run economic development of Africa. Nunn (2008), Nunn and Wantchekon (2011), Whatley and Gillezeau (2011), Fenske and Kala (2014), and Okoye and Pongou (2015) investigate the effects of the slave trade on various outcomes related to development. Gallego and Woodberry (2010), Wantchekon et al. (2015), and Okoye and Pongou (2015) analyze the

[8]Also see Storeygard (2016) who underscores the importance of road networks and connection to coastal ports for local economic performance, and Fourie and Herranz-Loncan (2015) who document the importance of the railway in South Africa.

effects of missionary activities. Dev et al. (2016) theoretically study path dependence in human capital accumulation and derive implications for the effect that differential historical exposure to missionary activities had on ethnic inequality in educational attainment in Nigeria. We compare the effect of colonial railways to that of missionary activities in Nigeria, finding that, while colonial railways have had an impact only in the North, historical exposure to missionary activities has had a long-run positive effect both in the North and in the South. This finding is consistent with the Fogelian hypothesis tested in this paper because in both regions, there were no pre-existing alternatives to missionary schooling, viewed as an institutional innovation.

The rest of the paper is organized as follows. Section 2 depicts the historical context of the construction of railroads in Nigeria. Section 3 presents our data and the various identification strategies and robustness checks that we use to assess the local impact of the railroads. The corresponding results are described in section 4. Section 5 analyzes the North-South differences in the long-run impact of the railway. Section 6 disscusses the dynamics of the path of the impacts of the railway and compares short- and long-run effects. Section 7 documents the mechanisms underlying the heterogeneity of the impact of railroads. The final section concludes.

## 2 Historical Background

> So vast an area as Nigeria, comprising in all some 380,000 square miles... cannot be commercially developed except by railways — *p. 19 Of the Colonial Report of Northern Nigeria, 1900-1901, as quoted in (Onyewuenyi, 1981, p.65)*.

Toward the end of the nineteenth century, after the area now known as Nigeria officially came under British control, the colonial government began to seek out ways of linking the interior of the country to its ports in order to facilitate export trade. The construction of the railway was seen as an effective means of moving goods and services from the interior of the country to the coast. Construction of the railway lines largely occurred between 1898 and 1930, with an additional extension completed after independence in 1964.[9] The railway was generally constructed to open up the country to export trade with Europe. Three specific reasons were given for the construction of the various lines: agricultural, mineral exploitation, and political or administrative reasons (Taaffe et al., 1963; Onyewuenyi, 1981).

Table 1 shows the dominant motivations for each of the lines constructed between 1898 and 1964. It establishes that the export of agricultural products was the main motive for the railway. Of all the segments shown in Table 1, only Zaria-Jos-Bukuru and Kaduna-Kafanchan were not constructed for agricultural exploitation reasons. In terms of spatial

[9] A rail line joining Abuja to Kaduna was built between 2011 and 2014. Since it was constructed after the dates of our outcome measures, it should have no major bearing on our results.

distribution, the colonial railroads were slightly more extensive in the Northern region which covers 4/5th of the country's area.[10]

## 2.1 Alternative Transportation Modes

Before the railways, transportation of goods was done through head portage, bicycles, animals, cart and inland waterways. In the North, there were caravan routes going through Timbuktu to major agglomerations such as Kano and Sokoto and on to North Africa (Cairo, Tripoli). One consequence of the railway, as we discuss later in the paper, was to redirect Northern trade from Trans-Saharan routes to the coastal ports.

The most important transportation mode for goods before the advent of railways were inland waterways. Many rivers, their tributaries, and creeks traverse the coastal plains of the country. In the South, between the coastal ports of Lagos and Opobo for example, the abundant creeks allow transportation of produce and many ports were installed along the way: Epe, Sapele, Warri, Forcados, Burutu, Brass, Degema, and Port Harcourt (Onyewuenyi, 1981). These river networks, as well as direct access to roads using bicycles, exposed the South to trade with Europe long before the railway was constructed.

While the country's two main rivers, the Niger and the Benue, run through the North, the rivers are navigable only for part of the year and for a fraction of the distance they cover. They are heavily dependent on water levels in the rainy season, and the Niger itself is filled with dangerous rapids. As a result, the only available means of consistent transportation from the Northern parts of the country to the coast was through roads, which were not viable because of the enormous distance and other dangers of road transportation in pre-colonial Nigeria.[11] For example, Hodder (1959) in a study of tin-mining in Jos estimated that the road journey from the mines to the coast took 35 days by road, and while this was tolerable for mining, it was not conducive to agricultural exports. These problems with river and road transportation meant that most areas of the North were cut off from export trade prior to the construction of the railway.

## 2.2 Railway Construction

The railway construction was done in three main phases. The first phase consisted of initial penetration lines. The origination points were the ports of Lagos, Zungeru and Baro for the Western line, and Port Harcourt for the Eastern line. The Western line originated in Lagos in 1898 and reached the Niger river at Jebba in 1909. The construction of the Eastern line began in 1913 in Port Harcourt and reached Enugu by 1916. In the second phase of the railway development, more interior centers were linked to the ports with lines such

[10] There is an average of 3.1 rail length meter per area square kilometer in the South and of 3.4 in the North.

[11] On average, Northern populated areas and coastal ports are more than 600 km apart.

as Baro-Kano and Enugu-Kaduna. By 1927, both main North-South links were established giving Northern areas access to the ports of Lagos and Port Harcourt. Building branch lines and extensions such as Zaria-Kaura Namoda or Kano-Nguru made up the third stage of railway development. At the end of this phase, in 1931, the railway was 3,067 $km$ long.

New centers of economic activities quickly appeared along the newly constructed railroads. By the time the main lines were built, more than 200 buying and selling stations had emerged along the railway lines (Onyewuenyi, 1981). One of the fastest growing centers was the coastal town of Port Harcourt which was chosen as a terminal node of the Eastern line in 1913 before it even existed as a town. Because of its deeper harbor and direct access to the hinterland, Port Harcourt had developed, by the 1930s, as the second largest port of the country, at the expense of previously established non-railway ports within the region such as Bonny, Opobo, or Degema. Similarly, in the "Lagos" area, the railway did not begin in Lagos itself but in another town known as Iddo because Lagos is an island which would have made construction more expensive. We exploit these local idiosyncrasies and discontinuities to motivate one of the empirical strategies we use to estimate the impact of the railway.

## 2.3 Growth of Export Agriculture Following the Railway Construction

> In the Northern Provinces, the history of export cotton production, like that of groundnut has been closely linked with the history of railway expansion, and it was not until the railway reached Kano in 1912 that the export cotton production attained any importance — *Lamb 1925, p. 19*.

The incentives to produce more than what was needed for consumption were weak in remote areas in the North of the country, especially in areas poorly connected to rivers. The advent of the railways dramatically changed the trade opportunities available to these areas. The railways were used almost exclusively for goods transportation as more than 90% of rolling stock units were devoted to goods service. Over the period 1901-1950, an average of 2/3 of these goods were agricultural products. According to the Colonial Reports of 1913, only a year after the first railways were built in the North of Nigeria, the value of Northern agricultural exports jumped by 150% (groundnuts by 666%, benniseed by 157%, gum arabic by 133%, cotton lint by 45%, hides and skins by 41%, and sheanut products by 20%). Acreage under cultivation increased at all station areas.

The railway stations allowed the concentration of markets along the railroads making possible the clustering of a traditionally scattered population and agricultural production. The development of export agriculture was initially limited to "the irradiation area of the railways, the inter-regional roads and auxiliary local roads" (Schätzl, 1973, p. 89). As we show in the next sections, the incredible boom in export agriculture that followed the railway construction had long-run consequences on the human and social development of people living in areas connected to the railways. This impact was concentrated in the North of the country, which was the main beneficiary of the introduction of railways.

Next, we describe the empirical strategies and data we use to estimate the long-term impact of the colonial railway in Nigeria, and how this impact differs according to initial market access.

# 3 Data and Empirical Strategy

## 3.1 Data

Data on colonial railroads in Nigeria come from the Digital Chart of the World (DMA, 1992). These data are combined with individual-level data from the 2008 Nigeria *Demographic and Health Survey* [DHS] (NBS and ICF International, 2008) to estimate the long-run impacts of railroads. The DHS uses a two-stage probabilistic sampling technique to select clusters in the first stage and households in the second stage. In general, DHS clusters are census enumeration zones, to which we will sometimes refer simply as localities. Using DHS-provided information on the geographical coordinates of each such locality, we match individuals to local areas[12] and rail networks.

The DHS provides information on each individual's characteristics including age, sex, migration status, religion, ethnicity, and area of residence. Individuals without specific information on ethnicity are dropped from the analysis. Information on each of our individual-level outcome variables - years of schooling, literacy, type of employment (professional or agricultural), and the frequency at which an individual reads newspapers, listens to the radio, and watches TV - is also available in the DHS. Household-level variables include household wealth and size as well as the gender and age of the household head.

We complement our outcome dataset with panel data on urban population density and city presence in 1900, 1960, 1980 and 2010, from Jedwab and Moradi (2015). This allows us to analyze urbanization outcomes and explore short- and long-run effects.

We collect detailed information on geographic, climatic, and soil conditions from the FAO GAEZ database (Fischer et al., 2008). Specifically, we gather information on average annual rainfall (in mm), average annual temperature (in degree Celsius), elevation (in meters), two important soil characteristics (nutrient retention capacity and workability)[13],

[12]The local government area is the smallest administrative unit in Nigeria with an average area of 1020 $km^2$, and median area of 705 $km^2$, and serves as the primary measure of the "local area" in which the individual lives. Their boundaries are obtained from the GADM database of global administrative areas (UC Berkeley, 2014).

[13]The soil characteristics are measures on a 4-point scale ranging from no or slight constraints (1) to very severe constraints (4).

and suitability for the production of cocoa, cotton, groundnuts, and oil palm.[14,15]

We also collect information on the presence of primary roads and major rivers in each local area. These data are supplied by DMA (1992) and are available in Hijmans et al. (2001). Lastly, we collect data on the presence of Christian mission stations in local areas in 1928 by combining maps published in Ayandele (1966) and Roome (1925). Altogether, we have information on over 30,000 individuals living in 22,798 households belonging to 845 clusters (localities) spread out over 550 local government areas in 37 states. These individuals belong to 30 major ethnic groups that make up over 90% of the country's population.

## 3.2 Identification Strategies

Comparing railway-connected and railway-unconnected areas does not necessarily yield the long-term impact of the railway because these areas might be different on other dimensions that are relevant to economic development. We use a mix of strategies to deal with the possible endogeneity of railway placement. We first use a fixed-effect approach that compares connected and unconnected areas within states that were targeted or not to host railroads. We then exploit an instrumental variable strategy that is based on straight lines between major railway nodes. We complement these strategies with a host of falsification exercises and identification checks using placebo lines, heterogeneity along the timing of and motives for rail line construction and various definitions of control groups.

### 3.2.1 Fixed Effects

The railway was intended to connect large areas suitable for agricultural and mineral exploitation to the coast. In order to avoid comparing targeted areas with non-targeted areas, we use in our regressions state fixed-effects as states are the closest administrative level to capture these areas of interest.

Within states, we compare areas that are close to the railway lines to areas that are further away. A concern with this strategy is that even within targeted areas, the railway might be placed in areas that are systematically different on dimensions relevant for economic development. Historical accounts of railway placement tend to indicate otherwise. Engineering decisions that took into account elevation and other considerations relevant to the cost of the railway construction were central to determining the exact location of the railway within a targeted state. For instance, the railway line that terminates in Lagos state actually terminated in a small locality called Iddo because the city of Lagos was an island

[14]Crop suitability is the average estimated agro-climatically attainable yield in kg/ha for rain-fed agriculture, using medium or low inputs within the local area because that is the dominant form of agriculture in Nigeria (FAO, 2016).

[15]These measures are provided by the FAO for cells of 30 $\times$ 30 arcseconds (approximately 0.9 $km^2$ at the equator).

and having the terminus in such a city would have been very expensive. A second example is the line that terminates in the "Oil Rivers" which is now the city of Port Harcourt. This location was chosen as the city did not even exist, as opposed to prosperous pre-colonial ports such as Bonny, Calabar, New Calabar, and Opobo, because of its deeper harbor and more direct access to the hinterlands.

Comparing observables within railway-connected and unconnected area lends further support to the claim that the exact placement of the railway within targeted areas does not appear to systematically be in areas more developped or more likely to be developped in the future. We refer to localities within 20 $km$ of a rail track as connected areas. Table 2 presents summary statistics for various observable characteristics in connected and unconnected areas. Baseline observables such as crop suitability, geo-climatic factors and soil constraints do not exhibit significant differences in observable characteristics between connected and unconnected localities. Strikingly, connected individuals are not more likely to live in local areas with mission stations, nor in local areas with primary roads, nor in areas crossed by a major navigable river. Connected areas had fewer cities in 1900 than unconnected areas and were not more likely to have cities before the introduction of the railway or to have larger urban populations.

Nevertheless, we control for all these geographic, climatic, population-based and other pre-railway obervables. For individual and household-level outcomes, we control for additional factors such as age, religion, ethnicity and household size in order to get more precise estimates of the railway effect.[16]

Specifically, our baseline specification for individual- and household-level outcomes is:

$$Y_{i,h,a,e,s} = \phi + \beta R_{i\leq 20} + \mathbf{X_a}\mathbf{\Lambda} + \mathbf{X_i}\mathbf{\Pi} + \mathbf{X_h}\mathbf{\Gamma} + \alpha_\mathbf{s} + \gamma_\mathbf{e} + \epsilon_{\mathbf{i,h,a,e,s}} \quad (1)$$

The parameter of interest , $\beta$, is the effect of living within 20 $km$ of a railway track on outcome $Y$. The outcome is measured for each individual $i$ in household $h$, who lives in local area $a$ in state $s$, and belongs to ethnic group $e$. The outcome variables are education (years of schooling, literacy), occupation (professional or agricultural worker), media access (newspaper, radio, TV), household wealth and the probability of living in an urban area. If the railway has a positive impact on local development today, then individuals who are closer to the railway should be more educated and more literate, and should have greater access to the media. Under this hypothesis, we would also expect individuals in railway areas to be more likely to live in urban areas and in wealthier households and to be non-farm workers.[17]

[16]While we use living within 20 $km$ of a rail line as the measure of connectedness, we confirm that observables are also balanced using other measures of connectedness such as an indicator for having the railway pass through the local area. We also divide the country into 40 $km \times 40$ $km$ grid cells and show that observables are balanced between connected and unconnected grid cells. These results are available upon request.

[17]We categorize an individual as being literate if they can read some or parts of a sentence. We deem

Individual, household, and local area observable characteristics are denoted by $\mathbf{X_i}$, $\mathbf{X_h}$, and $\mathbf{X_a}$, respectively. All regressions involving individual- or household-level outcomes include 37 state fixed effects ($\alpha_s$) and 30 ethnic group fixed effects ($\gamma_e$). Individual controls ($\mathbf{X_i}$) are age, age-squared, and an indicator for being Christian. At the household level, we control for gender and age of the household head, as well as for the size of the household. At the local area level, we control for average rainfall, temperature, soil nutrient retention capacity and workability, elevation, suitability for key colonial area cash crops (oil palm, cocoa, cotton, groundnut), and presence of a mission station in the local area as at 1928. In the remainder of the paper, these local area controls are referred to as baseline controls.

For urbanization outcomes (city presence and city growth), our baseline specification has the same structure as in (1) but the set of controls excludes individual and household-level controls:

$$Y_{a,s} = \phi + \beta R_{i\leq 20} + \mathbf{X_a}\mathbf{\Lambda} + \alpha_\mathbf{s} + \epsilon_{\mathbf{a,s}} \quad (2)$$

### 3.2.2 Instrumental Variable Approach

To further address endogeneity concerns, we adopt an instrumental variable approach, similar to that used in Banerjee et al. (2012) and Jedwab and Moradi (2015). We exploit the distance to straight lines joining major nodes and use it as an instrument for being connected to the railway line. We also exclude individuals living in nodes from the sample. The identification assumption is that, besides its correlation with the railway line, a straight line connecting nodes is unrelated to economic development.

The major nodes are chosen to be major historical cities existing at the time of the introduction of the railway such as Lagos, Abeokuta, Ibadan, Kano, and junctions in the middle of the country such as Kafanchan. We connect these major nodes in the spirit of the railway introduction, that is, by finding the minimal path to connect areas of interest to the coast for each railway line defined by periods of planning/construction (pre-1912, 1916-1930, and 1964). The result of this simple algorithm is shown in Figure 2.

The instrument for being within 20 $km$ of a rail line is defined as the distance to a node-joining straight line. All observations within the same local government area as a node are dropped. Only "intermediate" observations are used to estimate the specification.

### 3.2.3 Identificaton Checks

We use a number of exercises to confirm the causality of the effects we uncover.

them as utilizing media resources (newspaper, radio, TV) if they use these resources at some point during a month. We adopt broad and inclusive definitions in order to provide conservative estimates, and deal with concerns about arbitrary cutoffs for inclusion into these categories. Our results are in fact stronger if literacy is limited to individuals who can only read whole sentences, or to individuals who use media at least once a week.

**Alternative Transportation Technologies.** A natural concern with our empirical strategies is that distances to straight lines connecting nodes could be correlated with roads and other transportation technologies and that the impacts we bring to light might be unrelated to the railway line. We address this concern by demonstrating that the estimated impacts of the railway line are robust to the inclusion of other transportation technologies.

**Placebo Lines.** A way to test whether the effects we measure have to do with the railway and not merely with being close to lines joining nodes that could have been linked by the railway or any other important transportation technology is to look at areas closer to placebo lines. Placebo lines are segments that were surveyed and proposed for railway construction but were never actually constructed. Given the prohibitive costs of railway surveys, these segments were seriously considered. They were ultimately abandoned for a host of plausibly exogenous reasons that have to do with the turnover of officials in charge of the Nigerian railways and the conflicting interests of the British colonial decision-makers (unlikely to be related to local economic development).

As Figure 3 shows, the placebo network that we reconstruct covers an extensive part of the country. In the Southern part of the country, these lines were meant to connect already existing cities. Following independence, these very early cities were finally connected by roads. Thus, in this exercise, we control for the effects of alternative technologies (roads and rivers). We are aware of the fact that the placement of roads, presented in Figure 4, might be endogenous to the existing railways. Hence, the results of this identification check are only suggestive and should be interpreted with some caution. Our hypothesis is that if the effects that we uncover are indeed causal, one would expect them to disappear or reverse once we replace actual railway lines with placebo lines, in areas where placebo lines do not coincide with roads or waterways.

**Varying Control Groups.** Another concern with our identification strategies is that we might be merely identifying differences between localities within 20 $km$ of a railway and those very far away, such as clusters over 80 $km$ away from the railway line, which may not be good control clusters. This is because clusters that are very far away are more likely to be different in dimensions not captured by our control variables. To address this concern, we break down the control group into various distances: clusters within 20-40 $km$, 40-60 $km$, 60-80 $km$, and farther than 80 $km$ of the railway line.[18] We then re-estimate equation (1) using the disaggregated distances as different control groups, and exclude the indicator for individuals living beyond 80 $km$ of the railway from the regression. The coefficient on living within 20 $km$ of a railway line now represents the impact of the railway relative to individuals living beyond 80 $km$ of a railway. This strategy allows us to compare individual

[18] In our dataset, we find that 32% of individuals live within 20 $km$ of a railway track, 12.3% within 20-40 $km$, 13.57% within 40-60 $km$, 11.42% within 60-80 $km$, and about 30.61% farther than 80 $km$ of a rail line.

outcomes across multiple distances and to account for potential spatial spillovers beyond 20 $km$. A similar exercise consists of estimating equation (1) on samples limited to areas within 40 $km$, 60 $km$ and 80 $km$ of the railway. Both exercises yield similar results.

**Actual Lines versus Placebo Lines.** As explained above, one could also imagine randomly assigning surveyed localities to be connected to two groups: the group of localities connected to the railways, and the group of localities that were surveyed but not connected (placebo lines). This makes the use of areas close to placebo lines a powerful control group to check our identification. We implement this by re-estimating specification (1) on the clusters that are within 20 $km$ of rail or placebo lines. This analysis would effectively yield a causal effect of the railway if, as argued above, placebo lines were not constructed for exogenous reasons (or reasons not related to long-run economic development).

**Reasons for and Dates of Construction.** As discussed in the historical section, the railways were built at different time periods and for different reasons. If areas close to the railway were substantially different from areas further away, this might be revealed by the stated reason and timing of construction. For example, railways constructed for strictly agricultural reasons might have been routed through more productive localities and those constructed for political reasons might have been subject to political wrangling. In the same vein, lines constructed earlier might have been constructed to link vital productive localities. Following Table 1, we categorize reasons for construction into 3 non-mutually exclusive categories: agricultural, mineral, and political. There are several dates of construction which we aggregate into three categories: pre-1912 (before WWI), between 1912 and 1930, and post-independence (completed in 1964).[19]

We estimate the effect of the railway conditional on motive or date of construction, $M_R$, using the following regression equation:

$$Y_{i,h,a,e,s} = \phi + \beta R_{i\leq 20} \times M_R + \mathbf{X_a\Lambda} + \mathbf{X_i\Pi} + \mathbf{X_h\Gamma} + \alpha_\mathbf{s} + \gamma_\mathbf{e} + \epsilon_\mathbf{i,h,a,e,s} \quad (3)$$

We exclude the indicator for motive or date itself, because the presence of a motive indicates the presence of a rail line. A homogeneous impact of the railway along the timing and motives of construction dimensions would reassure us that we are identifying the effect of the railway and not merely the effect of being close to particular subsets of nodes.

[19]Note that while the date of construction is unique, the reasons given for construction are not. In fact, 97.2% of connected individuals are close to a rail line constructed for agricultural reasons, 30% live close to lines constructed for access to minerals, and 64.5% of individuals live close to a rail line constructed for political reasons. For construction dates, 36.7% of connected individuals live close to a rail line completed before 1912, 23.6% live close to a rail line completed between 1912-1930, and 39.7% live close to a rail line completed after independence.

**Other Identification and Robustness Checks.** We complement our empirical strategies with the following identification and robustness checks. We test the robustness of our results to other measures of connectedness such as continuous measures based on the distance to the rail lines or the distance to the closest railway station or an indicator for being in the same local government area as the rail line. We also divide the country into 40 $km \times$ 40 $km$ grid cells and use the indicator for being in the same grid cell as the rail line as another measure of connectedness to the railway.

We also test the robustness of our results to: limiting the sample to rural areas only, using Conley standard errors to correct for spatial autocorrelation, limiting the sample to migrant individuals or to non-migrants, excluding areas with mission stations, excluding local government areas with rail tracks or rail stations, and even excluding areas within 20 $km$ of a railway station.

# 4 Average Effect of the Railway: Countrywide Estimates

We first estimate the country-wide effect of railroads in Nigeria. This analysis is of independent interest in light of the current literature on the long-term impact of colonial railroads because it reveals the economic effect of railroads on individuals beyond their effect on urbanization and other aggregate-level outcomes.

## 4.1 Fixed Effects Results

Table 3 presents results of the estimating specification (1). Standard errors are clustered at the local government area level in order to deal with arbitrary correlation between localities (clusters) within local areas. As reported in Column 1, living within 20 $km$ of a rail line increases schooling attainment by 1.3 years on average. This is associated with a 13% increase in the probability of being literate, a 1.6% increase in the probability of working in a professional wage job, and a 7.4% decline in the probability of being an agricultural worker (Columns 2-4). Furthermore, being connected to the rail line is positively associated with media access. Individuals in connected areas are more likely to read newspapers, listen to the radio and watch TV (Columns 5-7). Finally, being connected to the rail line is associated with living in a wealthier household, and it increases the probability of living in an urban area by 18% (Columns 8-9). These results are all consistent with a strong positive impact of proximity to the rail line on individual and household development outcomes.

It is interesting that none of the geographic and climatic variables have a consistent impact on the outcomes. Similarly, we do not find that areas that are more suitable for oil palm, cocoa, cotton, or groundnut are more developed in the present. Importantly, this evidence supports our identification assumption that, within a state, geographic characteristics, and any pre-colonial advantages they might have conferred, are largely unrelated to contemporary development. However, and in accordance with previous studies, we find that missionary activity is strongly associated with development at the local level (Gallego and Woodberry, 2010; Nunn, 2014; Okoye and Pongou, 2014; Cage and Rueda, 2014; Wantchekon et al., 2015). Local missionary activity has a positive impact on years of schooling, literacy, occupational choices, media access, household wealth and urbanization. The impact of the railway and missionary activity, and the non-impact of geographic and climatic variables, are remarkable and speak to the importance of historical circumstances for development at the local level.

## 4.2 Instrumental Variable Estimates

Turning to instrumental variable results, we first present the first-stage estimation in Table 4. The estimates show that a doubling (100% increase) of the distance to a line joining nodes lowers the probability of being connected to the rail line by about 29.4%. Also important is the finding that missionary activity and most of the geo-climatic variables are uncorrelated with rail presence. An exception is elevation which is negatively correlated with probability of connectedness, consistent with a higher cost of building in elevated areas. These first stage results are also taken as evidence that the rail line was not strategically placed within states.

The 2SLS estimates of the impact of being connected to the railway are shown in Table 5. They are generally not statistically different from the OLS estimates. An exception is the estimated impact of being within 20 $km$ of a rail line on the probability of being an agricultural worker, which falls in magnitude from -7.4% (OLS) to -4.5% (IV). The IV approach, along with the other robustness checks results reported below, lend support to our causal interpretation of the impact of the railway.

## 4.3 Identification Checks Results

**Alternative Transportation Technologies.** As shown in the top panel of Table 6, we control for the presence of other transportation technologies, because if, for instance, roads were built close or along some of the railway lines, their impacts could be picked up by our estimates. Indeed, when roads and rivers are controlled for,[20], we find a robust long-term impact of the railroads. The fact that the results are not driven by the correlation of the railway with other transportation infrastructure is important as, otherwise, this might call into question the identification assumption behind the IV strategy. We find that increased distance to roads is negatively correlated with living in an urban area, household wealth, and other measures of local development. Rivers, on the other hand, are not positively

[20]Data on the presence of major roads and navigable rivers are obtained from DMA (1992) and are available in Hijmans et al. (2001).

related to development, consistent with our findings on other geographic variables. However, we do not take the estimated impacts of the road network as causal because they might have been constructed in response to the rail network. These results suggest nevertheless that the estimated impact of the railway cannot be explained away by the road network.

**Placebo Lines.** When we replace the actual railways by placebo rail lines (the surveyed lines that were eventually not constructed), and exclude railway connected areas, most estimates decrease dramatically and lose their statistical significance. As shown in the middle panel of Table 6, the estimates of the coefficients on outcomes such as schooling, literacy, professional occupation, reading newspapers and being an urban resident are significantly smaller and not statistically different from zero.

In the bottom panel, we control for the presence of other transportation technologies, because transportation technologies were built on some of the placebo segments, and their impacts could be picked up by our estimates. Indeed, when roads and rivers are controlled for, the results for the placebo effect, presented in the bottom panel of Table 6, show a clear picture. None of the coefficient estimates is significant at the 5% level.

By providing evidence that straight lines connecting pre-existing cities are not correlated with local development outside of localities connected to the railway line, these placebo results reinforce our causal interpretation of the effect of railroads.

**Varying Control Groups.** As shown in the top panel of Table 7, the impact of the railway on individuals living within 20 $km$ of the rail line is robust to the use of different control groups. The impact of being connected to the rail line does not change significantly when we compare individuals living within 20 $km$ of the rail line to those living within 20-40 $km$ or to those living farther away. In fact, the estimated impact of the railway is stronger when individuals living within 20 $km$ of a rail line are compared to those living a further 20 $km$ away at most. For example, relative to individuals living beyond 80 $km$ of a railway line, being connected is associated with an additional 1.16 years of schooling. However, when compared to those living just "outside", within 20-40 $km$, the impact of being connected is an increase of about 1.5 years of schooling $(1.158 + .343)$ although this difference is statistically insignificant. Thus, extending the control group to all distance groups actually produces a conservative estimate of the impact of the railway as there is no evidence of a positive spillover beyond 20 $km$ of a rail line.

**Actual Lines versus Placebo Lines.** Our results are robust to using within 20 $km$ of placebo lines as a control group, as shown in the second panel of Table 7. These results suggest that even compared to localities close to other straight lines joining major cities and proposed to be connected by rail, localities closer to the railway lines still have better development outcomes today. Overall, our results strongly suggest that being connected to the historical railway line has a positive impact on development in today Nigeria, even though the railway has deteriorated after the country's independence.

**Results by Reasons for and Dates of Construction.** The results of the estimation of specification (3) are shown in the botton two panels of Table 7. The base estimates correspond to distance to lines constructed post-independence (third panel) and to lines constructed for political reasons (fourth panel). We do not find that the estimates vary systematically by date or by reason of construction. The estimates are generally insignificant. Although there is a sizable negative impact of railways constructed for mineral reasons on schooling, literacy, and household wealth, this negative impact is not found for other outcomes. That the estimated railway impact does not vary significantly by timing nor by stated reason for construction is reassuring and further assuages concerns about endogenously placed railways within states.

## 4.4 Additional Identification Checks Results

We also confirm that the estimated impact of the railway is not sensitive to the measure of connectedness. Although the point estimates are not directly comparable, the impact of the railway is robust to the use of a continuous measure of closeness to the rail line.[21] The corresponding results are presented in the top panel of Table A1. Our results are also robust to using other measures of connectedness to the railway such as proximity to railway station (second panel), an indicator for being in the same local government area as the rail line (third panel), and an indicator for being in the same grid cell as the rail line (bottom panel of Table A1).

We continue to find an impact of the railway when we exclude individuals living in urban areas (top panel of Table A2). This important result indicates that our results are not merely driven by urbanization. In the second panel of Table A2, we show that results are robust to using Conley standard errors to correct for spatial autocorrelation.[22] We find that the Conley standard errors are not much different from the cluster-robust standard errors used throughout the paper.

We also estimate the differential impacts of railways by migration status (third and fourth panels of Table A2). While the impacts are larger for migrants, the estimates are not statistically different from the estimated impact on non-migrants.[23] This suggests that

[21] It is defined as $-log(1 + clusterdistance)$ where $clusterdistance$ is the survey cluster's distance to the railway network.

[22] We use the methodology described in Conley (1999), and follow the implementation by Rappaport (2007), with a cutoff of 100 $km$.

[23] Non-migrants are defined as individuals who indicate they have never lived anywhere else beside their current place of residence.

the long-term effects of railroads are not driven primarily by migrants who have higher ability, education or skills.

The fifth panel of Table A2 shows that our results are not driven by missionary activity: our results are robust to excluding areas without mission stations. If anything, the impacts of the railway is stronger in localities without mission stations. This is consistent with the view that the railway increased economic activity and urbanization in connected areas, as in Jedwab and Moradi (2015), and this in turn has had a positive impact on schooling, literacy, and media access. In areas with missionary activity, missions had a positive impact on schooling, literacy, and media access, possibly attenuating the impact of the railway on these outcomes.[24]

The sixth panel of Table A2 indicates that the impact of the railway is sustained when we exclude local areas that contain rail tracks. This is further evidence that the effect that we find is driven by connectedness to the railway line, and not merely by the presence of a railway line or station in the local area. Finally, an important part of the connectedness to the railway line is the proximity to railway stations. The bottom panel of Table A2 suggests that the impact of the railway is attenuated when we exclude areas within 20 $km$ of railway stations.

## 4.5 Urbanization Outcomes Results

In addition to individual and household-level outcomes, we explore urbanization outcomes, namely urban population density and city presence, using the methodology developed in Jedwab and Moradi (2015). We analyze the long-run effect on city presence and urban population (measured in 2010) of the presence of rail tracks within 20 $km$ of a grid cell, controlling for the 1900 population density Z-score, missionary presence, and state fixed effects. Using the standardized score (Z-score) ensures that we measure changes relative to the mean (controlling for time trends), and controlling for 1900 Z-score ensures that we capture relative city growth. The results are presented in the first column of Table 8. We find that the presence of a rail track within 20 $km$ of a cluster has a positive effect on both outcomes. Furthermore, the estimates indicate that controlling for 1960 rates of urbanization, the railway had not further impact on urbanization, suggesting that the impacts of the railway on urbanization largely occurred before independence in 1960.

Overall, the railway has had a robust impact on schooling, literacy, media access, household wealth, and urbanization. The estimated impact is robust to accounting for a host of factors and to a long list of identification checks. We next use this methodology to test Fogel's hypothesis, which is the main contribution of this paper. In particular, we study the differential impact of colonial railroads in the North and the South of Nigeria. We expect

[24]See the following studies for the impact of missionary activity in Africa: Gallego and Woodberry (2010); Nunn (2014); Okoye and Pongou (2014); Cage and Rueda (2014); Wantchekon et al. (2015).

to find little impact in the South because unlike the North, it had viable alternatives to the railways.

# 5 Testing Fogel's Hypothesis: North-South Differences in the Impact of the Rail Line

In his seminal work on the American railway, Fogel (1964) argues that the railway had a minimal impact on economic growth as it was mostly a substitute for planned canals.[25] An implication of this argument is that the local impact of the railway would be minimal in areas with viable alternatives to the railway. Given that the main market in the context of colonial Nigeria was Europe, the coastal South is a good case of a region with alternative access routes (e.g., waterways) and the North, a region with no viable alternatives. Consequently, the Northern and Southern parts of Nigeria present us with an ideal setting to test Fogel's hypothesis.

In addition to having alternative access routes to the European market, the South had established and operated trade routes prior to the construction of the railway, thanks to centuries of pre-railway European trade (Anene, 1966; Crowder, 1980; Falola and Heaton, 2008).[26] The South of Nigeria therefore presents us with a factual case of Fogel's counterfactual. By examining North-South differences, we test whether the impact of the railway differed based on pre-existing access to export markets.

## 5.1 Estimated Impact of the Railway in the North and in the South

Table 9 presents estimates of the impact of connectedness to the railway line on contemporary individuals living in the North and in the South of Nigeria. The top panel shows the fixed effects (OLS) results for both the North and the South, and the bottom panel shows the instrumental variable results estimated by two-stage least squares (2SLS). Both panels suggest that the local impact of the railway in the North is larger than the country-wide average impact. In the North, living within 20 $km$ of a rail line increases schooling attainment by about 2 years on average. This is associated with an 18% increase in the probability of being literate, a 2% increase in the probability of working in a professional wage job and a lower probability of being an agricultural worker. Furthermore, being connected to the rail line in the North is positively associated with media access, higher

[25] However, Donaldson and Hornbeck (2016), using updated data and measures of market access, find larger impacts of the railway, of a magnitude greater than any impact of counterfactual canals or other transportation infrastructure. Unlike the papers by Fogel (1964) and Donaldson and Hornbeck (2016), we do not estimate railway aggregate impacts. We focus instead on impacts on individuals and local areas.

[26] This is reflected in trade statistics. Between 1900-1904, the South was already exporting an annual average of 176,511 tons of palm produce to Europe, while the North was exporting a modest 475 tons of its main crop: groundnuts.

household wealth, and a 20% greater probability of residing in an urban area. The railway has virtually no impact on contemporary development outcomes in the South. There is a significant impact on household wealth and on the probability of watching TV when the model is estimated using the IV strategy, but this result is not robust. All coefficients are economically small and insignificantly different from zero in the South when estimated using OLS.

In contrast to the estimated impact of the railway, we find the impact of missionary activity to be significant in both the North and the South. This is informative about the mechanism behind the impact of these historical events. As Donaldson (2016) finds in India, the railway's impact is primarily due to its impact on trade costs. Arguably, these costs were different in the North of Nigeria because it lacked viable transportation alternatives for export purposes, compared to the South which had pre-existing routes to coastal markets.[27] However, as Wantchekon et al. (2015) and Cage and Rueda (2014) argue, the primary mechanism behind the impact of missionaries is local human capital development. Thus the impact of missionary activity would not be expected to vary based on pre-existing market access, because traders by themselves did not introduce schools in the Nigerian context (Ayandele, 1966).

Turning to urbanization outcomes, we find the same pattern. Although overall the railways have had a sustained economic impact in the country, their effects are only visible in the North, as shown in columns 2 and 3 of Table 8. Indeed, in the North, living within 20 $km$ of a rail line increased the Z-score of urban population by 0.127 in 2010. Similarly, living within 20 $km$ of a rail line increased the Z-score of city presence by 0.157 in 2010, an estimate significant at the 1% level. The equivalent estimates are not significantly different from zero in the South. Excluding two Southern nodes that are mining centers (column 4 of Table 8) leaves our results unchanged.

## 5.2 Differential Impact in the North and South: Robustness Checks

We carry out several robustness checks to confirm our findings on the differential impacts of the railroads in the North and the South of Nigeria. First, we find that the placebo results obtained as an identification check for the country-wide analysis are robust to separate estimations for the North and the South. In the top panel of Table A3, we show that the strong and positive effect of the railway in the North completely disappears once actual lines are replaced by placebo lines. Controlling for the presence of roads and rivers reduce the estimates even further. Interestingly, and as shown in the bottom panel of the same table, estimates of the placebo lines coefficients are significant for most of our outcomes in the South, whereas the local impact of the actual railway lines in this region is statis-

[27] In Section 6 on the mechanisms of the short- and long-term impact of the railroads, we directly show that, relative to other transportation alternatives, railroads significantly reduce the cost of trade in the North but not in the South.

tically insignificant. As before, we show that the placebo effect simply reflects the effect of alternative transportation technologies. Once we control for roads and rivers, most of the coefficient estimates become insignificant. We also analyze the impact of the railway using placebo lines as the control group for the South and the North. For each region, we estimate the effect of being within 20 $km$ of the railway line relative to being within 20 $km$ of the placebo line. The results are presented in Table A4. The effect of the railway on each outcome is large and statistically significant in the North, but is close to zero in the South. Overall, these results indicate that the positive impact of the railroads in the North and their non-impact in the South are causal.

A possible reason for the non-impact of the railway in the Southern local areas might be that, because of the discovery of oil in the South, oil cities might have eclipsed railway cities. To test for this possibility, we restrict the sample to non-oil-producing areas of the South.[28] Table A5 presents the results of this exercise. As was found with the full sample, there is no impact of connectedness to the railway on local economic development in the South when oil producing areas are excluded. There are no significant differences between Southern connected and unconnected areas in schooling attainment, literacy, occupational choices, media access, household wealth, and urbanization. This is true regardless of the estimation strategy, which provides evidence that our results in the South are not driven by oil cities.

Next, we examine the dynamics of the railway impact both in the North and in the South of Nigeria.

# 6 Dynamics and Persistence

A non-existent long-term effect of the railway in the South of Nigeria could mask a short-run effect of the railway that dwindled over time— especially after the demise of railroads — or a non-existent short-run effect that remained stable over time. This raises the question of the stability of the impact of the railway and the comparison of its short- and long-run effects. To explore the path of the railway impacts, we need to examine economic outcomes for which information is available at different points in time. The individual and household outcomes we have examined so far do not meet this criterion but they allow us to look at how the impact of the railway varies across different cohorts. We focus on individuals who have never changed their place of residence (non-migrants) to ensure that the estimates we find reflect conditions in the locality at time of birth. We look at individuals born during the peak of the railway (1948-1975), those born between 1975 and 1984, and the youngest individuals in our dataset, born between 1985-1993. These cohorts were chosen to ensure that observations are roughly of equal number within each category. The

[28]We exclude localities in the oil producing states which are Abia, Akwa, Ibom, Bayelsa, Cross River, Delta, Edo, Imo, Ondo, and Rivers.

results are shown in Table 10.

We first observe that in both parts of the country, the younger cohorts (1985-1993) are generally better educated, and have more media access than the older cohort. This is consistent with the general upward trend in education in Nigeria, with the Southern education trend having a steeper slope (Csapo, 1981; Ajayi et al., 1996). We find no support for the hypothesis that the impact of the railway was stronger for the older cohort, be it in the North or in the South. If anything, in the North, the impact of being connected to the railway line on schooling appears larger for the younger cohort.

For urbanization outcomes, we use panel data from Jedwab and Moradi (2015) on city presence and urban population density in 1900, 1960, 1980 and 2010 to analyze the impact of the railway on these outcomes. This allows us to more clearly separate short-term and long-term impacts.

In a first exploration of the path of effect of the railway, we look at the effect of the railway on 2010 urbanization outcomes after controlling for the 1960 population density Z-score. The results are presented in the three rightmost columns of Table 8. Controlling for the 1960 city presence or population density Z-score, when estimating the effects of the railway on urbanization, renders the effect either not significant or much smaller than before. We view this result as evidence that the effect of the railroad has not changed since independence in 1960.

We then explore shorter-run effects of the railway looking at urbanization outcomes in 1960 and 1980 and find that the effect of the railway is remarkably stable over time. As shown in Table 11, living within 20 $km$ of a rail line increased the Z-score of city presence by 0.104 in 1960 (column 1). Similarly, living within 20 $km$ of a rail line increased the Z-score of urban population by 0.175 in 1960 and 0.182 in 1980 (column 1, last two panels).

Although overall, the railways have had a sustained economic impact in the country, their effects are only visible in the North, which is consistent with what we have found for individual-level outcomes. Indeed, in the North, living within 20 $km$ of a rail line increased the Z-score of city presence by 0.128 in 1960, as shown in the top panel of column 2, with these estimates being statistically significant at the 1% level. Similarly, living within 20 $km$ of a rail line increased the Z-score of urban population by 0.122 in 1960 and .125 in 1980 (column 2, second and third panels). It follows that the stability of the railway effect in the overall sample carries through to the North subsample. In the South, railways did not have any statistically significant effect on urbanization outcomes, neither in the short run nor in the long run.

We perform several robustness checks. We also conduct estimations on reduced samples: in column 4 of Table 11, the point estimate in the South is not only insignificant, but drops significantly when we exclude 2 mining nodes (Enugu and Port Harcourt) which grew as coal and crude oil mining cities, respectively, and later as regional capitals. The estimated impact in the North is robust to the exclusion of all nodes (second to last column).

Overall, these results suggest that the estimated impact of the railway has remained relatively stable over time. In the next section, we show that the non-impact of the railway in Southern parts of Nigeria could be explained by the presence of viable alternatives and relatively higher market access, as hypothesized by Fogel (1964).

# 7 Mechanisms

The North and the South of Nigeria differed in many respects before the introduction of the railway. While we cannot rule out the role of all other possible differences between the two regions, the key differences as far as the impact of railroads is concerned are the availability of viable transportation alternatives connecting Southern areas to the ports of export at the coast and, related to it, the early existence of a more urban spatial equilibrium in the South involving ancient cities and major trading centers. We present evidence indicating that the key mechanism underlying the strong impact of the railroads in the North and the lack of impact in the South is that railroads were dispensable in the South where pre-existing transportation technologies enabled trade with Europe prior to colonization, whereas railroads played a key role in connecting the North to the coast and the European market.

The supportive empirical evidence for the argument above is consistent with a theoretical model in which, in the North, railroads encouraged the concentration of production factors in connected areas and induced a spatial equilibrium that persisted in the long run. This equilibrium persistence, even after railroads had become dysfunctional, owed to the fact that railroad locations helped to coordinate factor investments for each subsequent period. In the South, however, railroads did not change the pre-existing spatial equilibrium, and as a consequence, they did not have any short-run or long-run impact. We present a sketch of the model along with the supportive evidence below. Together, the theoretical argument and the corresponding empirical tests offer a micro-level foundation for, and a validation of, the mechanism underlying Fogel's hypothesis on the dispensability of new transportation technologies.

## 7.1 The Theoretical Argument

The theoretical discussion draws on insights from economic geography, and on studies that attempt to explain the determinants of long-run differences across locations in the concentration of economic activity. Following Bleakley and Lin (2012), we consider an economy in which an agent's indirect utility function, $V$, depends on the density, $X$, of production factors in the location in which the agent resides. We assume that there are strong increasing returns to scale on the one hand and congestion costs on the other hand. This assumption implies a hump-shaped indirect utility function in which the increasing returns to scale are larger than congestion costs over some range of factor density.[29] There is some level of utility $V^*$ that an agent can receive in some other location, and the long-run equilibrium factor density is determined by equating the function $V(X)$ to $V^*$. Our assumptions imply the existence of multiple equilibria as in Bleakley and Lin (2012). Our goal is to analyze the implications of these equilibria and shed light on how the concentration of economic activity varies across locations that differ only by the availability of transportation technologies.

We consider an economy with two locations, denoted 1 and 2, that are identical except for the availability of a transportation technology in location 1. The transportation technology increases market access by lowering the cost of trade, so that the returns to factor investment are greater in location 1 than in location 2. The discussion features two cases. The first case represents the South of Nigeria, and the second case represents the North.

The situation in the South is depicted in Figure 5. The left half (5a) features three curves out of which two are identical: the indirect utility of a marginal agent when no transportation technology is available ($V_0$), her indirect utility when a transportation technology $f$ is available ($V_f$), and her indirect utility when two alternative transportation technologies $f$ and $g$ are available ($V_{fg}$). We assume that $f$ was available first. For instance, the technology $f$ could be the waterways that connected the different ports in Southern Nigeria, and $g$, the railroads that were constructed after the waterways. We also assume that the demand for $g$ is negligible, so that $V_f = V_{fg}$. Here, the long-run equilibrium density $X^*$ is not changed by the new technology $g$. In other words, railway locations and non-railway locations do not differ in their short-run and long-run equilibrium density.

In Figure 5b, we assume that the new technology $g$ shifts the indirect utility function upward, resulting in an equilibrium density $H^*$ in railway locations. In this case, as long as the railways are functional, one should see a difference ($H^* - X^*$) in factor density between railway and non-railway locations. But after the railways become dysfunctional, $V_{fg}$ gradually falls back to $V_f$, indicating that there is no long-run difference in factor density between railway and non-railway locations, although the short-run difference is positive.

Our empirical evidence is more consistent with the scenario depicted in Figure 5a, as we observe neither a short-run nor a long-run difference in urbanization between railway and non-railway locations in the South of Nigeria.

Figure 6 depicts the situation in the North of Nigeria. Again, the initial situation is represented by $V_0$ when there is no transportation technology. As the North did not have a transportation alternative giving it access to the coast, there is no longer a curve represent-

[29]Bleakley and Lin (2012) explore other assumptions on the production technology that we do not explore in this paper as the predictions of the model under these alternatives assumptions are not supported by our data. To save space, we present the only mechanism that is empirically supported.

ing utility under a technology $f$. After the construction of the railroads, the indirect utility curve shifts from $V_0$ to $V_g$ in the railroad locations, and from $V_0$ to $V$ in the non-railroad locations, inducing equilibria $H_1^*$ and $L_1^*$, respectively. After the railways become obsolete, the vertical distance between $V$ and $V_g$ gradually falls, so that the two curves eventually coincide with $V_0$. The equilibrium in railway locations moves from $H_1^*$ to $H_2^*$ and the equilibrium in the non-railway location moves from $L_1^*$ to $L_2^*$. Therefore, there is a persistent difference in equilibrium factor density between the two locations.

This scenario depicts a situation in which railroads attracted the concentration of production factors in connected areas, inducing a spatial equilibrium that persisted in the long run even after railroads became dysfunctional. The initial sunk investment and natural advantages related to railroads helped to coordinate factor investments for each subsequent period, perpetuating a high-density equilibrium in railroad locations. Extensive empirical evidence on increased factor investments in railway areas in colonial Africa is presented in Jedwab and Moradi (2015), so we do not revisit the issue. Instead, we present empirical evidence supporting the claim that, consistent with our theoretical argument, initial market access, benefits and subsequent adoption rates of the railway differed between the North and the South.

## 7.2 Adoption Rates and Benefit of Rail by Key Regional Crops

> The Zungeru-Barijuko (Kaduna) to Baro would "traverse the greatest trade route in Nigeria, and render possible the export of cotton and other produce grown in the Nupe province and in Southern Zaria [Northern Nigeria]. **Without it cotton cannot,...be profitably exported from those districts**." — *p. 58 of Colonial Report, Northern Nigeria, 1902, as quoted in (Onyewuenyi, 1981, p.66)*.
>
> The increase in the export of this valuable product (hides) is most gratifying, and as communication with Northern Nigeria is facilitated, **it is expected to divert the greater proportion of this trade, which at present is said to go across to Tripoli**. — *p. 22 of Colonial Report, Southern Nigeria, Lagos, 1905 as quoted in (Onyewuenyi, 1981, p.66)*. [30]

The need for railroads for agricultural purposes was much greater in the North of Nigeria than in the South. The South had already been trading with Europe for centuries, through the slave trade and the palm oil trade which replaced it, while trade in most of Northern Nigeria was directed towards North Africa. The railway was crucial in diverting the trans-Saharan caravan trade to the Nigerian coast by lowering transportation costs. To see this, we compute the net benefit of shipping agricultural goods by railway relative to

[30] The emphasis highlighted in bold is ours.

other means such as waterways and roads, which were the two other modes of transportation in colonial Nigeria. The calculation is done over the period 1945-1949 for each of the key regional crops - groundnuts and cotton in the North and palm oil and cocoa in the South. The results are shown in Table 12. Data on prices, quantities, and distances come from World Bank (1955) and Onyewuenyi (1981). The cost of river shipments from the North is estimated as the cost of railing to Baro and then shipping by river to the Delta ports.[31] We use the railing distance as the shipping distance for rivers in the South, although this might be an overestimate given the proximity of the South to several rivers which lead to the coast.[32] While an overestimate of shipping distance through rivers, it helps to illustrate the fact that the railway could not compete with pre-existing means of transportation in the South even with implicit and explicit government subsidies.[33]

We estimate that cost reduction from shipping groundnuts and cotton by railway rather than by river to be 4% and 50%, respectively. The equivalent cost reduction from railing these goods instead of shipping them by road was 65% and 75%, respectively. These estimates are similar to the estimated reduction in Hodder (1959), who finds that the railway reduced the cost of shipping from the Jos mines by about 70% relative to road transportation. In comparison, however, railing palm oil and cocoa instead of shipping them by river would have *increased* their cost by 119% and 15%, respectively. Similarly, railing these crops versus shipping them by road would have *increased* their cost by 58% and 60%, respectively.

As regards to shipping goods to the coast, railroads were cheaper than alternative transportation modes in the North, whereas the latter were cheaper in the South. This rationalizes the high adoption rate of railroads in the North and its low adoption rate in the South. Figure 7 presents percentages of Northern and Southern goods shipped through the railway over the period 1931-1949.[34] On average, 96% of groundnuts and 81% of cotton were railed to the coast from the North, compared to only 18% of palm kernel, 31% of palm oil, and 26% of cocoa, the three main Southern crops. We also observe that the fraction of goods shipped to the coast through rail increased over time for Northern goods, while it declined for Southern goods as subsidies for the railways declined (except palm oil).[35]

Consistent with this, the volume of groundnut exports (the main export in the North) grew at an annual rate of 13.8%, from 475 metric tons in 1900-1904 to 268,409 metric

[31] Hence, the rail prices in parentheses in Table 12.

[32] Using average straight line distances from survey clusters in the DHS to various transportation nodes, we calculate that Southern populated areas are four times closer to rivers than Northern populated areas (23 km vs. 90 km). This discrepancy is worsened if we take into account the navigability of rivers which is much better in the South.

[33] See discussion in Onyewuenyi 1981, p. 89–93.

[34] Figure 7 plots the ratio of quantities of goods railed to the quantities of goods exported. The outlier above 100% is due to crop spoliage.

[35] The cocoa example is striking since the percentage shipped through rail shrank from a third of the production to virtually zero over the period 1931-1950.

tons in 1945-1949. Over the 1900-1949 period, during which the railway expanded, exports of palm produce (the main export in the South) grew at an average annual rate of 1.9% only. A non-trivial volume of palm produce was already being exported in the early twentieth century, which illustrates initial access to export markets.

The evidence presented here illustrates two important facts. First, producers in the South had viable transportation alternatives to the railway, and the railway did not substantially lower transportation costs in the region (it actually increased transportation costs). Second, and as a result of the first observation, adoption rates of the railway were substantially lower in the South. In the North, however, the railway substantially lowered transportation costs and increased market access as producers adopted it as the primary means of exporting commodities.

### 7.3 Key Factor: Distance to Ports of Export

If the highlighted heterogeneity of the impact of the railway has to do with availability of other transportation alternatives in the South giving it access to ports of export rather than other fundamental differences between Northern and Southern areas, one would expect that the railway impact increases with distance to ports of export and that the effect is not driven by the presence of early cities which were more prevalent in the South of the country. We explore this by conducting two exercises.

First, we measure the impact of the railway by distance to ports of export. For each individual, we compute the distance of her survey cluster to the closest port. We then split the sample into two subsamples: individuals with above-median distance to ports and individuals with below-median distance to ports. We estimate the impact of railroads on the outcome variables using both the fixed effects (OLS) and IV (2SLS) strategies for each subsample, separately. The results are presented in Table 13.[36]

The effect of the railway is generally larger in local areas that were further away from ports of export. The effect of the railway for individuals in local areas that had higher pre-railway European market access is economically small and mostly not statistically different from zero. Importantly, state fixed effects and ethnicity fixed effects are included in the regressions presented in Table 13. Hence, these findings are not driven by specific characteristics of a given state or by ancestral exposure to railways.

In the bottom panel of Table 13, we interact our main independent variable with ranges of distances to coastal ports: 0-200 $km$, 200-400 $km$, 400-600 $km$ and over 600 $km$. The local effects of railways are unequivocally stronger, the further individuals are from ports. These findings provide a consistent explanation of the non-impact of the railway

[36]We also divided the sample by mean distance to ports, and into various groups, within 100 $km$, 100-300 $km$, and other distances that included a balanced mix of districts in the South and the North. We consistently find that the impacts of the railway begin to emerge beyond 300 $km$ from the coast. These results are omitted for brevity but are available upon request.

in the South based on its geographical proximity to ports and to its existing waterways connecting the different ports and on the non-substantial change in its market access as a consequence of the introduction of railways.

Second, we measure the impact of the railway by proximity to early cities. If the heterogeneity of the effect of the railway that we uncover is driven by the fact that the South had more early cities, one would expect that once we interact proximity to the railway with proximity to early cities, the effect on areas close to both the railway and early cities would be null. Table 14 shows that this is not the case. For most development outcomes, the effect on areas close to both the railway and early cities is significantly positive. This suggests that the heterogeneity we highlight can not be explained away by the stronger presence of early cities in Southern parts of Nigeria, or by the fact that Southern Nigeria was already urbanized. This finding, along with the other pieces of evidence shown above, indicates the importance of opportunity costs in the transformative power of railroads. Unlike in the South, railroads were vital in the North of Nigeria to enable exporting trade with Europe and as a result, they had a tremendous effect in this region.

# 8 Concluding Remarks

In this paper, we test Fogel's influential hypothesis that new transportation technologies may be dispensable if pre-existing technologies are viable or can simply be improved. We contribute to the existing literature by using "real" (as opposed to "artificial") counterfactuals to shed light on this hypothesis. We show that colonial railroads did not have a homogeneous impact in all the areas that they connected. Consistent with the Fogelian view, we find that the railroads had very little economic impact in the South of Nigeria which, thanks to waterways, already had viable alternatives for the transportation of goods to exporting ports. The North, however, was lacking transportation technologies to export goods. The railways were essential in linking these Northern areas to the exporting coastal ports. This revolution transformed these areas, not only in the short run but also over time and until today.

We find that, in the North, individuals in areas that were connected to the railways are more likely to go to school, to be literate, to have media access, to work in higher-paying professions and to live in wealthier households than individuals in unconnected areas. Connected areas are also more urbanized than unconnected areas. We do not find any of these effects in the South, neither in the short run nor in the long run.

The key variable underlying our results is the level of market access that the various areas of interest enjoyed before the construction of railroads. Areas that had more market access had viable transportation technologies linking them to exporting ports and did not desperately "need" railroads. Consequently, these areas were not transformed by railroads.

The converse is true for areas that were connected to exporting ports thanks to the railway.

Our findings also indicate a strong path dependence in the positive effect of railways in the North of Nigeria and in their non-effect in the South. They are consistent with the theoretical argument that, in the North, railroads led to the concentration of production factors in connected areas and these initially advantaged areas helped to coordinate factor investments in each of the subsequent periods. This implied a persistence of the initial spatial equilibrium induced by the railroads, even after they became obsolete. In the South, the railway did not even change the existing spatial equilibrium in the short run. It is perhaps not surprising that its non-impact in this region also persisted over time.

# References


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

**Ajayi, J.F.A., L.K.H. Goma, and A.G. Johnson**, *The African Experience with Higher Education*, Association of African Universities, 1996.

**Anene, J.C.**, *Southern Nigeria in Transition, 1885-1906: Theory and Practice in a Colonial Protectorate*, Cambridge University Press, 1966.

**Ayandele, E.A.**, *The Missionary Impact on Modern Nigeria, 1842-1914: A Political and Social Analysis* Ibadan History Series, Longmans, 1966.

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

**Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang**, "Roads, Railroads and Decentralization of Chinese cities," Technical Report, Working paper 2012.

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

**Cage, Julia and Valeria Rueda**, "The Long-Term Effects of the Printing Press in Sub-Saharan Africa," Sciences Po Economics Discussion Papers 2014-15, Sciences Po Departement of Economics Oct 2014.

**Chaléard, Jean-Louis, Chantal Chanson-Jabeur, and Chantal Béranger**, *Le Chemin de Fer en Afrique*, KARTHALA Editions, 2006.

**Conley, T.G.**, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

**Crowder, M.**, *The Story of Nigeria* Faber paperbacks, Faber & Faber, 1980.

**Csapo, Marg**, "Religious, Social and Economic Factors Hindering the Education of Girls in Northern Nigeria," *Comparative Education*, 1981, *17* (3), 311–319.

**Dev, Pritha, Blessing U. Mberu, and Roland Pongou**, "Ethnic Inequality: Theory and Evidence from Formal Education in Nigeria," *Economic Development and Cultural Change*, 2016, *64* (4), 603–660.

**DMA, Defense Mapping Agency**, *Digital Chart of the World*, Fairfax, Virginia: Defense Mapping Agency, 1992. Available at `http://www.diva-gis.org/gdata`.

**Donaldson, Dave**, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2016, *Forthcoming.*

— **and Richard Hornbeck**, "Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, 2016.

**Faber, Benjamin**, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

**Falola, T. and M.M. Heaton**, *A History of Nigeria*, Cambridge University Press, 2008.

**FAO**, "Nigeria At a Glance," 2016. Available at http://www.fao.org/nigeria/fao-in-nigeria/nigeria-at-a-glance/en/, Accessed: February, 2016.

**Fay, Marianne and Charlotte Opal**, *Urbanization Without Growth: A Not So Uncommon Phenomenon* Policy Research Working Papers, World Bank, Policy Research Dissemination Center, 2000.

**Fenske, James and Namrata Kala**, "1807: Economic Shocks, Conflict and the Slave Trade," CSAE Working Paper Series 2014-02, Centre for the Study of African Economies, University of Oxford 2014.

**Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, and D. Wiberg**, *Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008)*, IIASA, Laxenburg, Austria and FAO, Rome, Italy, 2008.

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

**Fourie, Johan and Alfonso Herranz-Loncan**, "Growth (and Segregation) by Rail : How the Railways Shaped Colonial South Africa," 2015.

**Gallego, Francisco A. and Robert Woodberry**, "Christian Missionaries and Education in Former African Colonies: How Competition Mattered," *Journal of African Economies*, 2010, *19* (3), 294–329.

**Hijmans, R. J., M. Cruz, E. Rojas, and L. Guarino**, "DIVA-GIS, version 1.4. A Geographic Information System for the Management and Analysis of Genetic Resources Data.," http://www.diva-gis.org/gdata 2001.

**Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis**, "Very High Resolution Interpolated Climate Surfaces for Global Land Areas," *International Journal of Climatology*, 2005, *25* (15), 1965–1978.

**Hodder, B. W.**, "Tin Mining on the Jos Plateau of Nigeria," *Economic Geography*, 1959, *35* (2), 109–122.

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

**Jaekel, Francis**, *The History of the Nigerian Railway: Opening the Nation to Sea, Air and Road Transportation*, Vol. 1, Spectrum Books, 1997.

**Jedwab, Remi and Alexander Moradi**, "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," CEH Discussion Papers 031, Centre for Economic History, Research School of Economics, Australian National University Jan 2015.

— **and Dietrich Vollrath**, "Urbanization Without Growth in Historical Perspective," *Explorations in Economic History*, 2015, *58*, 1–21.

— **, Edward Kerby, and Alexander Moradi**, "History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya," *The Economic Journal*, 2015, p. Forthcoming.

**Lamb, P. H.**, "Past, Present and Future of Cotton-Growing in Nigeria," *Empire Cotton-Growing Review*, 1925, pp. 18–23.

**Law, Robin**, *From Slave Trade to 'Legitimate' Commerce: The Commercial Transition in Nineteenth-Century West Africa* African Studies, Cambridge, U.K.: Cambridge University Press, 2002.

**Michalopoulos, Stelios and Elias Papaioannou**, "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, 2014, *129* (1), 151–213.

**NBS and ICF International**, *2008 Nigeria Demographic and Health Survey [Data File]*, Vol. NGIR52FL.DTA and NGMR52FL.DTA, Calverton, Maryland: ICF International: Nigeria Bureau of Statistics and ICF International, 2008. Available at `http://dhsprogram.com/data/dataset/Nigeria_Standard-DHS_2008.cfm?flag=1`.

**Nunn, Nathan**, "The Long-Term Effects of Africa's Slave Trades," *The Quarterly Journal of Economics*, 02 2008, *123* (1), 139–176.

__ , "Gender and Missionary Influence in Colonial Africa," in Emmanuel Akyeampong, Robert H. Bates, Nathan Nunn, and James Robinson, eds., *African Development in Historical Perspective*, Cambridge: Cambridge University Press, May 2014.

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

**Okoye, Dozie and Roland Pongou**, "Historical Missionary Activity, Schooling, and the Reversal of Fortunes: Evidence from Nigeria," MPRA Paper 58052, University Library of Munich, Germany Aug 2014.

__ **and** __ , "Sea Change: The Competing Long-Run Impacts of the Transatlantic Slave Trade and Missionary Activity in Africa," Working Paper 2015.

**Onyewuenyi, Remy N**, *Railway Development and the Growth of Export Agriculture in Nigeria during the 1900-1950 Period.*, University of Ottawa (Canada)., 1981.

**Rappaport, Jordan**, "Moving to Nice Weather," *Regional Science and Urban Economics*, 2007, *37* (3), 375–398.

**Roome, William R.M.**, *Ethnographic Survey of Africa Showing the Tribes and Languages: Also the Stations of Missionary Societies*, Edward Stanford Ltd, 1925.

**Rothstein, Morton**, "Railroads and American Economic Growth: Essays in Econometric History. By Robert William Fogel. Baltimore, The Johns Hopkins Press, 1964. Pp. 296.," *Business History Review*, 3 1965, *39*, 130–132.

**Schätzl, Ludwig**, *Industrialization in Nigeria: A Spatial Analysis*, Vol. 81, Weltforum Verlag München, 1973.

**Storeygard, Adam**, "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

**Taaffe, Edward J, Richard L Morrill, and Peter R Gould**, "Transport Expansion in Underdeveloped Countries: A Comparative Analysis," *Geographical review*, 1963, *53* (4), 503–529.

**Taylor, George R.**, "Review of Railroads and American Economic Growth," *American Economic Review, IV*, 1965, pp. 890–2.

**UC Berkeley, Museum of Vertebrate Zoology**, *GADM Database of Global Administrative Areas, Version 2.7*, University of California, Berkeley, 2014. Available at `www.gadm.org`.

**Wantchekon, Leonard, Marko Klašnja, and Natalija Novta**, "Education and Human Capital Externalities: Evidence from Colonial Benin," *The Quarterly Journal of Economics*, 2015, *forthcoming.*

**Whatley, Warren and Rob Gillezeau**, "The Impact of the Transatlantic Slave Trade on Ethnic Stratification in Africa," *American Economic Review*, 2011, *101* (3), 571–76.

**World Bank**, "The Economic Development of Nigeria," Technical Report 11151, The World Bank, Baltimore, MD 1955.

# Figures and Tables

**Figure 1:** Rail Line Across Clusters and Local Areas

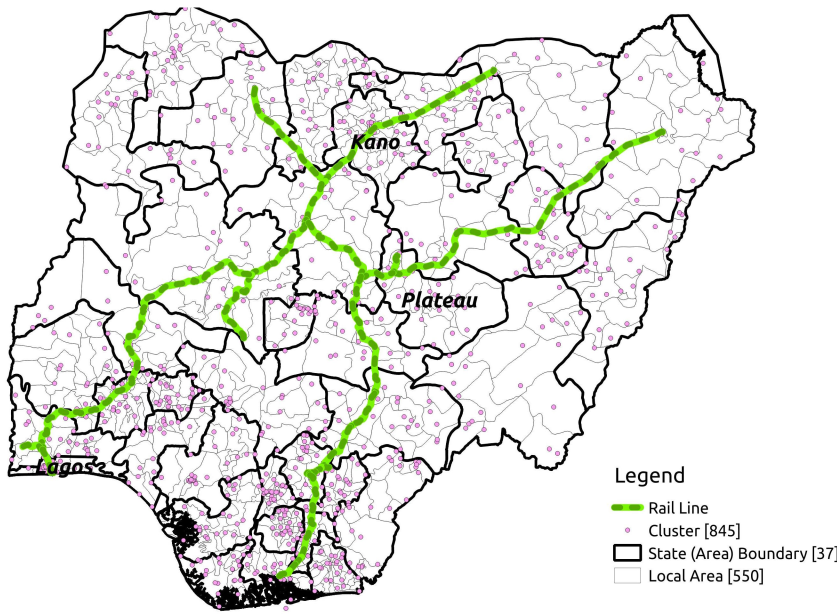



**Figure 2:** Straight Lines Between Nodes

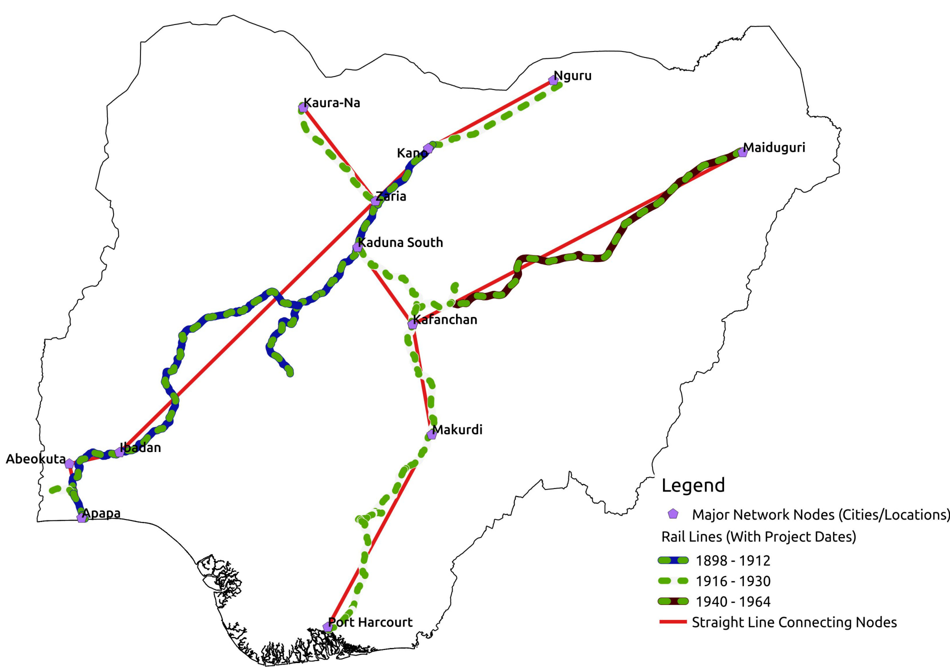


**Figure 3:** Rail Line, Ports and Placebo Lines

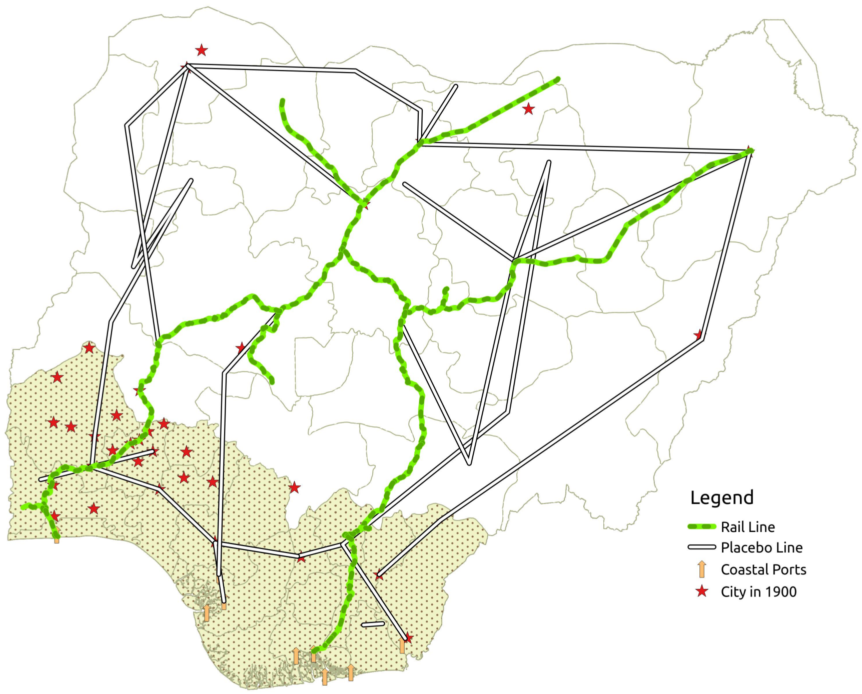



**Figure 4:** Rail Line, Roads and Placebo Lines

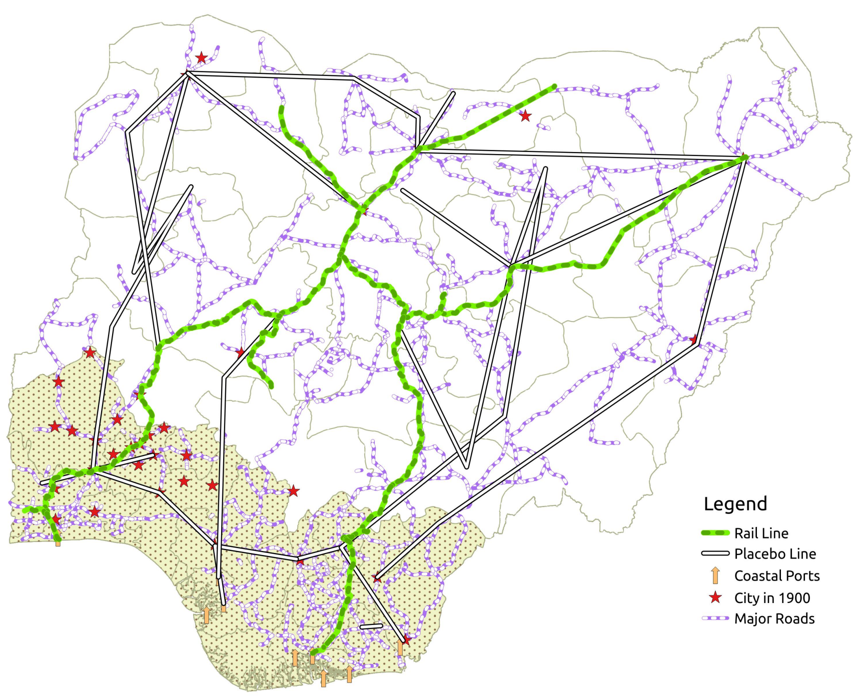


**Figure 5:** Equilibrium in the South of Nigeria

**(a)** $V_{fg} = V_f$ **(b)** $V_{fg} > V_f$

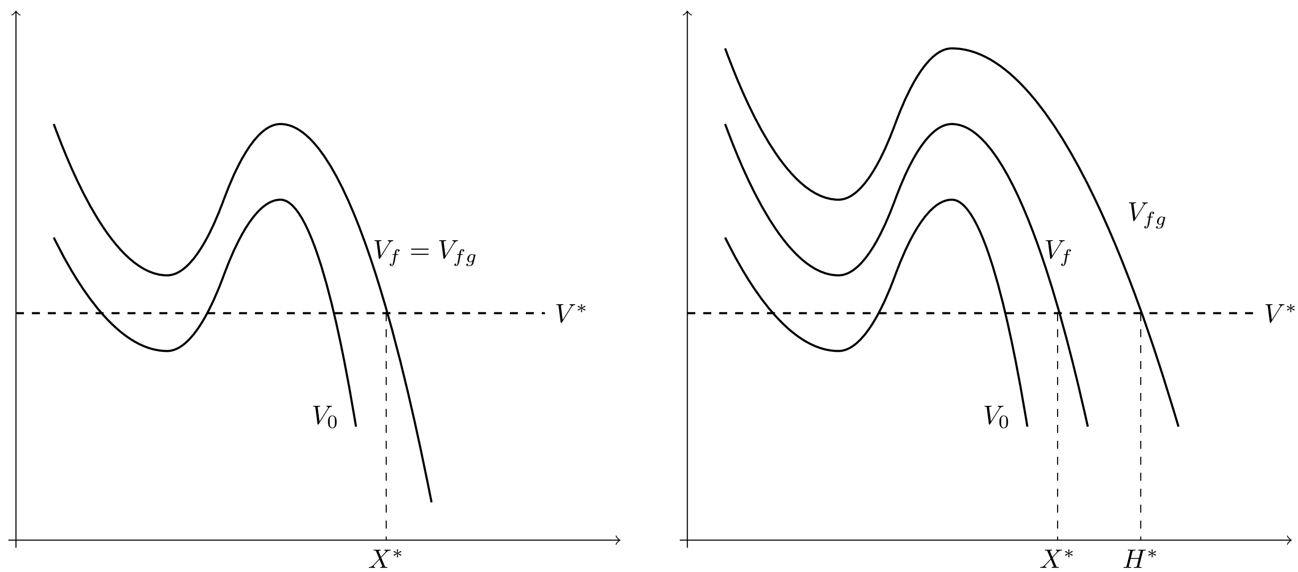



**Figure 6:** Equilibrium in the North of Nigeria

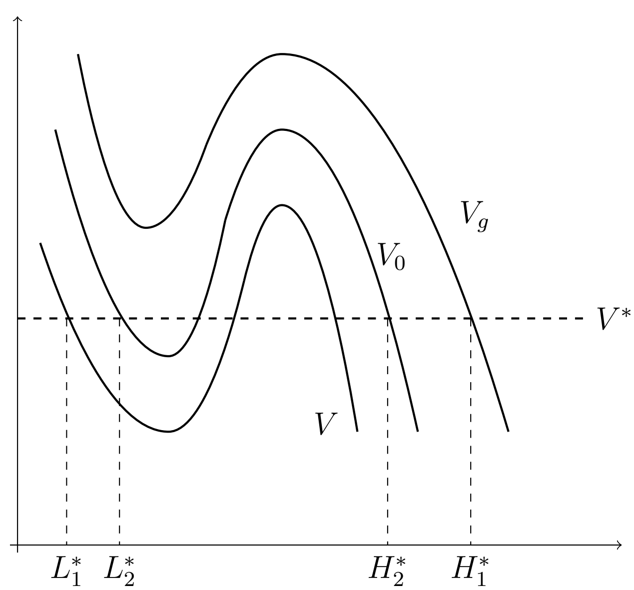
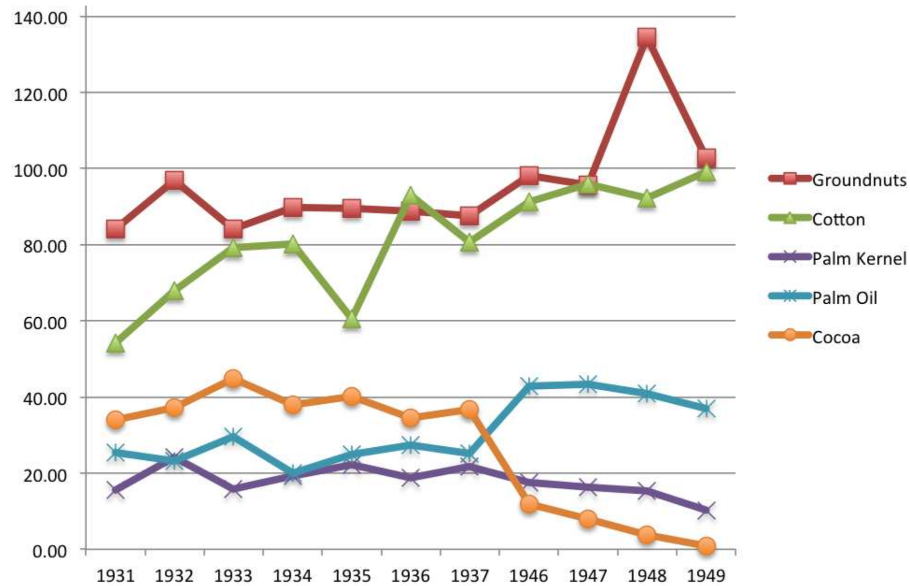


**Figure 7:** Railway Adoption for Northern and Southern Main Exports




**Notes:** Figure 7 plots the ratio of quantities of goods railed to the quantities of goods exported. The outlier above 100% is due to crop spoliage. During the 1931-1949 period, groundnuts and cotton were the main Northern exports and palm Kernel, palm oil and cocoa were the main Southern exports.

**Table 1:** History of Railway Construction in Nigeria

| Link | Date | Length(km) | Motivation |
|---|---|---|---|
| Lagos - Otta | 1898 | 32 | Administrative & Agricultural |
| Otta - Abeokuta - Ibadan | 1901 | 165 | Administrative & Agricultural |
| Ibadan - Ilorin | 1908 | 201 | Administrative & Agricultural |
| Ilorin - Jebba | 1909 | 96 | Administrative & Agricultural |
| Zaria - Jos - Bukuru | 1911 | 227 | Mineral |
| Jebba - Zungeru - Minna | 1912 | 233 | Administrative & Agricultural |
| Baro - Kano | 1912 | 573 | Administrative & Agricultural |
| Port Harcourt - Enugu | 1916 | 243 | Agricultural & Mineral |
| Enugu - Makurdi - Jos | 1927 | 596 | Agricultural & Mineral |
| Kaduna - Kafanchan | 1927 | 201 | Administrative & Mineral |
| Zaria - Gusau - Kaura Namoda | 1929 | 232 | Agricultural |
| Kano - Nguru | 1930 | 229 | Agricultural |
| Ifo - Ilaro - Idogo | 1930 | 39 | Agricultural |
| Jos - Maiduguri | 1964 | 645 | Agricultural |

**Notes:** The motivations for the railway construction are classified in three categories : administrative (political or military), agricultural exploitation, and mineral exploitation.
Source: Onyewuenyi 1981, p. 39.

**Table 2:** Observables in Areas Within and Outside 20 $km$ of Railway Tracks

| Variable | Within 20 $km$ of Rail | | | Outside 20 $km$ of Rail | | | Difference in Means | |
|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Median | Mean | S.D. | Median | Difference | T-stat |
| **Local Area Variables** | | | | | | | | |
| **Annual Rainfall (mm)** | 1471.09 | 587.97 | 1393 | 1502.82 | 713.16 | 1363 | -31.73 | (-0.40) |
| **Temperature (celsius)** | 26.36 | 0.62 | 26.4 | 26.38 | 0.84 | 26.5 | -0.02 | (-0.22) |
| **Soil Nutrients** | 1.53 | 0.69 | 1 | 1.38 | 0.64 | 1 | 0.15 | (1.50) |
| **Soil Workability** | 1.99 | 0.86 | 2 | 1.75 | 0.77 | 2 | .24** | (2.23) |
| **Elevation (Meters)** | 217.46 | 214.58 | 138 | 251.05 | 191.52 | 225 | -33.59 | (-1.21) |
| **Oil Palm Suitability (kg/ha)** | 2.86 | 2.05 | 4 | 2.44 | 2.08 | 4 | 0.42 | (1.78) |
| **Cocoa Suitability (kg/ha)** | 0.67 | 0.47 | 1 | 0.59 | 0.49 | 1 | 0.08 | (1.38) |
| **Cotton Suitability (kg/ha)** | 0.35 | 0.48 | 0 | 0.43 | 0.49 | 0 | -0.08 | (-1.31) |
| **Groundnut Suitability (kg/ha)** | 1.82 | 0.39 | 2 | 1.8 | 0.4 | 2 | 0.02 | (0.38) |
| **Mission Station** | 0.25 | 0.43 | 0 | 0.21 | 0.4 | 0 | 0.04 | (0.70) |
| **Primary Road in Area** | 0.37 | 0.48 | 0 | 0.44 | 0.5 | 0 | -0.06 | (-1.06) |
| **Major River in Area** | 0.37 | 0.48 | 0 | 0.33 | 0.47 | 0 | 0.05 | (0.88) |
| **City in 1890** | 0.075 | 0.26 | 0 | 0.072 | 0.26 | 0 | -0.003 | (-0.1) |
| **1900 City Population (Standardized)**[1] | .21 | 1.83 | -.2 | -.027 | .74 | -.2 | .234 | (1.19) |
| **Local Areas with Cities in 1900** | 12 | | | 27 | | | | |
| **Household Variables** | | | | | | | | |
| **Male Head** | 0.84 | 0.36 | 1 | 0.84 | 0.36 | 1 | 0 | (0.03) |
| **Age of Head** | 44.2 | 13.44 | 43 | 44.14 | 14.26 | 43 | 0.05 | (0.13) |
| **Household Size** | 5.83 | 3.8 | 5 | 5.91 | 3.5 | 5 | -0.09 | (-0.47) |
| **Individual Variables** | | | | | | | | |
| **Age** | 29.65 | 10.16 | 28 | 29.72 | 10.43 | 28 | -0.07 | (-0.41) |
| **Christian** | 0.52 | 0.5 | 1 | 0.54 | 0.5 | 1 | -0.02 | (-0.44) |
| **Observations** | 10386 | | | 28980 | | | 39366 | |
| **Number of Clusters** | 208 | | | 637 | | | 845 | |
| **Number of Local Areas** | 134 | | | 436 | | | 550 | |

**Notes:** *** $p < .01$, ** $p < .05$. Summary statistics of local government areas for clusters within and outside 20 $km$ of a rail track. T-statistics of differences in means are obtained from standard errors clustered at the local area level. For categorical variables, $1 = Yes$ and $0 = No$, so the mean represents the proportion in each area.
[1]This row compares the standardized scores of populations in 1900 cities in local government areas with and without the rail line.

**Table 3:** Effect of Proximity to Railway on Contemporary Outcomes (Fixed Effects)

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Rail Within 20 $km$** | 1.286*** | 0.128*** | 0.0161*** | -0.0737*** | 0.0720*** | 0.0610*** | 0.157*** | 0.637*** | 0.183*** |
| | [5.12] | [5.80] | [2.83] | [-4.55] | [4.79] | [4.55] | [5.73] | [6.33] | [3.62] |
| **Rainfall (mm)** | -0.0004 | 0 | 0 | 0 | 0 | 0 | 0 | -0.0001 | 0.0001 |
| | [-0.99] | [0.02] | [-0.83] | [-0.07] | [-0.59] | [-0.14] | [-0.23] | [-0.52] | [0.64] |
| **Temperature (celsius)** | -0.637** | -0.0235 | -0.0047 | 0.0567** | -0.0315 | -0.0028 | -0.0302 | -0.228* | -0.0009 |
| | [-2.20] | [-0.91] | [-0.62] | [2.57] | [-1.63] | [-0.12] | [-0.92] | [-1.76] | [-0.01] |
| **Soil Nutrients** | -0.279* | -0.0218* | -0.0024 | 0.013 | -0.0109 | -0.0031 | -0.0185 | -0.115* | -0.0637** |
| | [-1.92] | [-1.86] | [-0.67] | [1.14] | [-1.03] | [-0.32] | [-1.23] | [-1.85] | 0 |
| **Soil Workability** | 0.299** | 0.0267** | 0.0026 | -0.0076 | 0.0246*** | 0.0099 | 0.0207 | 0.0396 | [.] |
| | [2.26] | [2.37] | [0.78] | [-0.75] | [2.61] | [1.12] | [1.53] | [0.72] | |
| **Elevation** | 0 | 0.0002 | 0 | 0.0002 | 0 | 0.000305** | 0.0002 | 0 | 0.0003 |
| | [-0.01] | [1.06] | [1.11] | [1.24] | [0.43] | [2.42] | [0.94] | [-0.02] | [0.81] |
| **Oil Palm Suitability** | 0.317 | 0.0293* | 0.0057 | -0.0408* | 0.0305** | 0.0271** | 0.0402** | 0.130* | -0.016 |
| | [1.44] | [1.76] | [1.03] | [-1.91] | [2.18] | [2.50] | [2.31] | [1.76] | [-0.45] |
| **Cocoa Suitability** | -1.015 | -0.0877 | -0.0197 | 0.113* | -0.0884 | -0.0562 | -0.135** | -0.440* | -0.107 |
| | [-1.32] | [-1.43] | [-0.84] | [1.73] | [-1.59] | [-1.33] | [-2.23] | [-1.78] | [-0.73] |
| **Cotton Suitability** | -0.777 | -0.106** | -0.0089 | 0.0801* | -0.0054 | -0.0368 | -0.038 | -0.266 | -0.111 |
| | [-1.50] | [-2.35] | [-0.54] | [1.75] | [-0.16] | [-1.15] | [-0.72] | [-1.56] | [-0.92] |
| **Groundnut Suitability** | -0.4 | -0.0334 | -0.0003 | 0.0565* | -0.0538 | -0.0228 | -0.0771** | -0.252 | -0.225** |
| | [-1.34] | [-1.22] | [-0.03] | [1.74] | [-1.57] | [-0.96] | [-2.47] | [-1.63] | [-2.55] |
| **Missions in Local Area** | 1.092*** | 0.0769*** | 0.0226*** | -0.0949*** | 0.0852*** | 0.0389*** | 0.117*** | 0.467*** | 0.217*** |
| | [5.81] | [5.11] | [3.77] | [-5.62] | [5.43] | [3.41] | [5.82] | [5.74] | [4.92] |
| **Age** | 0.185*** | -0.00638*** | 0.00741*** | 0.00767*** | 0.00882*** | 0.00539*** | 0.0017 | 0.00936*** | 0.00386*** |
| | [12.18] | [-4.32] | [9.55] | [4.96] | [6.36] | [4.17] | [1.28] | [2.91] | [2.94] |
| **Christian** | 2.027*** | 0.147*** | 0.0596*** | -0.0411*** | 0.117*** | 0.0558*** | 0.0957*** | 0.274*** | 0.0196 |
| | [12.07] | [10.00] | [9.26] | [-2.59] | [8.26] | [5.28] | [5.21] | [4.80] | [0.73] |
| **Male Head** | -0.537*** | -0.0464*** | -0.00787* | 0.0041 | -0.0313*** | 0.0237*** | -0.0039 | 0.045 | -0.0265** |
| | [-6.39] | [-6.30] | [-1.95] | [0.62] | [-3.80] | [3.18] | [-0.44] | [1.52] | [-2.24] |
| **Age of Head** | 0.00927*** | 0.000828*** | -0.000329*** | -0.000781*** | 0.000818*** | -0.0002 | 0.0004 | -0.00475*** | -0.0002 |
| | [3.86] | [4.19] | [-3.83] | [-4.23] | [3.52] | [-0.96] | [1.46] | [-6.15] | [-0.45] |
| **Household Size** | 0.0184 | 0.00388*** | -0.000968** | -0.00181* | 0.001 | 0.00219** | 0.00716*** | 0.0358*** | 0.00415** |
| | [1.22] | [3.24] | [-2.20] | [-1.68] | [0.91] | [2.07] | [5.67] | [7.91] | [2.06] |
| **Ethnicity Fixed Effects (31)** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effect (37)** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.499 | 0.438 | 0.056 | 0.257 | 0.256 | 0.189 | 0.352 | 0.498 | 0.291 |
| **Control Means** | 6.186 | 0.566 | 0.046 | 0.228 | 0.277 | 0.752 | 0.521 | 2.932 | 0.282 |

**Notes:** $*p < .1$, $** \ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. Table shows estimates of the impact of being within 20 $km$ of a railway line on various individual outcomes. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the local government area (county). Data on Christian mission stations come from historical maps, as described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

## **Table 4:** First-Stage Estimates based on Distance to Line Joining Nodes

| Dependent Variable is an Indicator for Being Within 20 $km$ of Railway Line | | | | | |
|---|---|---|---|---|---|
| **Distance to Line** | **Rainfall (mm)** | **Temperature (celsius)** | **Soil Nutrients** | **Soil Workability** | **Altitude** |
| -0.294*** | 0.0000331 | 0.000363 | -0.0261 | 0.0178 | -0.000441* |
| [-14.12] | [0.50] | [0.02] | [-0.65] | [0.95] | [-1.74] |
| **Oil Palm Suitability** | **Cocoa Suitability** | **Cotton Suitability** | **Groundnut Suitability** | **Missions in Local Area** | **Age** |
| 0.00193 | 0.106 | 0.116 | 0.0153 | 0.0183 | 0.000383 |
| [0.09] | [0.95] | [1.47] | [0.54] | [0.72] | [0.39] |
| **Christian** | **Male Head** | **Age of Head** | **Household Size** | | |
| -0.0149 | -0.00816 | 0.000122 | 0.00094 | | |
| [-0.67] | [-1.35] | [0.68] | [1.03] | | |
| **Ethnicity Fixed Effect** | **State Fixed Effect** | **Observations** | **Adjusted $R^2$** | **AP F-stat** | |
| Yes | Yes | 37262 | 0.679 | 199.5 | |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The instrument (Distance to Line) is a measure of the distance of the individual's cluster to a straight line joining selected nodes of the railway line as obtained from Table 4 of Onyewuenyi (1981), but selecting nodes based on historical importance. All nodes of the railway line, regardless of historical importance, identified as all clusters within the local government area with the rail station, are dropped from the regressions as they may have been endogenously chosen. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network . Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Surveys.

## **Table 5:** Effect of Proximity to Railway on Contemporary Outcomes (2SLS)

| | **Schooling** | **Literacy** | **Professional** | **Agricultural** | **Read Paper** | **Listen to Radio** | **Watch TV** | **Wealth** | **Urban Resident** |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Rail Within 20 $km$** | 1.391*** | 0.128*** | 0.0129 | -0.0509** | 0.0758*** | 0.0631*** | 0.173*** | 0.635*** | 0.114 |
| | [3.66] | [3.95] | [1.57] | [-2.15] | [3.05] | [2.86] | [4.24] | [4.06] | [1.48] |
| **Ethnicity Fixed Effect** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effect** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Observations** | 37228 | 36928 | 37089 | 37089 | 37068 | 37168 | 37159 | 37262 | 37262 |
| **Centered $R^2$** | 0.502 | 0.44 | 0.057 | 0.26 | 0.261 | 0.187 | 0.354 | 0.493 | 0.294 |
| **Control Means** | 6.186 | 0.566 | 0.046 | 0.228 | 0.277 | 0.752 | 0.521 | 2.932 | 0.282 |
| **KP UnderId Test** | 86.03 | 86.06 | 86.21 | 86.21 | 86.07 | 86.18 | 86.24 | 86.09 | 86.09 |
| **KP rk Wald F** | 199.2 | 198.1 | 200 | 200 | 197.6 | 201.7 | 199.8 | 199.5 | 199.5 |
| **AR Wald F** | 12.65 | 14.12 | 2.495 | 4.568 | 8.621 | 7.712 | 16.63 | 15.14 | 2.089 |
| **SW LM S** | 11.56 | 12.95 | 2.461 | 4.367 | 8.379 | 7.051 | 14.49 | 13.43 | 2.051 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The instrument (Distance to Line) is a measure of the distance of the individual's cluster to a straight line joining nodes of the railway line selected based on historical importance. All nodes regardless of historical importance, identified as all clusters within the local government area with the rail station, are dropped from the regressions as they may have been endogenously chosen. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 6:** Falsification Exercises: Other Transportation Means and Placebo Lines

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | **Other Modes of Transportation** | | | | | |
| **Rail Within 20 $km$** | 0.965*** | 0.101*** | 0.0111* | -0.0519*** | 0.0503*** | 0.0382*** | 0.120*** | 0.454*** | 0.126*** |
| | [4.18] | [5.07] | [1.92] | [-3.24] | [3.52] | [2.99] | [4.76] | [5.04] | [2.59] |
| **River in Local Area** | -0.0014 | -0.0003 | -0.0001 | 0.0001 | -0.000314* | -0.0002 | -0.000942*** | -0.00191** | -0.0004 |
| | [-0.60] | [-1.23] | [-1.51] | [0.60] | [-1.92] | [-1.48] | [-3.83] | [-2.04] | [-0.75] |
| **Distance to Road** | -0.0417*** | -0.00339*** | -0.000596*** | 0.00281*** | -0.00266*** | -0.00290*** | -0.00450*** | -0.0229*** | -0.00718*** |
| | [-6.60] | [-5.60] | [-4.01] | [5.49] | [-5.77] | [-6.18] | [-6.14] | [-8.18] | [-4.82] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38768 | 38869 | 38866 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.506 | 0.444 | 0.057 | 0.263 | 0.266 | 0.195 | 0.367 | 0.53 | 0.318 |
| | | | | **Placebo Lines** | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Within 20 $km$ of Placebo** | 0.285 | 0.0224 | 0.0087 | -0.0265* | 0.0236 | 0.0367*** | 0.0409** | 0.238*** | 0.0609 |
| | [1.62] | [1.53] | [1.56] | [-1.72] | [1.63] | [2.93] | [2.01] | [2.83] | [1.47] |
| **Observations** | 28597 | 28389 | 28502 | 28502 | 28479 | 28540 | 28532 | 28621 | 28621 |
| **Adjusted $R^2$** | 0.526 | 0.45 | 0.055 | 0.255 | 0.244 | 0.19 | 0.358 | 0.475 | 0.253 |
| | | | | **Placebo Lines and Other Modes of Transportation** | | | | | |
| **Within 20 $km$ of Placebo** | 0.115 | 0.0098 | 0.0061 | -0.013 | 0.0115 | 0.0223* | 0.0244 | 0.142* | 0.0235 |
| | [0.66] | [0.68] | [1.10] | [-0.87] | [0.81] | [1.90] | [1.23] | [1.79] | [0.59] |
| **River in Local Area** | -0.359* | -0.026 | -0.0052 | 0.0241 | -0.0249 | -0.0288** | 0.0003 | -0.0843 | -0.120** |
| | [-1.72] | [-1.54] | [-0.94] | [1.21] | [-1.45] | [-2.13] | [0.01] | [-1.03] | [-2.41] |
| **Distance to Road** | -0.0277*** | -0.00206*** | -0.000424*** | 0.00226*** | -0.00198*** | -0.00236*** | -0.00327*** | -0.0176*** | -0.00538*** |
| | [-4.62] | [-3.67] | [-2.74] | [4.20] | [-4.05] | [-4.59] | [-4.57] | [-6.28] | [-3.42] |
| **Observations** | 28597 | 28389 | 28502 | 28502 | 28479 | 28540 | 28532 | 28621 | 28621 |
| **Adjusted $R^2$** | 0.53 | 0.452 | 0.056 | 0.259 | 0.247 | 0.194 | 0.364 | 0.497 | 0.282 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates the impact of being within 20 $km$ to a placebo line joining proposed nodes. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distances to rail network and roads are computed using DHS data and information on rail and road networks. Presence of river is computed using information on presence of river within the local government area. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 7:** Effect of Railway By Different Control Groups, Reasons and Dates of Construction

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | **Across Multiple Distances** | | | | | |
| **Rail Within 20** $km$ | 1.158*** | 0.117*** | 0.0207*** | -0.0668*** | 0.0818*** | 0.0697*** | 0.173*** | 0.680*** | 0.185*** |
| | [3.80] | [4.32] | [2.76] | [-3.16] | [4.18] | [3.66] | [5.19] | [5.36] | [2.97] |
| **Rail Within 20-40** $km$ | -0.343 | -0.0288 | 0.0036 | 0.0137 | 0.0106 | 0.004 | 0.0058 | 0.0172 | -0.0666 |
| | [-1.42] | [-1.40] | [0.51] | [0.67] | [0.59] | [0.21] | [0.22] | [0.16] | [-1.26] |
| **Rail Within 40-60** $km$ | 0.0216 | -0.001 | 0.0064 | -0.0109 | 0.0099 | 0.0142 | 0.0508* | 0.125 | 0.065 |
| | [0.09] | [-0.05] | [0.91] | [-0.51] | [0.54] | [0.83] | [1.79] | [1.17] | [1.15] |
| **Rail Within 60-80** $km$ | 0.0242 | 0.0073 | 0.0083 | 0.0273 | 0.017 | 0.019 | -0.0043 | 0.0066 | 0.0687 |
| | [0.11] | [0.47] | [0.96] | [1.50] | [1.01] | [1.37] | [-0.21] | [0.07] | [1.34] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted** $R^2$ | 0.499 | 0.438 | 0.056 | 0.258 | 0.256 | 0.189 | 0.352 | 0.498 | 0.297 |
| | | | | **Using Placebo as Control Group** | | | | | |
| **Rail Within 20** $km$ | 0.955*** | 0.126*** | 0.00638 | -0.0520** | 0.0555*** | 0.0383** | 0.132*** | 0.507*** | 0.142** |
| | [2.99] | [4.52] | [0.60] | [-2.18] | [2.70] | [2.17] | [3.27] | [3.56] | [2.04] |
| **Observations** | 19602 | 19420 | 19519 | 19519 | 19505 | 19569 | 19578 | 19617 | 19617 |
| **Adjusted** $R^2$ | 0.476 | 0.419 | 0.059 | 0.287 | 0.27 | 0.183 | 0.351 | 0.531 | 0.374 |
| | | | | **By Date of Construction** | | | | | |
| **Rail Within 20** $km$ | 1.387*** | 0.139*** | 0.0170** | -0.0743*** | 0.0775*** | 0.0697*** | 0.172*** | 0.677*** | 0.244*** |
| | [4.38] | [4.93] | [2.20] | [-4.33] | [3.82] | [3.91] | [5.00] | [5.46] | [3.58] |
| **X 1912** | -0.092 | -0.0202 | -0.0019 | -0.0048 | -0.0135 | -0.0251 | -0.0431 | -0.0925 | -0.0735 |
| | [-0.26] | [-0.68] | [-0.19] | [-0.29] | [-0.60] | [-1.47] | [-1.16] | [-0.73] | [-1.06] |
| **X 1930** | -0.546 | -0.0364 | -0.0019 | 0.0236 | -0.0137 | -0.0006 | -0.0163 | -0.0658 | -0.16 |
| | [-1.16] | [-0.91] | [-0.16] | [0.71] | [-0.42] | [-0.02] | [-0.32] | [-0.34] | [-1.36] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted** $R^2$ | 0.5 | 0.439 | 0.056 | 0.258 | 0.256 | 0.189 | 0.352 | 0.498 | 0.296 |
| | | | | **By Reason For Construction** | | | | | |
| **Rail Within 20** $km$ | 1.375*** | 0.139*** | 0.0171** | -0.0733*** | 0.0780*** | 0.0711*** | 0.175*** | 0.684*** | 0.241*** |
| | [4.36] | [4.91] | [2.24] | [-4.33] | [3.85] | [3.99] | [5.06] | [5.53] | [3.58] |
| **X Agriculture** | 0.0135 | -0.005 | -0.0003 | -0.008 | -0.0096 | -0.0148 | -0.0208 | 0.012 | -0.0728 |
| | [0.04] | [-0.16] | [-0.03] | [-0.49] | [-0.43] | [-0.87] | [-0.55] | [0.09] | [-1.05] |
| **X Mineral** | -0.953** | -0.0804** | -0.0057 | 0.0409 | -0.0231 | -0.02 | -0.0652 | -0.328* | -0.188 |
| | [-2.00] | [-2.23] | [-0.44] | [0.98] | [-0.64] | [-0.83] | [-1.25] | [-1.67] | [-1.43] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted** $R^2$ | 0.501 | 0.439 | 0.056 | 0.258 | 0.256 | 0.189 | 0.353 | 0.499 | 0.296 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, ** $p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local (official or artificial) area level in brackets. The table estimates the impact of the railway across different distances to the railway line in the first panel, compares rail localities to localities close to proposed (but not constructed) lines in the second panel, and shows estimates by date of construction (between 1898-1912, 1913-1930, and the excluded category is the railway constructed in 1964) in the third panel. The fourth panel estimates the impact of the railway by reason for construction. The possible reasons are agricultural, mineral, or political (military) development (the excluded category). All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 8:** Long-Run Effects of Railway on Urbanization Outcomes

| | All | North | South | South 2 Mining Nodes Excluded | All | North | South |
|---|---|---|---|---|---|---|---|
| | | | | | Controlling for 1960 Urbanization | | |
| | **Dependent Variable: Z-score of City Presence in 2010** | | | | | | |
| **Rail Within 20 $km$** | 0.106** | 0.157*** | -0.0606 | -0.0796 | 0.0587 | 0.0974** | -0.0662 |
| | [2.07] | [2.96] | [-0.46] | [-0.65] | [1.37] | [2.15] | [-0.63] |
| **1900 City Z-score** | 0.182*** | 0.201*** | 0.172*** | 0.171*** | | | |
| | [13.53] | [12.37] | [9.33] | [9.20] | | | |
| **1960 City Z-score** | | | | | 0.454*** | 0.472*** | 0.432*** |
| | | | | | [51.77] | [51.16] | [29.45] |
| **Observations** | 7708 | 6153 | 1555 | 1553 | 7708 | 6153 | 1555 |
| **Adjusted $R^2$** | 0.135 | 0.086 | 0.173 | 0.174 | 0.289 | 0.234 | 0.34 |
| | **Dependent Variable: Z-score of urban population in 2010** | | | | | | |
| **Rail Within 20 $km$** | 0.179*** | 0.127** | 0.134 | -0.0319 | 0.0158 | 0.0138 | 0.000511 |
| | [2.71] | [2.36] | [0.75] | [-0.40] | [0.81] | [1.06] | [0.01] |
| **1900 Pop. Z-score** | 0.404*** | 1.304*** | 0.244*** | 0.243*** | | | |
| | [3.36] | [3.26] | [4.73] | [4.71] | | | |
| **1960 Pop. Z-score** | | | | | 0.907*** | 0.945*** | 0.833*** |
| | | | | | [16.58] | [13.49] | [7.11] |
| **Observations** | 7708 | 6153 | 1555 | 1553 | 7708 | 6153 | 1555 |
| **Adjusted $R^2$** | 0.236 | 0.437 | 0.242 | 0.313 | 0.839 | 0.925 | 0.697 |

**Notes:** $*p < .1$, ** $p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the grid cell level in brackets. The Table estimates the impact of proximity to the railway line on urbanization in 2010 (measured as city presence and urban population) within $10km \times 10km$ local grid cells. The Z-score is simply the standardized score of the variable of interest, computed as the difference from the mean divided by the standard deviation. We estimate the impact of the railway as described in the paper (including all controls). We also control for the presence of mission stations within the grid cell and update the measure of rail connectedness to reflect the line completed after 1960. All regressions include state of residence (37) fixed effects. Climatic and geographic controls are measured as the average within the grid cell.

**Table 9:** Effect of Proximity to Railway in North and South of Country

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | **Estimation Strategy: Fixed Effects** | | | | | |
| | | | | **North** | | | | | |
| **Rail Within 20 $km$** | 1.817*** | 0.179*** | 0.0217*** | -0.0955*** | 0.0975*** | 0.0961*** | 0.214*** | 0.831*** | 0.235*** |
| | [5.47] | [5.88] | [3.90] | [-4.91] | [5.28] | [5.12] | [6.08] | [6.69] | [3.97] |
| **Missions in Local Area** | 1.209*** | 0.109*** | 0.0074 | -0.105*** | 0.0670** | 0.0301 | 0.104*** | 0.409*** | 0.206** |
| | [2.95] | [3.17] | [0.80] | [-2.78] | [2.55] | [1.20] | [2.59] | [2.70] | [2.57] |
| **Observations** | 19620 | 19451 | 19554 | 19554 | 19650 | 19650 | 19650 | 19650 | 19650 |
| **Adjusted $R^2$** | 0.444 | 0.391 | 0.061 | 0.386 | 0.27 | 0.196 | 0.298 | 0.451 | 0.247 |
| | | | | **South** | | | | | |
| **Rail Within 20 $km$** | 0.167 | 0.029 | 0.0038 | -0.0268 | 0.0082 | -0.0052 | 0.0457 | 0.219 | 0.065 |
| | [0.50] | [1.29] | [0.29] | [-0.90] | [0.32] | [-0.33] | [1.13] | [1.41] | [0.77] |
| **Missions in Local Area** | 1.148*** | 0.0728*** | 0.0313*** | -0.0989*** | 0.0999*** | 0.0492*** | 0.135*** | 0.534*** | 0.236*** |
| | [6.09] | [5.04] | [4.17] | [-5.25] | [5.29] | [4.31] | [6.29] | [5.97] | [4.60] |
| **Observations** | 19315 | 19163 | 19235 | 19235 | 19320 | 19320 | 19320 | 19320 | 19320 |
| **Adjusted $R^2$** | 0.216 | 0.187 | 0.048 | 0.194 | 0.156 | 0.113 | 0.166 | 0.351 | 0.329 |
| | | | | **Estimation Strategy: 2SLS** | | | | | |
| | | | | **North** | | | | | |
| **Rail Within 20 $km$** | 2.096*** | 0.189*** | 0.011 | -0.0729** | 0.0972*** | 0.101*** | 0.233*** | 0.895*** | 0.157* |
| | [3.94] | [4.05] | [1.31] | [-2.45] | [3.07] | [3.40] | [4.41] | [4.40] | [1.67] |
| **Observations** | 18678 | 18519 | 18619 | 18619 | 18576 | 18654 | 18636 | 18708 | 18708 |
| **Centered $R^2$** | 0.443 | 0.39 | 0.06 | 0.395 | 0.275 | 0.189 | 0.289 | 0.435 | 0.223 |
| | | | | **South** | | | | | |
| **Rail Within 20 $km$** | 0.523 | 0.0556* | 0.0115 | -0.0466 | 0.0498 | 0.0156 | 0.110* | 0.370* | 0.112 |
| | [1.15] | [1.68] | [0.69] | [-1.09] | [1.23] | [0.57] | [1.88] | [1.70] | [0.92] |
| **Observations** | 18550 | 18409 | 18470 | 18470 | 18492 | 18514 | 18523 | 18554 | 18554 |
| **Centered $R^2$** | 0.213 | 0.186 | 0.049 | 0.192 | 0.156 | 0.117 | 0.166 | 0.343 | 0.333 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates the impact of distance to the railway on Northern (West, East, Central) and Southern (West, East, South) Nigeria. The instrument (Distance to Line) is a measure of the distance of the individual's cluster to a straight line joining historically important nodes of the railway line as described in the text. All nodes, identified as all clusters within the local government area with the rail station regardless of historical importance, are dropped from the regressions as they may have been endogenously chosen. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 10:** Differential Impact by Cohort (Non-Migrants Only)

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | Estimation Strategy: OLS | | | | | |
| | | | | North | | | | | |
| **Rail Within 20 $km$** | 1.373*** | 0.137*** | 0.0164 | -0.0971*** | 0.0857*** | 0.100*** | 0.139*** | 0.631*** | 0.202*** |
| | [3.71] | [4.08] | [1.23] | [-2.80] | [3.49] | [4.42] | [3.65] | [5.11] | [3.33] |
| **X 1975-1984** | 0.456 | 0.0274 | 0.0141 | -0.0051 | 0.0043 | -0.0607** | 0.0480* | 0.0368 | 0.0145 |
| | [1.34] | [0.88] | [0.94] | [-0.18] | [0.18] | [-2.28] | [1.90] | [0.70] | [0.61] |
| **X 1985-1993** | 0.646** | 0.0812*** | -0.0133 | 0.0209 | 0.0196 | 0.0014 | 0.0665** | 0.139* | 0.0378 |
| | [1.99] | [2.66] | [-0.84] | [0.70] | [0.68] | [0.05] | [2.34] | [1.69] | [1.29] |
| **1975-1984** | 0.406** | 0.0032 | -0.0035 | 0.0141 | 0.0612*** | 0.0044 | 0.0396* | 0.143*** | 0.0304 |
| | [2.00] | [0.14] | [-0.45] | [0.75] | [3.26] | [0.23] | [1.96] | [2.76] | [1.60] |
| **1985-1993** | 1.579*** | 0.0905** | 0.0078 | -0.0035 | 0.142*** | 0.0277 | 0.126*** | 0.241** | 0.0598 |
| | [4.39] | [2.21] | [0.47] | [-0.11] | [3.91] | [0.83] | [3.82] | [2.42] | [1.50] |
| **Observations** | 9392 | 9309 | 9353 | 9353 | 9348 | 9374 | 9371 | 9407 | 9407 |
| **Adjusted $R^2$** | 0.375 | 0.335 | 0.028 | 0.378 | 0.207 | 0.208 | 0.273 | 0.398 | 0.234 |
| | | | | South | | | | | |
| **Rail Within 20 $km$** | -0.355 | 0.0174 | -0.0214 | -0.0055 | -0.0128 | -0.018 | -0.0075 | 0.039 | 0.0189 |
| | [-0.73] | [0.42] | [-1.04] | [-0.10] | [-0.35] | [-0.65] | [-0.13] | [0.23] | [0.19] |
| **X 1975-1984** | 0.635** | 0.0132 | 0.0242 | 0.0006 | 0.0256 | 0.0214 | 0.0393 | 0.192*** | 0.0636** |
| | [2.00] | [0.38] | [1.45] | [0.01] | [0.79] | [0.88] | [1.08] | [2.79] | [2.10] |
| **X 1985-1993** | 0.299 | -0.0378 | 0.0187 | 0.0286 | 0.0402 | 0.0006 | 0.0152 | 0.157*** | 0.0477* |
| | [0.89] | [-0.99] | [1.10] | [0.71] | [1.30] | [0.02] | [0.43] | [2.79] | [1.82] |
| **1975-1984** | 1.822*** | 0.0795*** | 0.009 | -0.0395 | 0.146*** | 0.0412* | 0.0561** | 0.0963* | 0.0142 |
| | [7.40] | [3.49] | [0.65] | [-1.62] | [5.55] | [1.92] | [2.33] | [1.83] | [0.68] |
| **1985-1993** | 2.909*** | 0.142*** | -0.0034 | -0.0456 | 0.210*** | 0.0994*** | 0.0857** | 0.172** | 0.0446 |
| | [9.17] | [4.42] | [-0.17] | [-1.43] | [4.63] | [2.94] | [2.34] | [2.32] | [1.32] |
| **Observations** | 7375 | 7328 | 7354 | 7354 | 7357 | 7363 | 7366 | 7377 | 7377 |
| **Adjusted $R^2$** | 0.262 | 0.23 | 0.045 | 0.216 | 0.158 | 0.142 | 0.188 | 0.334 | 0.288 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, ** $p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates, by cohort, the impact of distance to the railway on Northern (West, East, Central) and Southern (West, East, South) Nigeria. The omitted cohort are older individuals born between 1948-1974. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distances to rail network and roads are computed using DHS data and information on rail network. Presence of river is computed using information on presence of river within the local government area. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 11:** Short-Run Effects of the Railway on Urbanization Outcomes

| | All | North | South | South<br>2 Mining Nodes Excluded | All | North<br>All Nodes Excluded | South |
|---|---|---|---|---|---|---|---|
| | | | | Dependent Variable: Z-score of City Presence in 1960 | | | |
| **Rail Within 20 $km$** | 0.104** | 0.128*** | 0.0127 | -0.0313 | 0.065 | 0.0861** | -0.0149 |
| | [2.21] | [2.96] | [0.08] | [-0.21] | [1.45] | [2.14] | [-0.10] |
| **Mission in Grid** | 1.245*** | 1.130*** | 1.367*** | 1.400*** | 1.244*** | 1.040*** | 1.440*** |
| | [5.16] | [3.48] | [3.86] | [3.92] | [5.11] | [3.31] | [3.93] |
| **1900 City Z-score** | 0.400*** | 0.412*** | 0.389*** | 0.388*** | 0.397*** | 0.406*** | 0.386*** |
| | [16.27] | [9.30] | [12.53] | [12.42] | [15.38] | [7.75] | [12.23] |
| **Observations** | 7708 | 6153 | 1555 | 1553 | 7685 | 6139 | 1546 |
| **Adjusted $R^2$** | 0.256 | 0.167 | 0.318 | 0.323 | 0.253 | 0.145 | 0.324 |
| | | | | Dependent Variable: Z-score of urban population in 1960 | | | |
| **Rail Within 20 $km$** | 0.175*** | 0.122** | 0.149 | 0.0359 | 0.120* | 0.0908* | 0.0498 |
| | [2.69] | [2.30] | [0.97] | [0.33] | [1.94] | [1.75] | [0.44] |
| **Mission in Grid** | 0.932*** | 0.493** | 1.123*** | 1.162*** | 0.946*** | 0.504** | 1.195*** |
| | [4.18] | [2.01] | [3.47] | [3.58] | [4.15] | [1.99] | [3.58] |
| **1900 Pop. Z-score** | 0.506*** | 1.345*** | 0.355*** | 0.355*** | 0.489*** | 1.434** | 0.353*** |
| | [3.61] | [2.65] | [4.44] | [4.43] | [3.45] | [2.31] | [4.41] |
| **Observations** | 7708 | 6153 | 1555 | 1553 | 7685 | 6139 | 1546 |
| **Adjusted $R^2$** | 0.318 | 0.441 | 0.413 | 0.464 | 0.332 | 0.452 | 0.465 |
| | | | | Dependent Variable: Z-score of urban population in 1980 | | | |
| **Rail Within 20 $km$** | 0.182*** | 0.125** | 0.145 | 0.0218 | 0.116* | 0.0843* | 0.0354 |
| | [2.78] | [2.39] | [1.00] | [0.25] | [1.92] | [1.77] | [0.40] |
| **Mission in Grid** | 1.013*** | 0.646** | 1.124*** | 1.165*** | 1.020*** | 0.649** | 1.197*** |
| | [4.52] | [2.51] | [3.68] | [3.80] | [4.48] | [2.47] | [3.81] |
| **1900 Pop. Z-score** | 0.470*** | 1.416*** | 0.302*** | 0.301*** | 0.446*** | 1.486** | 0.300*** |
| | [3.26] | [2.67] | [4.14] | [4.12] | [3.08] | [2.25] | [4.11] |
| **Observations** | 7708 | 6153 | 1555 | 1553 | 7685 | 6139 | 1546 |
| **Adjusted $R^2$** | 0.283 | 0.448 | 0.382 | 0.453 | 0.305 | 0.474 | 0.454 |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the grid cell level in brackets. The Table estimates the impact of proximity to the railway line on urbanization in 2010 (measured as city presence and urban population) within $10km \times 10km$ local grid cells. The Z-score is simply the standardized score of the variable of interest, computed as the difference from the mean divided by the standard deviation. We estimate the impact of the railway as described in the paper (including all controls). We also control for the presence of mission stations within the grid cell and update the measure of rail connectedness to reflect the line completed after 1960. All regressions include state of residence (37) fixed effects. Climatic and geographic controls are measured as the average within the grid cell.

**Table 12:** Benefits of Shipping by Rail For Key Regional Crops

| | Northern Crops | | Southern Crops | |
|---|---|---|---|---|
| **Esimates** | **Groundnuts** | **Cotton** | **Palm Oil** | **Cocoa** |
| **Rail Price (pence per ton km)** | 1.95 | 1.37 | 3.95 | 2.08 |
| **River Price (pence per ton km)** | .9 (+ 3.1 rail) | 2.5 (+ 3.1 rail) | 1.8 | 1.8 |
| **Road Price (pence per ton km)** | 5.6 | 5.6 | 2.5 | 1.3 |
| **Distance Rail (km)** | 1127 | 1159 | 61 | 193 |
| **Distance River (km)** | 575 river (552 rail) | 575 river (584 rail) | 61 | 193 |
| **Cost Reduction From Rail** | | | | |
| **As % of River Cost** | -1.4 | -51.1 | 119.4 | 15.6 |
| **As % of Road Cost** | -65.2 | -75.5 | 58 | 60 |

Table 12 calculates the benefit of the railway over the period 1945-1949. Data on prices, quantities, and distances are obtained from sources described in text. For river shipments in the North, the cost is estimated as the cost of railing to Baro and then shipping by river to the Delta ports, hence the rail price and distance in parentheses . We use the railing distance as the shipping distance for rivers in the South, although this might be an overestimate given the proximity of the South to several rivers which lead to the coast.

**Table 13:** Effect of Proximity to Railway By Distance to Coastal Port

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | **Estimation Strategy: OLS** | | | | | | | | |
| | **Above Median Distance to Port (192 $km$)** | | | | | | | | |
| **Rail Within 20 $km$** | 1.697*** | 0.166*** | 0.0205*** | -0.0822*** | 0.0923*** | 0.0937*** | 0.200*** | 0.768*** | 0.205*** |
| | [5.47] | [5.81] | [3.88] | [-4.25] | [5.34] | [5.32] | [5.96] | [6.49] | [3.58] |
| **Observations** | 22101 | 21924 | 22025 | 22025 | 21991 | 22066 | 22055 | 22131 | 22131 |
| **Adjusted $R^2$** | 0.489 | 0.425 | 0.065 | 0.358 | 0.274 | 0.201 | 0.324 | 0.462 | 0.243 |
| | **Below Median Distance to Port (192 $km$)** | | | | | | | | |
| **Rail Within 20 $km$** | -0.122 | 0.0062 | -0.0079 | -0.0255 | -0.0184 | -0.0207 | 0.0163 | 0.155 | 0.129 |
| | [-0.36] | [0.27] | [-0.52] | [-0.85] | [-0.67] | [-1.31] | [0.37] | [0.91] | [1.46] |
| **Observations** | 16834 | 16690 | 16764 | 16764 | 16777 | 16803 | 16811 | 16839 | 16839 |
| **Adjusted $R^2$** | 0.203 | 0.173 | 0.048 | 0.195 | 0.164 | 0.104 | 0.169 | 0.359 | 0.374 |
| | **Interaction with Distance to Port** | | | | | | | | |
| **Rail Within 20 $km$** | 0.29 | 0.0303 | 0.0073 | -0.0518* | 0.0279 | -0.0055 | 0.0483 | 0.293* | 0.105 |
| | [0.92] | [1.44] | [0.61] | [-1.84] | [1.08] | [-0.37] | [1.23] | [1.94] | [1.26] |
| **X 200-400 $km$** | 1.495** | 0.128** | 0.0276 | -0.0775 | 0.0608 | 0.118*** | 0.159** | 0.4 | 0.0439 |
| | [2.20] | [2.23] | [1.65] | [-1.08] | [1.38] | [2.87] | [1.99] | [1.58] | [0.29] |
| **X 400-600 $km$** | 1.411** | 0.133** | 0.0042 | -0.0656 | 0.0862* | 0.152*** | 0.205*** | 0.432 | 0.348** |
| | [1.98] | [2.34] | [0.22] | [-1.33] | [1.85] | [3.66] | [2.68] | [1.52] | [2.17] |
| **X $\geq$ 600 $km$** | 1.588*** | 0.162*** | 0.0125 | -0.0136 | 0.0686** | 0.0860*** | 0.167*** | 0.601*** | 0.092 |
| | [2.91] | [3.59] | [0.92] | [-0.41] | [2.01] | [3.14] | [2.79] | [2.73] | [0.83] |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates the impact of distance to the railway by to distance to a coastal port (Bonny, Burutu, Calabar, Degema, Lagos, Opobo, Port Harcourt, Sapele, Warri). All nodes, identified as all clusters within the local government area with the rail station, are dropped from the regressions as they may have been endogenously chosen. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table 14:** Effect of Railway By Proximity to Early Cities

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | **By Reason For Construction** | | | | | | | | |
| **Rail Within 20 $km$** | 1.123*** | 0.110*** | 0.0176*** | -0.0687*** | 0.0700*** | 0.0613*** | 0.143*** | 0.599*** | 0.158*** |
| | [4.49] | [4.73] | [3.36] | [-4.04] | [4.40] | [4.49] | [4.82] | [5.67] | [3.05] |
| **X Within 20 $km$ of 1900 city** | -0.53 | -0.0247 | -0.0284* | 0.0551* | -0.0697** | -0.0514** | -0.069 | -0.393** | -0.117 |
| | [-1.19] | [-0.68] | [-1.79] | [1.93] | [-2.40] | [-2.27] | [-1.45] | [-2.34] | [-1.27] |
| **Within 20 $km$ of 1900 city** | 1.755*** | 0.138*** | 0.0401*** | -0.119*** | 0.128*** | 0.0868*** | 0.197*** | 0.858*** | 0.329*** |
| | [7.11] | [6.37] | [3.40] | [-4.93] | [5.87] | [5.57] | [6.30] | [7.40] | [5.44] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38768 | 38869 | 38866 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.506 | 0.444 | 0.058 | 0.263 | 0.267 | 0.192 | 0.366 | 0.522 | 0.325 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates the impact of the railway, by proximity to a city in 1900. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using DHS data and information on rail network. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

# Appendix: Other Results and Robustness Checks

**Table A1:** Robustness to Other Measures of Connectedness

| | (1) Schooling | (2) Literacy | (3) Professional | (4) Agricultural | (5) Read Paper | (6) Listen to Radio | (7) Watch TV | (8) Wealth | (9) Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| **Closeness to Railway Lines** | | | | | | | | | |
| **Closeness to Rail** | 0.436*** | 0.0399*** | 0.00618*** | -0.0237*** | 0.0256*** | 0.0195*** | 0.0531*** | 0.212*** | 0.0552*** |
| | [5.88] | [5.74] | [3.34] | [-5.48] | [5.86] | [4.59] | [6.68] | [7.21] | [3.87] |
| **Observations** | 38935 | 38614 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.5 | 0.438 | 0.056 | 0.256 | 0.257 | 0.189 | 0.353 | 0.5 | 0.29 |
| **Proximity to Railway Station** | | | | | | | | | |
| **Rail Station Within 20 $km$** | 1.572*** | 0.148*** | 0.0187*** | -0.0972*** | 0.0896*** | 0.0504*** | 0.178*** | 0.706*** | 0.230*** |
| | [5.17] | [5.52] | [2.67] | [-6.10] | [5.34] | [3.25] | [5.64] | [6.29] | [4.19] |
| **Observations** | 38935 | 38614 | 38789 | 38789 | 38768 | 38869 | 38866 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.5 | 0.438 | 0.057 | 0.259 | 0.263 | 0.189 | 0.355 | 0.497 | 0.295 |
| **Presence of rail tracks in local area** | | | | | | | | | |
| **Rail in Local Area** | 0.635** | 0.0663*** | 0.0114** | -0.0270* | 0.0398*** | 0.0413*** | 0.0909*** | 0.374*** | 0.0882* |
| | [2.48] | [2.99] | [2.13] | [-1.74] | [2.60] | [3.04] | [3.22] | [3.69] | [1.82] |
| **Observations** | 38935 | 38614 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.494 | 0.433 | 0.056 | 0.253 | 0.254 | 0.188 | 0.345 | 0.484 | 0.279 |
| **Presence of rail tracks in grid cell** | | | | | | | | | |
| **Rail in Grid** | 0.907*** | 0.0859*** | 0.00999** | -0.0504*** | 0.0553*** | 0.0379*** | 0.117*** | 0.436*** | 0.0782* |
| | [3.81] | [3.94] | [2.10] | [-3.15] | [3.85] | [2.63] | [4.69] | [4.61] | [1.80] |
| **Observations** | 38935 | 38614 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.498 | 0.437 | 0.055 | 0.253 | 0.255 | 0.189 | 0.351 | 0.498 | 0.28 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local area level in brackets. The table estimates the impact of an individual's closeness to the railway line on various development outcomes, defined as the log of the inverse of 1 plus the distance of the individual's cluster to the railway line. Rail in Grid is equal to 1 if the cluster is located within a $40km \times 40km$ grid with a railway line, and Rail in Local Area is equal 1 if a railway line crosses the individual's local government area. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distances to rail network and roads are computed using DHS data and information on rail and road networks. Presence of river is computed using information on presence of river within the local government area. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table A2:** Additional Robustness Checks

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Rural Only** | | | | | | | | | |
| **Rail Within 20 $km$** | 0.757*** | 0.0785*** | 0.00932** | -0.0477*** | 0.0437** | 0.0453*** | 0.0983*** | 0.456*** | — |
| | [2.66] | [3.10] | [2.05] | [-2.75] | [2.53] | [2.93] | [3.17] | [3.96] | |
| **Observations** | 25440 | 25205 | 25342 | 25342 | 25320 | 25394 | 25384 | 25455 | |
| **Adjusted $R^2$** | 0.524 | 0.447 | 0.046 | 0.289 | 0.237 | 0.208 | 0.34 | 0.457 | . |
| **Spatial Correlation-Robust Conley Standard Errors** | | | | | | | | | |
| **Rail Within 20 $km$** | 1.286*** | 0.128*** | 0.0161*** | -0.0737*** | 0.0720*** | 0.0610*** | 0.157*** | 0.637*** | 0.183*** |
| | [4.66] | [5.21] | [3.10] | [-3.89] | [4.33] | [4.02] | [5.51] | [5.72] | [3.68] |
| **Observations** | 38935 | 38614 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 | 38970 |
| **Adjusted $R^2$** | 0.498 | 0.438 | 0.056 | 0.257 | 0.262 | 0.189 | 0.355 | 0.498 | 0.291 |
| **Migrants** | | | | | | | | | |
| **Rail Within 20 $km$** | 1.401*** | 0.140*** | 0.0209*** | -0.0740*** | 0.0707*** | 0.0754*** | 0.193*** | 0.747*** | 0.200*** |
| | [5.41] | [5.97] | [2.84] | [-4.42] | [4.23] | [4.69] | [6.62] | [7.05] | [3.79] |
| **Observations** | 22168 | 21977 | 22082 | 22082 | 22186 | 22186 | 22186 | 22186 | 22186 |
| **Adjusted $R^2$** | 0.535 | 0.49 | 0.066 | 0.281 | 0.291 | 0.204 | 0.379 | 0.531 | 0.33 |
| **Non-Migrants** | | | | | | | | | |
| **Rail Within 20 $km$** | 1.143*** | 0.114*** | 0.009 | -0.0697*** | 0.0743*** | 0.0469*** | 0.119*** | 0.532*** | 0.162*** |
| | [3.92] | [4.65] | [1.35] | [-3.21] | [4.09] | [3.13] | [4.00] | [5.18] | [3.00] |
| **Observations** | 16767 | 16637 | 16707 | 16707 | 16784 | 16784 | 16784 | 16784 | 16784 |
| **Adjusted $R^2$** | 0.458 | 0.39 | 0.038 | 0.277 | 0.214 | 0.188 | 0.323 | 0.435 | 0.237 |
| **Mission Station Not in Official Local Area** | | | | | | | | | |
| **Rail Within 20 $km$** | 1.369*** | 0.133*** | 0.0208*** | -0.0716*** | 0.0718*** | 0.0549*** | 0.165*** | 0.683*** | 0.194*** |
| | [4.86] | [5.27] | [3.42] | [-4.16] | [4.29] | [3.64] | [5.15] | [5.83] | [3.52] |
| **Observations** | 30627 | 30358 | 30509 | 30509 | 30654 | 30654 | 30654 | 30654 | 30654 |
| **Adjusted $R^2$** | 0.511 | 0.445 | 0.055 | 0.268 | 0.269 | 0.192 | 0.35 | 0.494 | 0.288 |
| **No Rail Line in Official Local Area (Excludes Nodes and Stations)** | | | | | | | | | |
| **Rail Within 20 $km$** | 0.961** | 0.08932** | 0.0178* | -0.0761*** | 0.0649** | 0.0259 | 0.143*** | 0.614*** | 0.223** |
| | [2.35] | [2.37] | [1.68] | [-3.49] | [2.34] | [1.20] | [3.41] | [3.82] | [2.43] |
| **Observations** | 30179 | 29888 | 29998 | 29998 | 29982 | 30055 | 30046 | 30134 | 30134 |
| **Adjusted $R^2$** | 0.517 | 0.447 | 0.056 | 0.255 | 0.25 | 0.189 | 0.358 | 0.49 | 0.279 |
| **Excluding Areas Within 20 $km$ of Rail Station** | | | | | | | | | |
| **Rail Within 20 $km$** | 0.486* | 0.0567** | 0.0122* | -0.0204 | 0.0314* | 0.0543*** | 0.0813*** | 0.337*** | 0.0723 |
| | [1.79] | [2.58] | [1.92] | [-0.99] | [1.67] | [3.21] | [2.71] | [2.94] | [1.09] |
| **Observations** | 32862 | 32619 | 32749 | 32749 | 32726 | 32802 | 32799 | 32893 | 32893 |
| **Adjusted $R^2$** | 0.526 | 0.453 | 0.057 | 0.259 | 0.267 | 0.189 | 0.366 | 0.505 | 0.3 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $** \ p < 0.05$, $***p < 0.01$. T-statistics for Conley standard errors clustered at the official local government area level in brackets. The table estimates the impact of distance to the railway adjusting for spatial correlation (Conley standard errors), and also computes the effect across various individual and local characteristics. Conley standard errors are computed with a cutoff of 100 $km$. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distances to rail network and roads are computed using DHS data and information on rail and road networks. Presence of river is computed using information on presence of river within the local government area. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

**Table A3:** Falsification Exercise: Placebo Lines Estimates in North and South

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | | **North** | | | | |
| **Within 20 $km$ of Placebo** | 0.0575 | 0.0093 | -0.0005 | -0.0164 | -0.0118 | 0.0315 | 0.0383 | 0.218* | 0.0569 |
| | [0.26] | [0.46] | [-0.07] | [-0.93] | [-0.85] | [1.63] | [1.31] | [1.96] | [1.03] |
| **Observations** | 15049 | 14928 | 15003 | 15003 | 14974 | 15022 | 15007 | 15070 | 15070 |
| **Adjusted $R^2$** | 0.486 | 0.403 | 0.063 | 0.396 | 0.291 | 0.184 | 0.283 | 0.449 | 0.219 |
| | | | | | **Controlling for Presence of Rivers and Roads** | | | | |
| **Within 20 $km$ of Placebo** | -0.0481 | -0.0001 | -0.0016 | -0.007 | -0.015 | 0.0169 | 0.0245 | 0.134 | 0.0169 |
| | [-0.21] | [-0.00] | [-0.26] | [-0.39] | [-1.04] | [0.90] | [0.86] | [1.26] | [0.31] |
| **River in Local Area** | -0.144 | -0.0299 | 0.009 | 0.0113 | 0.0152 | -0.032 | -0.0014 | -0.068 | -0.120* |
| | [-0.42] | [-1.05] | [1.17] | [0.47] | [0.69] | [-1.41] | [-0.04] | [-0.60] | [-1.85] |
| **Distance to Road** | -0.0126** | -0.0009 | -0.000256* | 0.00113** | -0.0006 | -0.00159*** | -0.00184** | -0.0105*** | -0.00398** |
| | [-2.25] | [-1.55] | [-1.66] | [2.24] | [-1.53] | [-2.77] | [-2.53] | [-3.96] | [-2.49] |
| **Observations** | 15049 | 14928 | 15003 | 15003 | 14974 | 15022 | 15007 | 15070 | 15070 |
| **Adjusted $R^2$** | 0.487 | 0.404 | 0.063 | 0.398 | 0.291 | 0.186 | 0.286 | 0.462 | 0.247 |
| | | | | | **South** | | | | |
| **Within 20 $km$ of Placebo** | 0.626** | 0.0474** | 0.0138 | -0.0511* | 0.0581** | 0.0416** | 0.0553* | 0.300** | 0.0892 |
| | [2.28] | [2.37] | [1.57] | [-1.95] | [2.18] | [2.54] | [1.91] | [2.29] | [1.34] |
| **Observations** | 13548 | 13461 | 13499 | 13499 | 13505 | 13518 | 13525 | 13551 | 13551 |
| **Adjusted $R^2$** | 0.213 | 0.19 | 0.048 | 0.184 | 0.129 | 0.125 | 0.16 | 0.287 | 0.29 |
| | | | | | **Controlling for Presence of Rivers and Roads** | | | | |
| **Within 20 $km$ of Placebo** | 0.371 | 0.0305 | 0.0088 | -0.0331 | 0.0333 | 0.0271* | 0.0351 | 0.185 | 0.0535 |
| | [1.46] | [1.65] | [1.00] | [-1.32] | [1.38] | [1.87] | [1.21] | [1.49] | [0.83] |
| **River in Local Area** | -0.445** | -0.0136 | -0.0183** | 0.0294 | -0.0482* | -0.0182 | 0.0027 | -0.0919 | -0.128* |
| | [-1.99] | [-0.81] | [-2.47] | [0.94] | [-1.95] | [-1.41] | [0.09] | [-0.77] | [-1.78] |
| **Distance to Road** | -0.0730*** | -0.00527*** | -0.00116*** | 0.00520*** | -0.00691*** | -0.00440*** | -0.00693*** | -0.0363*** | -0.00822** |
| | [-5.47] | [-5.57] | [-3.04] | [3.48] | [-6.05] | [-6.06] | [-4.90] | [-5.90] | [-2.30] |
| **Observations** | 13548 | 13461 | 13499 | 13499 | 13505 | 13518 | 13525 | 13551 | 13551 |
| **Adjusted $R^2$** | 0.231 | 0.199 | 0.05 | 0.194 | 0.142 | 0.133 | 0.174 | 0.341 | 0.317 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $**\ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the official local government area level in brackets. The table estimates the impact of being within 20 $km$ to a placebo line joining proposed nodes that were not constructed. All regressions include ethnicity (31) and state of residence (37) fixed effects. Distances to rail network and roads are computed using DHS data and information on rail and road networks. Presence of river is computed using information on presence of river within the local government area. Climatic and geographic controls are measured as the average within the official local government area. Data on Christian mission stations come from maps described in the text. All other variables are taken from the 2008 Nigeria Demographic and Health Survey.

## **Table A4:** Placebo as Control Group in North and South

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | | **North** | | | | |
| **Rail Within 20 $km$** | 1.688*** | 0.185*** | 0.0236*** | -0.0645** | 0.0864*** | 0.0770*** | 0.197*** | 0.726*** | 0.168* |
| | [4.51] | [5.11] | [3.34] | [-2.40] | [4.02] | [3.46] | [4.12] | [4.55] | [1.94] |
| **Observations** | 14334 | 14186 | 14268 | 14268 | 14256 | 14314 | 14318 | 14347 | 14347 |
| **Adjusted $R^2$** | 0.502 | 0.452 | 0.064 | 0.342 | 0.326 | 0.206 | 0.381 | 0.576 | 0.415 |
| | | | | | **South** | | | | |
| **Rail Within 20 $km$** | -0.238 | 0.032 | -0.0223 | -0.0249 | -0.000394 | -0.0156 | 0.027 | 0.159 | 0.0993 |
| | [-0.52] | [0.98] | [-0.93] | [-0.65] | [-0.01] | [-0.75] | [0.41] | [0.66] | [0.84] |
| **Observations** | 15577 | 15434 | 15509 | 15509 | 15509 | 15555 | 15565 | 15590 | 15590 |
| **Adjusted $R^2$** | 0.385 | 0.334 | 0.058 | 0.236 | 0.221 | 0.161 | 0.28 | 0.476 | 0.382 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $** \ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the official local government area level in brackets. The table estimates separately for Southern and Northern Nigeria the impact of being within 20 $km$ of the railway line relative to being within 20 $km$ of the placebo line. All regressions include ethnicity (31) and state of residence (37) fixed effects. Climatic and geographic controls are from Fischer et al. (2008) and Hijmans et al. (2005), and are measured as the average within the local area. Data on Christian mission stations come from maps published by Ayandele (1966) and Roome (1925). All other variables are taken from the 2008 Nigeria Demographic and Health Survey (NBS and ICF International, 2008).

## **Table A5:** Robustness of No Effect in South Excluding Crude Oil Producers

| | Schooling | Literacy | Professional | Agricultural | Read Paper | Listen to Radio | Watch TV | Wealth | Urban Resident |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | | | | | **South: Excluding Oil Producers** | | | | |
| | | | | | **Estimation Strategy: OLS** | | | | |
| **Rail Within 20 $km$** | -0.02 | 0.0209 | -0.0048 | -0.0082 | -0.0019 | -0.0116 | 0.0268 | 0.167 | 0.0054 |
| | [-0.05] | [0.78] | [-0.27] | [-0.25] | [-0.06] | [-0.60] | [0.48] | [0.83] | [0.06] |
| **Observations** | 9315 | 9225 | 9278 | 9278 | 9319 | 9319 | 9319 | 9319 | 9319 |
| **Adjusted $R^2$** | 0.286 | 0.251 | 0.055 | 0.227 | 0.192 | 0.103 | 0.228 | 0.432 | 0.341 |
| **All Baseline Controls** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Ethnicity Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **State Fixed Effects** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Notes:** $*p < .1$, $** \ p < 0.05$, $***p < 0.01$. T-statistics for standard errors clustered at the local government area level in brackets. The table estimates the impact of distance to the railway in Southern Nigeria (West, East, South) excluding oil producing areas. Oil producing areas are the historically oil producing states (Abia, Akwa Ibom, Bayelsa, Cross River, Delta, Edo, Imo, Ondo, Rivers). All regressions include ethnicity (31) and state of residence (37) fixed effects. Distance to rail network is computed using data from the 2008 Nigeria Demographic and Health Survey and information on rail network taken from DMA (1992). Climatic and geographic controls are from Fischer et al. (2008) and Hijmans et al. (2005), and are measured as the average within the local area. Data on Christian mission stations come from maps published by Ayandele (1966) and Roome (1925). All other variables are taken from the 2008 Nigeria Demographic and Health Survey (NBS and ICF International, 2008).