

The long-term impact of Italian colonial roads in the Horn of Africa, 1935-2015.

16 May 2020

Abstract

This paper examines how a first-mover advantage in transportation, linked to the early construction of paved roads, affects the spatial distribution of economic activity in developing countries in the long-run. The extensive road network, built by the Italian government across the Horn of Africa during the occupation of Ethiopia (1935-1941), provides an ideal quasi-natural experiment to study this issue. The results show that Italian paved roads rendered areas located within 10 km of them significantly more populated, urbanized and luminous around 2010. The analysis of the mechanisms suggests how early road building lifted first-mover locations out of isolation and allowed for net welfare gains (and not simply displacement of economic activity), thanks to a reduction in transport costs and specialization. Instead of showing a persistently higher density equilibrium, to this day, first-mover locations continue to diverge from the control group, thanks to a coordination mechanism that led to an oversupply of governmental facilities in the post-colonial period.

Key words: roads, urbanization, colonialism

JEL codes: N70, N77, O18, R12

1 Introduction

Transport infrastructure is pivotal to development strategies across the globe and absorbs a significant share of spending from governments and development agencies, especially in low-income countries (World Bank, 2007). While extensive research on the localized effect of increases in market access through the construction or improvement of transport facilities exists (see Redding and Turner (2015)), results vary significantly as context-specific characteristics, such as the type of infrastructure, the developmental stage of the country and the time horizon of the study can yield heterogeneous outcomes. For roads in particular, these range from large and positive short-run effects on economic growth (Ghani, Goswami, and Kerr, 2016), to zero gains (Chandra and Thompson, 2000), or even negative net effects (Faber, 2014), if re-location or core-periphery dynamics prevail. Exploring this issue is particularly important for former colonial settings, where scholars have shown sizeable long-term effects of European railway building (Jedwab, Kerby, and Moradi, 2017) but for which little evidence exists on the impact of road infrastructure.

This article analyzes the repercussions of the construction of the extensive road network, built by the Italian government across Eritrea, Somalia and Ethiopia, between 1935 and 1940, for the distribution of the economic activity around 2010. In doing so, the paper contributes to the literature on intercity transportation by studying how a first-mover advantage provided by paved roads (asphalt motorways) can alter the distribution of economic activity in the long-run (as defined by a 80-year period) and how the effect changes over this time span. It also discusses the mechanisms that allowed the effect to emerge in the first place and seeks to explain how it strengthened after the monopoly in transportation, initially enjoyed by the colonial infrastructure, evaporated.

The setting of this study presents some unique features that make it akin to a natural experiment and, thus, allow for a causal interpretation of the results. First, all major Italian roads were created by paving the military tracks that had been built in the context of the Second Italo-Ethiopian War of 1935-1936. These tracks had been laid out following least-cost paths to connect a handful of politically and militarily strategic urban centers that were scattered across the immense Ethiopian territory. This feature allows identification grounded in

an inconsequential unit approach, given the plausibly quasi-random allocation of infrastructure between targets. Second, the Italian rule in Ethiopia lasted for just five years, which allows for effectively isolating the role of infrastructure from other confounding factors, such as human capital formation, changes to the land tenure system, and the array of other investments that could be associated with a prolonged colonial rule.

Obtaining a reliable measure of economic activity to trace changes across space and over time is a critical empirical challenge for the paper, especially given the lack of micro-level statistics for the selected countries. I address this issue in four steps. First, I employ data on population density between 1920 and 2000, a plausible proxy for economic activity (Henderson et al., 2017, pp. 370–1) and productivity (Ciccone and Hall, 1996). Second, to capture changes in urbanization, I collect information on the location of cities with more than 10,000 residents between 1925 and 2015. Third, as population density and urbanization are both imperfect proxies for welfare, I include data on night-time light density, an increasingly popular proxy for GDP in developing countries (Henderson, Storeygard, and Weil, 2012; Michalopoulos and Papaioannou, 2013), available in a consistent format between 1992 and 2013. Fourth, I employ information on wealth and occupational structure from the 2011 Demographic and Health Survey (DHS) of Ethiopia. I combine these outcome data with an array of historical information on the extent and quality of major transport facilities over the entire period of the study, other types of colonial investments, pre-colonial infrastructure and geographical features through a grid. This contains 15,288 non-nodal cells, each measuring roughly 11x11 km (0.1x0.1 decimal degrees) and covers Eritrea, Ethiopia, and Italian Somalia (see Figure 1c).

I start by testing, through regression analysis, whether proximity to colonial roads correlates with significantly higher levels of economic activity today, and whether this translates into gains in welfare. The estimates show that non-nodal cells located within 10 km of Italian roads are significantly more populous, luminous and urbanized 80 years after construction, relative to a baseline control group that includes all untreated cells of former Italian East Africa. The estimates are robust to various restrictions of the sample size. Second-mover locations, that received a paved motorway during the post-colonial period, do not show a significantly higher concentration of economic activity today, relative to untreated cells. I corroborate the grid-level analysis with individual-level data from the 2011 Demographic and Health Survey of

Ethiopia (DHS), and find that people living in treated locations are also wealthier. Crucially, the long-term differences in levels mask interesting time-series heterogeneity in growth rates. The effect of colonial roads on population density emerged by 1940. Rather than forming a stable equilibrium, economic activity increasingly concentrated in treated cells during the post-colonial era, and especially between 1990 and 2015, as reflected by the concentration of cities and luminosity.

To address endogeneity and to relax the assumption of random allocation between targets, I rely on two main robustness tests in line with the literature (Redding and Turner, 2015). First, I instrument colonial roads by least cost paths (based on ruggedness and land cover) connecting the network’s targets. Identification is based on the intuition that, conditional on the full set of controls, proximity to the least-cost paths between towns targeted by the Italian road network only affected long term development through colonial roads. Second, I implement a placebo test that compares the effect of Italian paved roads with unfinished roads, namely those tracks that the Italians were in the process of paving but could not complete due to the outbreak of World War II. Both approaches corroborate the fact, implicit in the inconsequential units approach’s identifying assumptions, that Italian motorways made first-mover locations better off in the long-run.

The case study is particularly relevant in the context of the literature, as related contributions that have looked at the long-term effect of transport infrastructure have mainly focused on railways which, compared to roads, are more expensive to build and to maintain.¹ This makes motorways a more flexible and popular development tool, particularly in developing countries (Ghani, Goswami, and Kerr, 2016), and renders the study of their long-term effect key for understanding patterns of spatial inequality in such settings. Those papers that have studied the effect of roads, in fact, have mainly considered their short-run impact and in countries at a relatively advanced developmental stage, where new transport infrastructure represented an improvement in connectivity and not a transition from a state of substantial isolation to one of market integration, as in context of this study (see for a review Redding

¹See Chaves, Engerman, and Robinson (2014) for a discussion of transport infrastructure building costs in the African context. Maintenance costs per mile of roads are lower than for railroads, see for instance <http://www.transport-watch.co.uk/facts-sheet-8-rail-versus-road-track-maintenance-costs>, last accessed 10th March 2019.

and Turner (2015)). Before the late 1930s, in fact, the Horn of Africa completely lacked asphalt roads, while railway and navigable waterways networks were negligible. As such, most regions faced prohibitive shipping costs (Benti, 2016, p.82), in particular between June and September, when the rainy season prevented wheeled transportation along the existing dirt roads. This situation, that Eaton and Kortum (2002) would define as autarchic, heavily reduced the gains from motorization, which was virtually absent before 1935 (Cobolli-Gigli, 1938, pp.39-41).

The exceptionally high, initial value of Italian motorways in the Ethiopian context is reflected in two findings. First, there is very little evidence of re-organization of economic activity. Untreated cells within a 20-60 km range from colonial roads are not significantly different from comparable, more distant locations. This holds true whether we look at economic outcomes around 2010 or population distribution in the late 1930s, which suggests net growth rather than simple re-shuffling of the existing economic activity. In this sense, my results depart from those contributions that have highlighted predominant relocation effects following transportation improvements (Chandra and Thompson, 2000; Buchel and Kyburz, 2020). Second, a comparison with related works on transport infrastructure that look at a similar time-span (Banerjee, Duflo, and Qian, 2012) reveals how the effect of Italian roads on GDP is larger, in relative terms.

These results are consistent with a model featuring welfare gains from falling barriers to trade and specialization along the lines of Ricardian comparative advantage, as proposed by Eaton and Kortum (2002) and Redding and Turner (2015). In their frameworks, agglomeration in historically better-connected locations is expected, as these experience an initial rise in real wages, following productivity gains from specialization that leads to in-migration. Contemporary data on employment from DHS surveys suggest that, in the Ethiopian context, specialization in first-mover cells materialized in a higher concentration of service sector jobs, with a decisive move away from agriculture. This paper is thus in line with those contributions that have highlighted localized economic (Storeygard, 2016; Ghani, Goswami, and Kerr, 2016) and welfare (Aggarwal, 2018; Stifel, Minten, and Koru, 2016) gains from road connectivity and, more broadly, with the literature on the economic effect of railways (Donaldson, 2018; Berger and Enflo, 2017; Banerjee, Duflo, and Qian, 2012), in particular those contributions

that have looked at the impact of colonial railroads in African countries (Jedwab and Moradi, 2016; Jedwab, Kerby, and Moradi, 2017).

More specifically, this paper complements those studies that have highlighted a long-run persistence of a high density equilibrium in proximity of railway lines, but not significantly faster economic growth several decades after construction (Berger and Enflo, 2017; Banerjee, Duflo, and Qian, 2012; Attack et al., 2010). I find that increasing economic divergence between first-mover cells and the control group occurred throughout the post-colonial period and continues to this date. This widening of the gap, which appears to be particularly acute between 1990 and 2015, is at odds with the notion that the transportation monopoly enjoyed by the colonial infrastructure evaporated soon after the end of the Italian occupation, when new motorways were built and parts of the colonial network deteriorated. As a progressively larger share of cells in the control group could also benefit from an easier access to markets thanks to the newly built transport infrastructure (while avoiding the congestion costs in first-mover locations), in fact, one might expect the equilibrium to stabilize or even convergence to occur.

Increasing divergence can be tentatively explained by coordination of post-colonial spending that tended to target historically better connected locations. In particular, I provide evidence of a higher concentration of hospitals and schools around 2010 in first-mover cells relative to similarly populated and urbanized areas in the control group, but similar to equally wealthy ones (as measured by light density). Controlling for the location of these facilities explains between 20 and 30% of the baseline effect of paved roads. The importance of post-colonial investments in reinforcing the first-mover advantage is confirmed by the fact that the effect of roads on luminosity is more concentrated in locations where Italian roads were constantly preserved throughout the post-colonial era. Conversely, colonial sunk investments, complementarity between extant and new roads and increasing returns to scale in first-mover locations seem to have played a limited role. In this respect, the findings are reminiscent of the factors coordination mechanism discussed by Bleakley and Lin (2012) and, for transport infrastructure, Jedwab, Kerby, and Moradi (2017). In turn, these results suggest that in the presence of particular conditions, such as an initial monopoly in transportation, which subsequently attracts governmental spending, a first-mover advantage provided by an early construction of motorways can lead to increasing divergence in the concentration of economic

activity in the long-run. Therefore, if governmental spending will start to be allocated more equitably across the area in the future, divergence might decelerate and the equilibrium, eventually, stabilize. By contrast, increasing density might occur if the oversupply of public facilities will continue.

More broadly, this article adds to the debate on the role played by colonialism in shaping contemporary spatial inequality. Scholars have emphasized the role of institutions (Banerjee and Iyer, 2005), taxation (Huillery, 2014), human capital (Cogneau and Moradi, 2014), and railroads (Donaldson, 2018), while little has been said on the role of colonial roads. With a narrower focus, this paper also relates to the Italian colonialism-focused literature, which has often dismissed the importance of Italian colonial investments in general and for the post-colonial economic history of the Horn of Africa, in particular (Labanca, 2002; Pankhurst, 1971).

The remainder of the paper is structured as follows. Section 2 provides a brief description of the historical framework. Sections 3 and 4 illustrate the empirical strategy and the data, respectively. The main results and robustness checks are presented in Sections 5 and 6. Section 7 discusses the mechanisms. Section 8 concludes.

2 Italian road construction in the Horn of Africa

Prior to the Prime Minister Benito Mussolini’s decision to expand the Italian colonial territories through the annexation of the Ethiopian Empire in 1934, the Horn of Africa had seen little investment in transportation. Both Eritrea and Italian Somalia lacked the economic potential to justify extensive road (or railroad) building and, moreover, there was no territorial contiguity between them (see Figure 1a). The Ethiopian monarchs had only managed to construct a handful of short gravel roads around Addis Abeba² (Benti, 2016, p.82).³ This shortage of modern facilities was aggravated by the heavy rainfalls between June and Septem-

²City names use the Italian spelling as reported in the employed primary sources (Piccioli (eds.), 1937–1943)

³The completion of the Djibuti-Addis Abeba railway in 1917 only connected the capital to the foreign port of Djibuti. Ethiopian international trade before 1935 was therefore mainly focused on luxury goods and military items (Shiferaw, 2019). Other regions were left isolated due to the lack of a road network and secondary railway tracks.

ber, which rendered unpaved caravan tracks completely impractical for a significant part of the year. Before the Italian invasion, therefore, wheeled transportation was hardly feasible and shipping costs were prohibitive (Cobolli-Gigli, 1938, pp.39-41).

The Italo-Ethiopian war completely transformed transportation dynamics in the area: between 1934 and 1935, the Italian military engineers started building new road tracks (mostly gravel) that connected the core of Eritrea and Italian Somalia with the Ethiopian border to allow troops and supplies to reach the front and, subsequently, the main Ethiopian cities (Badoglio, 1937; Graziani, 1938). These military facilities were constructed either by improving existing caravan trails (see Figure A.1) or, if dirt roads did not previously exist or their gradient was excessively high, often from scratch. When Addis Abeba was occupied on May 5, 1936, just a small portion of the Ethiopian territory was actually under Italian control. The Italian garrisons patrolled major cities such as Addis Abeba, Gimma, Gondar, Dessié, Harar, and only a few other towns. In the absence of a modern road network, especially during the rainy season, it was very difficult for the Italian army to supply and preserve these strongholds, which were constantly targeted by the intense guerrilla warfare of Ethiopian patriots (Dominioni, 2008). In this precarious situation, the Italian government started the construction of a large road network to connect the main cities through asphalt roads, aiming to protect its supply chain from guerilla warfare in isolated locations (Cobolli-Gigli, 1937, p.1480). Large parts of this new road network were created by paving the existing military tracks built during the invasion. Thus, new roads mostly connected strategically important cities following least cost paths (Emmenegger, 2012, p.16).

The works were carried out with a largesse of means and with surprising speed so that, before the start of the rainy season of 1937, most of the main arteries, such as the Asmara-Gondar, Asmara-Dessié-Addis Abeba, Addis Abeba-Gimma and Addis Abeba-Lechemti roads, were completed. In the following years (1937-1939), new trunks were added to the network, such as the Addis Abeba-Debra Tabor and the Assab-Dessié roads (see Piccioli (eds.) (1937–1943)). Other major arteries were started but left unfinished (left unpaved, that is) because of the drastic cut in the budget due to the World War II preparations, and, ultimately the outbreak of the war (Del Boca, 1986). As shown in Figures 1c and A.1, by 1940, these major roads formed an extensive network of about 7,000 km. Of this, roughly 3,500 km were paved

(Benti, 2016, p.94). In the empirical analysis, I focus on the effect of these paved roads.

The new asphalt motorways led to drastic reduction in transport costs. For instance, the cost of shipping 100 kg of goods from the port of Massaua to Addis Abeba declined from 500 Italian Lire (LIT) in 1936-7, before the completion of the pavement, down to 120 LIT in 1939, a decrease of roughly 75% (Cecini, 2007, p.145). This figure, however, might significantly underestimate the actual decline that occurred. The 500 LIT/cwt, in fact, refers to cost of shipping on the improved military gravel road that was available in 1937, after the war. The transport costs along the older dirt roads would have been at least twice as large. Moreover, the creation of the new Assab-Dessié motorway led to further cost reduction, down to 60 LIT/cwt for shipping from the coast to the capital, thanks to the shorter length of the route (Cecini, 2007, p.145). Thus, the actual cost of moving goods from the Eritrean coast to Addis Abeba might have declined as much as 95% within 4 years. A reduction in the same order of magnitude (between 75% and 95%) arguably benefited all areas connected by the network and lends strong support to the idea that the new Italian infrastructure lifted entire regions out of virtual autarky.

The Italian rule did not last for long. In 1941, the last Italian strongholds surrendered to the combined British and Ethiopian forces. During the war, parts of the network's surface, major bridges, and tunnels were destroyed (Del Boca, 1986). The bulk of the colonial road network survived, but no maintenance or new construction plans were undertaken until 1951, due to a lack of capital and economic stagnation. From 1951, instead, several infrastructure-improvement plans were implemented: initially (during the 1950s), parts of the Italian roads were refurbished, while from the late 1950s, new construction works started. Eventually, the virtual monopoly in transportation that Italian roads had enjoyed during the colonial era and throughout the 1950s waned due to the construction of a number of new roads, poor maintenance of Italian routes, and the deterioration of parts of the network.⁴ The scale of post-colonial road building, in particular, was substantial, and especially so from the 1990s. By 2008, the Ethiopian road network alone totaled more than 110,000 km of roads, including 20,000 km of federal highways, which constitutes more than a 10-fold increase from the colo-

⁴Somalia and Eritrea became independent countries and the parts of the network that connected them to Ethiopia fell out of use for periods of time. See Figure A.1.

nial times (Emmenegger, 2012, p.9).

From this brief historical framework, three key elements emerge. First, the construction of colonial roads led to a dramatic fall in transport costs and lifted large parts of the area out of virtual isolation. Second, Italian motorways were designed to serve contingent strategic needs. As they were modeled based on military tracks whose purpose was to connect important strategic targets following least cost routes, construction between termini can be plausibly considered random. Third, Italian roads remained largely operational in the post-colonial period as they were maintained and extended. However, they progressively lost their relative advantage in transportation, due to the construction of new arteries and the decay of parts of the original network. These characteristics make the Horn of Africa an ideal setting to study the effect of a first-mover advantage in transportation on the distribution of economic activity in a developing country and over the long run.

3 Empirical strategy

The first research question that this paper considers is whether the Italian road network had a long-term impact on the concentration of economic activity of Eritrea, Italian Somalia, and Ethiopia. I start by running regression analysis on a spatially explicit gridded dataset, composed of 15,288 non-nodal cells, each measuring 0.1x0.1 degrees (roughly 11x11 km, see Figure 1c). In the baseline regression, the grid excludes islands and all cells located within 20 km of the network’s termini. For each observation, I know the average population density in 2000 and luminosity in 2013, together with the Euclidean distance of each cell’s centroid from the closest Italian motorway in 1939 and city with more than 10,000 inhabitants in 2015. I also add various pre-colonial, colonial, environmental, and geographical controls. I estimate Equation 1 through OLS:

$$\begin{aligned}
y_{i,e} = & \alpha + \beta_1 DC_{ColonialRoad0-10km,i,e} \\
& + \beta_2 DC_{ColonialRoad10-20km,i,e} \\
& + \delta_e + x'_{i,e} \gamma + \epsilon_{i,e}
\end{aligned} \tag{1}$$

$y_{i,e}$ represents the two main dependent variables, population density in 2000 (normalized by standard deviations) and luminosity in 2010 (in logarithm), both measured in cell i and

ethnic group e . I also provide results from a probabilistic model, similar to Equation 1, but using an indicator for urban agglomeration as dependent variable. This is a dummy variable that equals one if a cell's centroid is located within 20 km of a cities with more than 10,000 inhabitants.

$DColonialRoad_{0-10km,i,e}$ and $DColonialRoad_{10-20km,i,e}$ are the explanatory variables of interest, two dummies that equal one if cell i in ethnic group e is located within a 10 km and a 10 to 20 km radius of colonial paved roads, respectively. If the Italian motorways have affected economic concentration and income distribution persistently, I expect colonial roads to show positive and statistically significant coefficients. ($\beta_1, \beta_2 > 0$).

δ_e is the baseline set of fixed effects, the 53 ethnic groups in which Oriental Italian Africa was divided, according to Murdock (1967). Ethnic fixed effects are common in the literature on African development and account for unobservables at the ethnic group level, which is particularly important given the extensive empirical evidence on the long-lasting association between ethnic characteristics and long-term development (Michalopoulos and Papaioannou, 2013). A cell is assigned to an ethnic group if its centroid falls within its boundaries. Ethnic boundaries do not overlap in Murdock, so each cell can only belong to one group.

$x'_{i,c}$ is a vector representing the full set of controls. These can be divided into three groups, namely pre-colonial, colonial and geographical controls. Pre-colonial ones include dummies for the centroids being located within 10 km of, or between 10 and 20 km away from 1935 tracks and population density in 1930. As better access to telecommunications could improve total factor productivity (TFP) and growth potential (Wong, 2004), I also include a dummy for being located within 10 km of telephonic and telegraphic lines that existed prior to the outbreak of the war.⁵ Controls for colonial investments include zero-to-10 km distance cut-offs from other facilities of the colonial era: railroads, primary and secondary hospitals, schools for locals, colonial provincial capitals, Italian unpaved tracks, and provincial administrative capitals. I also include a dummy for being in an area of mining interest during the colonial era. Additionally, I introduce a standard set of geographical controls for the following: lati-

⁵In all regressions with dummies indicating presence of cities after 1936 as dependent variable, I also control for cities in 1935.

tude, longitude, altitude, annual average rainfall, FAO land suitability, being located within 10 km of perennial rivers, and less than 30 km from the coast. Finally, I include malaria rates, ruggedness, and temperature. Table 1 reports descriptive statistics for the variables included in the baseline model.

$\epsilon_{i,e}$ is the error term. To make estimates robust to heteroskedasticity and spatial correlation, I cluster the standard errors by 202 colonial provinces (“Residenze”) in all regressions. Conley errors are also calculated with large distance cut-offs for the baseline specifications.

Two assumptions are needed for the OLS to provide unbiased estimates and thus inform about the long-term effect of Italian roads. First, allocation between targets must be orthogonal to the underlying characteristics of the connected areas, in line with the inconsequential unit approach. Second, paved roads must not have strong spillover effects in neighboring cells. Violations of either of these assumptions could equally lead to downward or upward bias. In the remainder of the paper I relax both assumptions. First, I implement an IV estimation based on Least Cost Paths between targets and a placebo test based on unfinished roads in Section 6. Second, in Section 7.1, I include additional cut-off dummies to the model to capture potential spillovers beyond the 20 km range, thus effectively restricting the control group to cells located more than 60 km away.

4 Data

The described analysis required a novel dataset containing information on a vast array of dependent, historical and geographical variables. In this section, I focus on the main dependent and explanatory variables: population density, agglomeration data and Italian colonial investments, my main dependent and explanatory variables. Appendix 10.2 provides a detailed description of each variable and of the data sources.

The main dependent variable, employed for the baseline estimation, is decennial population density for the 1920-2000 period. For the years 1960, 1970, 1980, 1990 and 2000 (last available), this is obtainable through the United Nations Environment Programme / Global Resource Information Database (UNEP/GRID)-Sioux Falls dataset in raster format. I cal-

culate the average value of the raster cells' centroids that fall into a grid-cell i . This source distributes district-level population across a raster surface by looking at the variation of an “accessibility index” that incorporates data on agglomeration and cost distances between cells. As the variation primarily comes from micro-level statistics, this methodology offers fairly reliable information on population distribution. However, as assumptions are needed to project the available population statistics at the raster cell's level (and transport infrastructure is incorporated in the cost-distance analysis), it is of paramount importance to check the robustness of the estimates with alternative measures of economic activity, such as light density, urbanization, and living standards. As all grid cells have the same size, after the population density data are aggregated at the grid level, differences in density closely approximate differences in the population count. The unit of measurement is people per square km. Figure 1c shows the distribution of population in 2000 across the sample area.

For the years 1920, 1930, 1940 and 1950, micro-level data of population density do not exist. As a first step to solve this issue, I employ reconstructed population densities from the History Database of Global Environment (HYDE), which combines historical population statistics with cropland and pasture information to reconstruct population distribution in raster format (Klein Goldewijk et al., 2011). This source gives an approximation of varying quality on the population distribution across the area before 1960. Relatively precise reconstruction is possible for Eritrea and Italian Somalia, for which more reliable historical statistics are available. By contrast, the central regions of Ethiopia are more problematic, as very limited availability of population statistics in the pre-colonial and colonial period exists and the reconstructed data are marred by spatial correlation.

To distinguish between urban agglomeration and population density, and also to address the limited reliability of the information from HYDE, I back up the available population density information with the location of villages, towns, and cities from cartographic material. A series of maps edited by Achille Dardano, the director of the Cartographic Office of the Ministry of the Colonies, allowed me to obtain information on the location of all villages (100 to 4,000 inhabitants), towns (4,000 to 10,000 inhabitants), and cities (more than 10,000

inhabitants) in Eritrea, Italian Somalia, and Ethiopia for the years 1925⁶, 1935⁷, 1938⁸, and 1939.⁹ For Ethiopia alone, I was also able extend the data on cities' location to the years 1968¹⁰, 1981¹¹, 1988¹², 2000, 2010 and 2015¹³. The location of cities in 2015 is displayed in Figure 1d. For villages and towns between 1925 and 1939, I compute binary indicators that equal one if a cell contained a small agglomeration. For cities between 1925 and 2015, I calculate the distance between each cell's centroid and the closest city and then create binary indicators for being located within a 20 km cut-off of the city's centroid.

Information on colonial investments is mainly from “Gli Annali dell’Africa Italiana”, a review edited by the Italian Ministry of the Colonies, that describes the public investments implemented by the Italian Government in Oriental Italian Africa. The location of colonial paved and unpaved roads in 1939, which is used to defined treatment, control and placebo cells, comes from the map *Carta progressiva delle costruzioni stradali*, contained in the fourth issue of the 1939 edition. This source details the road sections that had been completed by 1939 and distinguishes the road surface quality between paved (asphalt) or unpaved (gravel). Figures 1c and 1b show the extension of the Italian road network as well as the different types of roads. I also collect data on the location of Italian hospitals, clinics and schools for the indigenous population by geo-referencing a series of maps from the first 1940 issue. The distribution of other investments from the colonial period, such as the location of colonial railroads, administrative capitals, boundaries and mining areas, come from diverse sources, as reported in Appendix 10.2.

Pre-colonial tracks come from Achille Dardano’s 1935 map¹⁴, to the best of my knowledge, the most accurate map ever drawn for pre-1936 Ethiopia. I calculate linear distances between each cell's centroid and the closest unpaved track and then define binary indicators for 0-10 km and 10-20 km distance cut-offs. Finally, a cross-section of pre-colonial ethnic boundaries

⁶ Achille Dardano (1925). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico

⁷ Achille Dardano (1935). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico

⁸ Consociazione Turistica Italiana (CTI) (1938?). *Africa Orientale Italiana (6 maps)*

⁹ Achille Dardano (1939). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico

¹⁰ Union of Soviet Socialist Republics (USSR) (1968). *Ethiopia (Map 1:2,500,000)*. Cartography Department - URSS Ministry Committee

¹¹ Ethiopian Mapping Agency (1981). *National Atlas of Ethiopia (preliminary version)*

¹² Ethiopian Mapping Authority (1988). *National Atlas of Ethiopia*

¹³ Available at <https://www.africapolis.org/data>, last accessed on 10th of November 2019

¹⁴ Achille Dardano (1935). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico

is obtained from Murdock (1967). Each cell is assigned to the ethnic boundary that fully contains its centroid and can, by definition, belong to one ethnic group only, similar to colonial administrative boundaries and mining areas.

5 Main results

5.1 The long-run effect of colonial roads

Are historically better-connected locations richer today relative to cells with similar characteristics in the control group? In order to answer the core question of this paper, I begin by estimating Equation 1 through OLS. Table 2 reports the estimates from the main specification: standardized population density (*z-scores*) in 2000 is employed as dependent variable in columns 1 to 4, whereas the results for log light density at night in 2013 are displayed in columns 5 and 6. Column 7 reports the marginal effects from a probabilistic model, with a binary variable for being located within 20 km of a city with more than 10,000 inhabitants in 2015 as dependent variable. Ethnic fixed effects are included in all specifications, while standard errors are clustered by the 202 colonial provinces. Network’s targets (cells within 20 km of termini) are excluded in all columns.¹⁵ Column 1 has no controls, column 2 includes pre-colonial factors, while column 3 adds colonial ones. Columns 4, 5 and 7 report the estimates from the fully restricted model, that includes geographical controls, for the three dependent variables. Column 6 makes the estimates for light density conditional on population in 2000.

As column 4 shows, cells located within 10 km of colonial paved roads are, on average, 0.198 standard deviations more densely populated, once the full set of controls has been included. This effect is economically very large, as it implies a statistically significant deviation from the mean of about 20% of the standard deviation, or 15 points. This result implies that treated locations have an average population density of about 55 inhabitants sq/km, compared to the average, for the whole area, of roughly 40 (see Table 1). The estimates are robust to the introduction of the different sets of controls in columns 2 to 4. Compared to first-wave colonial infrastructure, the effect of second-wave motorways, that were built between 1941 and 1990, is dismal. Table A.9 reveals how, after controlling for population density before

¹⁵Columns 2, 3 and 6 of Table A.1 report the same estimation with the full sample, without the baseline set of fixed effects and directly controlling for the distance from targets, respectively.

construction, the coefficient of newly built, post-colonial paved roads on population density in 2000 is between 95% and 75% smaller than the one of Italian motorways.

Population density is a good approximation if one wants to measure the concentration of economic activity but, especially in developing countries, it leaves room for uncertainty when it comes to disentangling density, productivity and living standards. The replication of the previous exercise with luminosity and location of urban centers relaxes these constraints, and makes it possible to say more about actual levels of development and welfare. Column 5 shows how, conditional on the full set of controls, cells located within 10 km of Italian motorways are, on average, about 62%¹⁶ more luminous than the control group, thus suggesting higher GDP levels (Henderson, Storeygard, and Weil, 2012). Once contemporary population density is added to the equation as a control in column 6, the coefficient of interest remains positive and significant at the 1% level, albeit a bit smaller at roughly 51%. This is a key finding as, by comparing treated and untreated locations with similar population, I provide a rough estimate of GDP per capita differences. In other words, not only do early-connected cells show a higher concentration of economic activity, but they are also significantly richer relative to the control group. Finally, column 7 gives convincing evidence that historically better-connected locations in Ethiopia are also more urbanized today. In fact, conditional on the location of cities in 1935, treated cells are about 20% more likely to be located within 20 km of a city with more than 10,000 inhabitants. Cells in the 10-20 km range, display a positive association with the different outcomes. This correlation, however, is not robust across all specifications.

Population density, luminosity, and urbanization are good proxies for economic concentration and productivity. Nevertheless, one might still worry that the estimates do not reflect higher welfare. In particular, the luminosity data might suffer from significant measurement error and urbanization in developing countries might not necessarily be associated with productivity gains if cities are mainly consumption outlets affected by overpopulation (Gollin, Jedwab, and Vollrath, 2016). Therefore, I also test my working hypothesis using the welfare index from the 2011 male individual Demographic and Health Survey of Ethiopia, a variable that provides a good proxy for living standards. In Table 3, I re-estimate Equation 1, this time looking at 9,669 Ethiopian males living more than 20 km away of the network's targets.

¹⁶ $(exp\beta - 1) * 100$

Consistent with the results from Table 2, the data reveal how, conditional on the full set of controls, Ethiopian males in first-mover areas have a 10% higher probability of being above the sample mean’s wealth index, relative to residents in villages with similar characteristics but located farther away from Italian roads (column 4).¹⁷

In conclusion, Tables 2 and 3 provide evidence that historically better-connected cells are not only more populous but also significantly richer, as shown by proxies of GDP, urbanization and income. While it is difficult to compare the magnitude of these results to other studies given the differences in units of observations, employed dependent variables and specification choices, my results suggest a relatively strong effect, albeit highly localized. For instance, Banerjee, Duflo, and Qian (2012, p.26) estimate an average 19% reduction in GDP when moving between Chinese municipalities in the 25th and the 75th percentile of distance from railroad lines, which they benchmark against a 242% average GDP growth over the period of their analysis (roughly 20 years). For the Horn of Africa, the estimates that employ light density data suggest a 51% increase in per capita GDP when switching from the control to the treatment group (column 6 from Table 2), relative to an average GDP growth for the 1992-2013 period of 174%, in non-nodal cells. Albeit very localized, the gap is much larger in relative terms, and masks a staggering difference in total growth rates between treated and untreated locations in the Horn of Africa, which grew 910% and 140% respectively, over this 20-year window.¹⁸

5.2 The effect over time

One fundamental difference between this article and the related literature on the effect of motorways is that this case study allows to trace the evolution of the effect of paved roads over a 80-year period. Figure 2 focuses on Ethiopia, the only country for which I have been able to obtain relatively consistent data on density and agglomeration over the entire period of the study. Panel A plots the coefficients from Equation 1 with decadal population density, location of cities with more than 10,000 inhabitants, villages and luminosity as dependent variables, for the available years between 1920 and 2015.

¹⁷Panel B from Table 5 shows that individuals in treated locations also have better chances of being employed in higher-paying, third-sector jobs and are less likely to work in agriculture.

¹⁸In Ethiopia, the largest and fastest-growing economy in the area the gap is even larger. Here, the coefficient shows a 82% GDP increase when switching from control to treatment (column 4, panel A, Table 5), relative to an average GDP growth in non-nodal cells of 220% between 1992 and 2013.

By looking at population density estimates, one can see how the effect emerges rapidly in the wake of the occupation and it is already evident in 1940. This immediate concentration of economic activity is consistent with the available qualitative evidence, which portrays rapid agglomeration alongside main roads in the form, for instance, of barracks for workers, grocery stores and gas stations (Cobolli-Gigli, 1938; Benti, 2016). As the quality of reconstructed population density data from HYDE is poor, I use newly collected data on the location of all villages and small towns in 1925, 1935, 1938 and 1939 to verify that the jump in the coefficient between 1930 and 1940 is not driven by data back-projection. The marginal effects for villages (small clusters below 10,000 inhabitants) show a story consistent with population density estimates and confirm strong population growth in treated locations immediately after construction.

The effect on population density remains stable between 1940 and 1950, a decade of financial hardship for the independent Ethiopian government, which was unable to implement any relevant infrastructural development program and only peaks in 1960.¹⁹ The subsequent jump in population concentration during the 1950s coincided with the first period of moderate growth (averaging 3% per annum) for the Ethiopian economy (Shiferaw, 1995) and with the “First Highway Plan”. Launched by the Imperial Ethiopian government between 1951 and 1957, the plan largely focused on refurbishing the existing Italian network (Emmenegger, 2012). The effect possibly moderately declined between 1960 and 1980, during a period that saw the construction of a number of new motorways and the deterioration of parts of the colonial network. As reported in Table A.7, by 1973, 17% of the cells in the control group had gained access to a new motorway (roughly 600 cells), while more than 50% of treated cells (roughly 300 cells) had lost access to paved roads by 1981, due to a significant deterioration of the asphalt surface (see also Figure A.1). In other words, the Italian network had arguably lost its monopoly in transportation by 1980. After that date and during a period of major conflicts and political uncertainty (Eritrean War of Independence and the Ethiopian Civil War) the effect increased by about 25% over the 20-year period, reaching 0.25 standard deviations in 2000, up from 0.2 in 1980.

¹⁹This increase, however, should only be interpreted as indicative as the data sources for 1950 and 1960 are different and the region of Shewa is only included in the sample from 1960 onward due to missing data in the previous years.

A different pattern emerges from the marginal effects of proximity to Italian roads on the location of Ethiopian cities with more than 10,000 residents. Treated locations only start showing higher urbanization from the late 1960s, roughly 20 years after the end of the colonial occupation. Most importantly, the size of the coefficients grows sharply thereafter, reaching a 20% higher probability of finding a city in treated locations in 2015, well after the Italian infrastructure had lost its monopoly in transportation. The process of faster urbanization is mirrored by the rapid increase in light density between 1992 and 2013, which suggests significantly faster GDP growth in first-mover locations. The described patterns hold when restricting to cells located within 10 km of a road (both gravel and asphalt) in 2011 (see Figure A.3a), which suggests that divergence of treated locations occurred irrespective of the selected control group. The marked increase in urban density and luminosity between 2000 and 2015 suggests that population might have also risen in the past 20 years. Faster urbanization and GDP growth during the last three decades in first-mover locations is also visible in panel b, which reports the coefficient of paved roads for the last available year of the different dependent variables from panel a, this time including controls for their lags. While population density in 1960 largely explains its geographical distribution in 2000, thus suggesting path-dependence, the location of cities and light density in previous decades does not account for contemporary levels.

All together, these results reveal increasing concentration of economic activity in first-mover location after 1980. This is a new and puzzling result. The related literature, in fact, highlights fast urban growth and welfare gains immediately after a fall in transport costs (Atack et al., 2010; Storeygard, 2016) and persistence of higher density equilibria in the long run (Berger and Enflo, 2017), but not differentially faster growth in first-mover locations 80 years after the loss of the monopoly in transportation. Section 7.2 discusses some of the mechanisms that might have reinforced the effect in the long-run.

6 Causality

Despite the consistency of the estimates across different development indicators, unobservables could be driving the positive correlation between proximity to colonial roads and concentration

of economic activity. While supported by the historical setting, in fact, the inconsequential unit approach’s assumption of random allocation between targets might not accurately reflect actual building procedures. To address potential endogeneity in this direction and thus relax the described assumption, I implement an IV and a placebo test based on a least-path network and unfinished roads, in the spirit of the standard identification strategies from the related literature on transportation (Redding and Turner, 2015). Table 4 focuses on population density, the main dependent variable, while Table A.3 replicates the tests for luminosity in 2013 and location of cities in 2015. All together, the robustness tests provide strong evidence that Italian roads caused treated cells to be denser and wealthier today.

6.1 *IV estimation: Least Cost Path (LCP)*

The employed instrumental variable takes advantage of the peculiar characteristics of the Italian road network. As described in Section 2, to accomplish the occupation of Ethiopia rapidly and to secure steady supplies for the urban centers of the newly conquered territory, the Italian army sought to connect major cities as quickly and cheaply as possible by building unpaved tracks that tended to follow least-cost routes. In most cases, the contracted civil engineers that followed found themselves under significant time and budget pressure and simply opted for paving these military tracks. Thus, because of the strategic nature of the army’s facilities, Italian motorways also ended up following least-cost routes closely. In this respect, the Italian case differs from the vast majority of road construction plans that, where possible, tend to connect areas with higher economic potential. I take this feature explicitly into account by instrumenting proximity to colonial roads with distance from least-cost paths (a dummy for being within a 0-10 km range of them), connecting the road network’s targets. The latter are selected based on the targeted towns as reported in the Italian construction plans²⁰ and tend to coincide with pre-1935 large cities and ports.

I generate least cost paths based on ruggedness and terrain cover, in combination with information on cost differentials for building paved roads across various terrains as reported in contemporary construction manuals. By using data provided in coeval primary sources, I also adjust these differences to the specific building cost differentials that the Italians faced

²⁰See Cobolli-Gigli (1937). In case a trunk was not completed, the two end-points of the segment are selected as targets.

in the Horn of Africa during the 1930s. Ruggedness and land cover account for a great deal in the cost variation for building new arteries, especially in territories mostly lacking pre-existing facilities. I expect my instrument to be a strong predictor of road location. For the instrument to be valid, conditional on the full set of controls, factors other than proximity to Italian paved roads should not affect long-term economic concentration. Figure 1b shows the least-cost-path network, while section 10.2 from the appendix describes the calculations in detail.

Column 2 from Table 4 reports the first stage of the 2SLS estimation and shows that the selected instrument is, consistently with the historical narrative, a good predictor for proximity to colonial paved roads. The second stage (column 3) shows a positive and statistically significant coefficient of roughly the same magnitude. The Kleibergen-Paap F statistic is large, which in turn indicates that the least-cost network is a strong instrument. The fact that the IV coefficient is about 10% larger than baseline points to a moderate downward bias of the OLS estimates.

6.2 *Placebo: unfinished roads*

The placebo test exploits incomplete roads that were in the process of being paved, but could not be finished due to the outbreak of WWII. This approach is a more convincing falsification exercise than examining both projected paved roads or unpaved tracks alone. That is, projected roads that had not been started by 1939 could simply show lower agglomeration in 2000 due to lower initial potential, which might have been a motivating factor in the delayed start of the works. Similarly, lorry tracks might have been left unpaved deliberately, due to the lower economic potential of the regions they crossed. Plausibly, however, cells in proximity to unfinished roads had characteristics similar to treated ones, as these tracks were also part of the original network, and were left incomplete only due to the diversion of funds from civil engineering to war preparation. As shown in Table A.5, which reports t-tests for observable pre-colonial characteristics, treated and placebo cells were statistically indistinguishable before the construction of Italian network. If anything, placebo ones were characterized by higher land suitability and concentration of cities. Figure 1b shows the placebo lines, while Appendix 10.2 describes the sources that allowed the identification of placebo routes.

Column 4 from Table 4 shows how, compared to the rest of the sample area, placebo cells are not more densely populated in 2000. As the presence of treated cells in the control group might invalidate the placebo exercise, in column 5, I drop observations located within 20 km of paved roads, so to compare placebo cells with untreated locations only. Both coefficients are small and statistically indistinguishable from zero. Figure A.3b shows changes in concentration of economic activity in placebo cells in Ethiopia for the different outcomes and over the entire period of the study. After starting on par with treated locations, placebo cells witnessed a relative reduction in population and light density since the 1950s. Urbanization seems to have increased in the last 15 years, possibly due to the building of new transport facilities. However, placebo coefficients are hardly statistically distinguishable from zero.

6.3 *Additional robustness checks*

Additional robustness checks are reported in Appendix 10.1. Column 5 of Table A.1 shows that the main results for the different outcomes hold if only locations within a 10 km-radius of roads in 2011 are selected. This is particularly important for population density, as access to transport infrastructure is a component of the algorithm employed by UNEP to create raster files from population statistics. Columns 6 and 7 show that the results are robust to controlling for distance from targets and excluding arid cells (below the median of annual rainfall), respectively. The estimates are also robust to an array of additional specifications with varying sets of fixed effects and restrictions to the control group (see Tables A.1, A.2 and A.3). Finally, Table A.4 replicates the estimates reported in Table 2 using 30x30 km grid cells, and shows that the results are not sensitive to the size of the artificial units of observation.

7 Mechanisms and discussion

The results presented so far show that the construction of Italian roads and the related first-mover advantage have generated a persistent gap and an increasing divergence in the concentration of economic activity between the treatment and the control group. This section focuses on Ethiopia alone, for which I have more complete data, and discusses the channels at work. A thorough study of the mechanisms over a 80-year period would require a panel approach, rather than the dynamic cross-sectional one employed in this paper. As various

limitations in the data prevent the use of such techniques, the conclusions from this section should not be considered as definitive, but rather as suggestive evidence on the mechanisms at play.

7.1 Growth and re-organization

The outlined findings would be consistent with a “gains from falling barriers to trade” explanation *à la* Eaton and Kortum (2002). In their framework, a move from autarky to openness increases specialization and wages, reduces the price of consumer and intermediate goods, thus allowing welfare gains along the lines of Ricardian comparative advantage. While not all these variables are observed in this context, higher GDP (as proxied by light density), real wages (as proxied by the DHS’ welfare index) and urbanization point in this direction. One crucial issue with this interpretation, however, relates to the general equilibrium effects. The estimated coefficients could reflect re-organization of economic activity from neighboring to connected cells, rather than net growth from specialization (Redding and Turner, 2015).

Several related contributions have highlighted zero net growth in coincidence with large infrastructural investments, due to the prevailing pattern of relocation of the existing economic activity (Chandra and Thompson, 2000; Buchel and Kyburz, 2020). Investments in transportation, however, are characterized by decreasing marginal returns, with the largest advantages being associated with the first investments due to higher relative gains in market access (Fernald, 1999; Jedwab and Moradi, 2016). The Italian road network certainly constituted the first development push in transport infrastructure across the area and, as discussed in Section 2, led to a dramatic drop in transport costs. In this sense, the Ethiopian case is comparable to railway building in colonial India (Donaldson, 2018), Ghana (Jedwab and Moradi, 2016) and Kenya (Jedwab, Kerby, and Moradi, 2017), where few transport facilities existed before the railroads and net gains from transport investments have been identified.

I start by examining spatial spillovers in adjacent cells for contemporary outcomes. If, over the period of this study, relocation from adjacent locations was an important channel, we would expect to measure lower economic activity in those cells that were not directly connected by Italian transport infrastructure but were closest to treated ones. In Table 5, I operationalize this intuition by introducing controls for several distance cut-offs (20-30 km,

30-40 km, 40-50 km and 50-60 km) and a single stacked cut-off (20-60 km) in separate specifications, thus following Berger and Enflo (2017). Negative and significant coefficients would be indicative of relocation dynamics. The control group in this exercise is composed of more remote locations, which are implicitly assumed not to have been affected by Italian roads located more than 60 km away. Panel A reports the estimates for population density, luminosity and urban location. Panel B shows marginal effects from Probit estimations that look at spatial differences in sectoral employment through binary variables indicating occupation in agriculture, manufacturing and services from the 2011 Ethiopian DHS survey.

All together, the results lend little support to the relocation hypothesis. All cut-off dummies for population and urban density in panel A show small and insignificant coefficients, similarly to the lumped cut-offs in columns 3 and 9. Positive spillovers in the 10 to 20 km cut-off can be observed for both outcomes. A small luminosity drop occurs in the 20-30 km and 50-60 km cut-offs (column 5) and for the lumped 20-60 km bin in column 6. These estimates, however, imply a very small reduction (8% for the lumped coefficient), relative to the 0 to 10 km cut-off (85%).

When looking at differences in sectoral employment in panel B, no sizeable negative effect of road building in the 20 to 60 km range emerges. While residents in first-mover locations are more likely to be employed in services and less so in the agricultural sector, neighboring locations do not show a more backward employment structure relative to remote areas. Interestingly, there is significantly lower concentration of workers in manufacturing in the 10-20 km range, which is consistent with the idea of the secondary sector being more severely affected by relocation dynamics due to its heavier reliance on more costly intermediate goods (Chandra and Thompson, 2000). This finding, however, is more consistent with a core-periphery model (Faber, 2014) than with localized reshuffling, as significantly higher employment in manufacturing is not observed in the 0-10 km cut-off but only in the targets (not shown).

The picture that emerges from contemporary data might mask significant time-series heterogeneity. Strong relocation from adjacent to treated cells might have occurred immediately after construction and this effect might have disappeared in the long run through counterbalancing dynamics, for instance sub-urbanization (Baum-Snow et al., 2017). While evidence in

this direction would not change the overall interpretation of the long-run effects, an initial impoverishment of adjacent cells would make for a more nuanced reading of the results. Table 6 reports the estimates from the same test, this time focusing on the location of villages in 1938 and 1939.²¹ As visible from columns 2 and 3, relocation from nearby cells does not emerge as the main driving force behind early agglomeration. Both the separate cut-off bins (columns 2 and 5) and the stacked one (columns 3 and 6) show small and insignificant coefficients.²²

These results suggest that the effect in 1940 was either linked to the settlement of foreign migrants or, alternatively, to an even migration from across the entire country. The secondary literature on Italian East Africa does lend support to an interpretation based on net population growth between 1936 and 1940 due to the settlement of large numbers of Italian migrants (Del Boca, 1986; Benti, 2016). Although precise micro-level data is lacking, information on the location of Italian schools and branches of the Fascist party allow us to form an idea about the scattering of Italian communities across the Ethiopian territory. The results of this test are reported in columns 7 and 8 of Table 6. Consistent with the hypothesis of foreign immigration explaining part of the increase in population density after 1936, treated cells show a higher likelihood of being located in a district with at least one office of the fascist party or an elementary school for Italian pupils. The size of Italian migration to Ethiopia was certainly driven by the abnormal public spending undertaken by Mussolini’s government during the late 1930s. While this might raise questions about external validity, it also reminds of the clustering of European and Asian settlers along railways lines in Kenya that took place throughout the colonial period (Jedwab, Kerby, and Moradi, 2017) and suggests that large gains in market access might trigger welfare gains sufficient to attract migrants from abroad (Eaton and Kortum, 2002).²³

²¹These are more precise in measuring settlement density than the reconstructed population density data used in Figure 2. In fact, as previously pointed out, the data on villages are more reliable for granular-level analysis of the early post-construction period than reconstructed population densities from HYDE, which are instead marred by spatial correlation.

²²To be precise, some cut-offs dummies do show negative signs, but the coefficients are small (1% probability of seeing a reduction in the settlement in neighboring locations) and statistically insignificant. Moreover, both stacked bins’ coefficients are roughly 10 times smaller than the coefficient of interest (1-4% vs 10-12%). As the reported estimates are based on a binary outcome, these could fail to capture significant variations in village size (rather than disappearance of villages). I replicate the exercise in Panel A of Table 6 by running an ordinal probabilistic regression that captures changes in relative size based on village ranking. In Panel B, I re-run the exercise from Table 6, this time also including 1935 villages as a control to account for path dependence in settlement location.

²³Figure A.2 uses data on cities and population growth to show trends in the 20 to 60 km cut-off area throughout the period of the analysis. While confirming limited relocation between neighboring and treated

All together, the different outcomes portray a picture characterized by net economic growth linked to early infrastructural investments. If relocation played a role in determining the described long-term effect, its importance was marginal. Thus these findings lead to an optimistic assessment on the long-term localized gains from large infrastructural plans. This conclusion, however, should be weighted against the peculiar features of the setting in which Italian investments came into play. The latter was characterized by a virtual lack of modern roads and prohibitive transport costs before 1935. In that, the case study differs from other settings where alternative forms of transport infrastructure were already available. This assessment, therefore, only applies to first-wave infrastructure development and, in particular, to countries at a very early developmental stage. Second-wave motorways, built in Ethiopia during the post-colonial period, do not show a robust and statistically significant association with economic density today (see Table A.9).

7.2 Long-term effect

A second important issue for the interpretation of the results concerns the mechanisms behind the increasing divergence between first-mover locations and the control group, which took place during the post-colonial period. The early increase in market access, linked to the advantage in transportation provided by Italian motorways, in fact, explains the initial increase in population in their proximity. Once a higher density equilibrium has formed, this is likely to persist as long as existing transport facilities continue to operate. Jedwab and Moradi (2016) show how such an equilibrium can survive even after the transportation advantage becomes obsolete, thanks to centripetal agglomeration forces and the lock-in effect of cities (Bleakley and Lin, 2012; Fujita, Krugman, and Venables, 1999). By contrast, the increasing divergence in the concentration of economic activity between treatment and control group, visible in Figure 2, is more difficult to explain.

First, the equilibrium should stabilize in the medium-run as real wages equalize (Redding and Turner, 2015), which is a common finding in the related literature on transportation (Berger and Enflo, 2017; Banerjee, Duflo, and Qian, 2012). Second, catch-up growth and a

cells in the colonial and early post-colonial periods, the estimates also highlight slower population growth during the 1960s, in coincidence with the emergence of the first cities next to Italian roads (see Figure 5.2) and following a decade (the 1950s) that saw sustained population growth in the 0 to 60 km range.

certain degree of convergence might occur between control and treatment, as new cells are given access to asphalt motorways after 1950 and these can also benefit from lower transport costs (Eaton and Kortum, 2002). Third, treated areas were also affected by significant negative shocks after the end of the Italian occupation, namely the deterioration of parts of the network²⁴ and the outflow of Italian immigrants (Del Boca, 1986), which could have also led to convergence by reducing the initial advantage of first-mover locations.²⁵ Treated cells, instead, continued to experience markedly faster growth and increasing concentration of economic activity throughout the post-colonial era. If some degree of convergence can perhaps be observed in the 1960s and 1970s, in fact, the gap between first-mover cells and the control group decisively widened in the past four decades across different indicators.

A model featuring increasing returns to scale and agglomeration effects could provide an explanation of the persistently higher growth rates in first-mover cells, if these had the characteristics of an industrial core (Krugman, 1991). This interpretation, however, does not fit the available empirical evidence. To begin with, treated areas show a very limited concentration of manufacturing activities, which are typically the main beneficiaries of increasing returns. In the Ethiopian case, secondary activities are mainly located in the network’s targets (not shown), while historically better connected locations only feature a higher concentration of service-sector jobs (panel B, Table 5). Following Faber (2014), I test whether treated cells far from the targets and where medium-size towns were located in 1935 show higher levels of luminosity or manufacturing activity today. Results in this direction would be indicative of the existence of a core-periphery structure between first-mover cells, which are relatively more distant from the targets and had a higher pre-1935 market potential, and their hinterland. If the average effect across the entire network masks heterogeneity along these lines, agglomeration forces in more remote locations might be the engine behind the observed divergence. Table A.8, however, shows no evidence in this direction.

In this context, a first explanation might relate to the effect of non-fully depreciated colo-

²⁴This was due a lack of resources for maintenance, and to severe changes in the strategic commercial preferences after decolonization. After Somalia and, eventually, Eritrea became independent countries, connections with their capitals and ports were no longer a priority for the Ethiopian government. Linkages with the ports of Djibouti and Assab were preferred instead. As a consequence, at its lowest point after the communist coup of 1974, only 48% of the Italian road network in Ethiopia still had a paved surface. See Table A.7.

²⁵90% of treated locations is still within 10 km from a paved motorway in 2011, while 17% of the control group is within 10 km of a paved motorway in 2011. See Table A.7.

nial facilities, which I discuss in Section 7.2.1. Second, post-colonial road building might have reinforced the effect of colonial infrastructure, for instance if new roads acted as complements as well as substitutes of extant roads, a possibility that I explore in Section 7.2.2. Third, I examine the hypothesis that colonial infrastructure might have coordinated the allocation of complementary post-colonial investments in Section 7.2.3.

7.2.1 Sunk investments

A first hypothesis that might account for increasing spatial inequality could relate to the effect of sunk investments. If colonial facilities were disproportionately concentrated next to Italian roads and their value has not fully depreciated, first-mover locations might be able to sustain faster economic growth as long as these keep providing a competitive edge to early connected locations. If this is the predominant dynamic, we might expect growth to slow down in the treatment group in the near future.

The Italian government concentrated a wealth of complementary facilities next to the motorways (e.g. schools, hospitals, governmental buildings, unpaved tracks), which likely reinforced the agglomeration effect that roads had through market access. If sunk investments played a predominant role in the recent growth spurt of treated locations, one would expect that a growing share of the effect of Italian roads should be captured by controls for sunk colonial facilities. The top panel of Figure 3 plots the coefficients of roads over time for both the unrestricted and restricted model (Equation 1) with respect to sunk investments, using population density as dependent variable. The bottom panel reports estimates from the Probit model with the binary indicator for presence of cities with more than 10,000 inhabitants as the dependent variable.

The effect of roads on population density stabilizes at around 0.3 standard deviations from 1980 onward, while the relative effect of sunk investments reduces simultaneously to about 5% of the unrestricted estimates. The coefficients for presence of cities reported in the bottom panel show a steady increase in the size of the unrestricted coefficient over time and a simultaneous reduction in the share of the effect, which is explained by controls for colonial sunk investments. This decline is particularly accentuated during the urban take-off starting from the 1980s, which leads to the virtual coincidence of the restricted and unrestricted coefficients

in the 2000s.

While one should be wary of the fact that we might be unable to properly measure the value of the sunk investments (being both unable to account for gaps in the monetary value of different facilities of the same type and by possibly omitting certain types of sunk investments), the described patterns provide convincing evidence against the hypothesis of sunk investment playing a central role in supporting the recent economic take-off of first-mover cells. The small and declining portion of the unrestricted effect that is explained by sunk colonial facilities, in fact, suggests a limited role of sunk investments in explaining the described recent growth spurt in first-mover areas.

7.2.2 Post-colonial maintenance and road building

A second hypothesis relates to the interaction of extant Italian motorways and post-colonial construction works. As it is visible from Figure [A.1](#), there is substantial heterogeneity both in terms of where new motorways were built and which parts of the Italian network were consistently maintained, between 1940 and 2011.

Post-colonial maintenance and construction works might help to explain the observed long-term divergence through two distinct channels. On the one hand, by extending the network, new motorways might have acted as complements (as well as substitutes) of Italian infrastructure, this way increasing the transportation advantage of first-mover locations. On the other hand, targeted maintenance might have also enhanced the value of selected trunks through unobservables, for instance by enlarging the carriageway, improving access roads or co-locating more complementary facilities relative to other motorways (i.e. fuel stations). While we cannot directly control for this possibility, if indicators of the level of preservation during the post-colonial era were to fully capture the effect of Italian roads, we would not be able to exclude the possibility that unobserved improvements alone allowed first-mover locations to maintain a competitive edge throughout the post-colonial period.

Table [7](#) tests these hypotheses by looking at population in 2000, luminosity in 2013 and the location of cities in 2015. Two dummies are interacted with the colonial roads indicator and are introduced in columns 2, 4 and 6, respectively. The dummy “never collapsed” is

a binary indicators that equals one if the Italian road next to a given cell was consistently maintained during the 1940-2011 period. “Extended” takes the value of one if a trunk of the Italian network was either prolonged or intersected by a newly built paved motorway by 2011 (see Figure A.4).

The inclusion of these interaction terms in the baseline model reveals how the effect of Italian roads on the different indicators is relatively homogeneous across the network. In other words, the measured higher density is not limited to continuously maintained parts of the Italian motorways. Interestingly, introducing the maintenance dummy in column 6 reduces both the magnitude and the significance of the paved roads coefficient on light density, which emphasizes the importance of selective refurbishment works in explaining the long-term effect of colonial transport infrastructure on GDP. Even in this case, however, the coefficient of proximity to Italian roads remains significant. Receiving both regular maintenance and road extension does not appear to provide an additional advantage to first-mover locations, irrespective of the selected outcome. The interacted dummy variables, in fact, show negative and insignificant coefficients in all columns. While this test only allows for a very imprecise measurement of the aggregate effect of road extensions (as even the extension of a distant trunk would lead to an increase in market access across the entire network (Donaldson and Hornbeck, 2016)), the extension of a given trunk should benefit locations along this artery more than distant ones.

All together, these results lend little support to the hypothesis that complementarity between colonial and post-colonial motorways played a key role in reinforcing the described long-term effect. The estimates, however, reveal a potentially important role of selective maintenance in explaining the long-term effect of Italian roads on GDP, as supported by the fact that the measured effect is relatively stronger in areas that maintained an asphalt motorways throughout the entire post-colonial period. While it is expected that the continuing operation of colonial infrastructure should lead to better long-term outcomes, it is possible that maintenance also entailed unobserved technical improvements which, in turn, might have contributed to sustain the recent increase in growth rates relative to the control group by increasing the value of first-wave infrastructure. However, albeit reduced in magnitude, the main effect survives this control, which hints to the fact that mechanisms other than

heterogeneity in post-colonial road building might also be in place.

7.2.3 Coordination

As the results relating to selective maintenance suggest, historical infrastructure might have attracted governmental spending on complementary facilities, in turn leading to spatial coordination and thus reinforcing the first-mover advantage in transportation. Coordination might operate through two distinct channels. First, the early positive effect of roads on population concentration could lead the government to target higher-density locations with later expenditures (Jedwab and Moradi, 2016). Alternatively, roads might attract investments directly and lead to an oversupply of facilities. For instance if, in absence of reliable micro-level data, colonial transport facilities signaled location suitability for the allocation of government’s spending, or if governments simply targeted historically more developed areas as a path-dependent policy objective (Jedwab, Kerby, and Moradi, 2017; Davis and Weinstein, 2002).

While colonial infrastructure has not become obsolete, the monopoly in transportation of Italian roads faded in the post-colonial period, so they depreciated in relative terms. In line with Bleakley and Lin (2012), an abnormal density of post-colonial facilities in treated cells, relative to comparably dense ones and conditional on colonial facilities, would point to an oversupply of governmental spending as being a key mechanism in sustaining the increasing divergence between first-mover locations and the control group. By contrast, no difference would indicate the existence of a self-reinforcing higher equilibrium, unrelated to governmental spending.

Table 8 tests this hypothesis, in the spirit of (Jedwab, Kerby, and Moradi, 2017). Conditional on colonial sunk investments and pre-colonial tracks, columns 1, 3 and 5 show higher density of hospitals, schools and feeder roads around 2010 next to Italian motorways. Columns 2, 4 and 6 of panel A make the estimates conditional on population in 1960, when the 2000 higher-density equilibrium crystallized (see Figure 2). Column 6 suggests that feeder roads built in the 2010s targeted historically more populous cells, hence revealing a “investments follow the people” story. Feeder roads might have acted as complements to Italian paved roads, thus facilitating the trade of local agricultural products and accelerating the process of

agglomeration, but they were allocated based on the distribution of the population in 1960. By contrast, columns 2 and 4 suggest that governmental expenditures targeted treated cells more than similarly populated ones in the control group. The same findings emerge when we compare similarly urbanized cells (panel B). Interestingly, panel C shows that controlling for light density in 2010 fully accounts for the distribution of schools and hospitals, which suggests a strong association between concentration of governmental spending and GDP levels across Ethiopia. First-mover cells were already allocated a disproportionate number of governmental facilities in 1981 and 1988, before Ethiopia witnessed the described fast GDP growth during the last two decades (see panel A, Table A.10). Finally, controlling for the complementary facilities from Table 8 in the baseline regressions for Ethiopia (columns 1, 4 and 7 in panel A of Table 5) reduces the coefficient of paved roads between 20 and 30%, depending on the selected outcome.

While one should be wary of the fact that this conclusion is based on very limited data (which only loosely approximate actual differences in the spatial allocation of governmental spending²⁶), these results suggest a central role for post-colonial public investments in reinforcing the first-mover advantage in transportation and thus counteracting the negative effects of the demise of parts of the network, the departure of the Italian migrants, the loss of the monopoly in transportation and congestion costs. The lack of data on the allocation of private capital and the sorting of firms is another major limitation of this study (Ghani, Goswami, and Kerr, 2016). At least until 1991, however, the Ethiopian economy was dominated by the public sector, due to the heavily centralized and planned economic development strategies that characterized the colonial (1936-1941), imperial (1941-1974) and Derg (1974-1991) regimes (Shiferaw, 1995). In this context, public capital likely played a more important role than in comparable settings, especially considering the historically distinct agricultural features of the country's economy and the limited evidence for manufacturing concentration in first-mover locations today (see Table 5).

If an oversupply of governmental facilities was indeed responsible for the recent growth spurt in first-mover locations, we might expect divergence to stop if governmental spending

²⁶Important post-colonial investments in infrastructure are not considered, for instance the construction of the electric grid.

starts being distributed more evenly across the Ethiopian territory. By contrast, first-mover locations might diverge further if higher concentration of public facilities were to continue in the future. A higher density equilibrium, instead, is likely to persist irrespective of the allocation of complementary governmental facilities, in light of the persistence of the first-wave infrastructure and also considering the existence of the denser urban network, which is increasingly likely to take advantage of agglomeration forces.

8 Conclusion

In this paper, I have exploited the quasi-natural experiment provided by Italian road building in the Horn of Africa, to study the long-term effect of a first-mover advantage in transportation on the distribution of economic activity in developing countries. Regression analysis, using a 11x11-km-grid dataset and covering Eritrea, Ethiopia, and Italian Somalia, reveals that cells within a 10 km radius of Italian paved roads are more developed today, both in terms of population concentration and welfare, compared to more distant locations. A least cost path instrumental variable estimation and a falsification exercise confirm that Italian roads rendered treated cells richer 80 years after their construction. Despite losing their advantage in transportation early in post-colonial period (due to the construction of new motorways and the deterioration of parts of the Italian network), first-mover locations witnessed much faster urbanization and GDP growth relative to the control group in the last three decades. Thus, the results emphasize how treated locations have not simply maintained a higher density equilibrium, but have experienced increasing concentration of economic activity throughout the post-colonial era.

The setting offers the possibility to study a transition from virtual autarky (no asphalt roads existed in the area before 1935) to openness. Arguably, the consequent high value of colonial roads is a major driver behind the net economic gains in first-mover areas (the effect is not counterbalanced by an impoverishment of adjacent locations due to a re-shuffling of the existing economic activity) and the large effect on GDP compared to related studies. The results are consistent with welfare gains from falling barriers to trade (Eaton and Kortum, 2002) which, in this case, led to a specialization in the service sector and away from the agricultural one. Today, first-mover locations enjoy a higher concentration of governmental

facilities (hospitals and schools) relative to similarly populated and urbanized ones in the control group. A coordination mechanism that led to an oversupply of governmental spending throughout the post-colonial period, is a plausible explanation for the capacity of treated cells to maintain faster urbanization and GDP growth 70-80 years after construction. Therefore, while a higher-density equilibrium is likely to persist, divergence between first-mover locations and the control group may or may not continue depending on the future allocation of governmental spending.

This article has focused on the relationship between transportation infrastructure and the long-term distribution of economic activity and has found a localized positive association between the two. However, it has not addressed the question of whether transport facilities were allocated optimally and what their aggregate impact was on the regional economy. Such matters remain open and continue to be essential issues in the colonial context. Future research should focus on these aspects.

References

- Aggarwal, Shilpa (2018). “Do rural roads create pathways out of poverty? Evidence from India”. In: *Journal of Development Economics* 133, pp. 375–395.
- Anas (2018). *Listino Prezzi 2018. Nuove Costruzioni Manutenzione Straordinaria*. Tech. rep. URL: https://www.stradeanas.it/sites/default/files/pdf/4.5/NC-MS_LISTINO%20PREZZI%202018.pdf.
- Atack, Jeremy et al. (2010). “Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850–1860”. In: *Social Science History* 34.2, pp. 171–197.
- Badoglio, Pietro (1937). *La guerra d’Etiopia*. Milan: Mondadori.
- Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). “On the road: Access to transportation infrastructure and economic growth in China”. In: *NBER Working Paper 17897*.
- Banerjee, Abhijit and Lakshmi Iyer (2005). “History, institutions and economic performance: the legacy of colonial land tenure systems in India”. In: *American Economic Review* 95.4, pp. 1190–1213.
- Baum-Snow, Nathaniel et al. (2017). “Roads, Railroads, and Decentralization of Chinese Cities”. In: *Review of Economics and Statistics* 99.3, pp. 435–448.
- Benti, Getahun (2016). *Urban growth in Ethiopia, 1887-1974: From the foundation of Finfinnee to the demise of the first Imperial era*. London: Lexington Books.
- Berger, Thor and Kerstin Enflo (2017). “Locomotives of local growth: The short-and long-term impact of railroads in Sweden”. In: *Journal of Urban Economics* 98, pp. 124–138.
- Bleakley, Hoyt and Jeffrey Lin (2012). “Portage and path dependence”. In: *The Quarterly Journal of Economics* 127.2, pp. 587–644.
- Buchel, Konstantin and Stephan Kyburz (2020). “Fast track to growth? Railway access, population growth and local displacement in 19th century Switzerland”. In: *Journal of economic geography* 20.1, pp. 155–195.
- Cecini, Stefano (2007). “La realizzazione della rete stradale in Africa orientale italiana (1936-41)”. In: *Dimensioni e problemi della ricerca storica* 1, pp. 113–156.
- Chandra, Amitabh and Eric Thompson (2000). “Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system”. In: *Regional Science and Urban Economics* 30.4, pp. 457–490.

- Chaves, Isaías N, Stanley L Engerman, and James A Robinson (2014). “Reinventing the wheel: The economic benefits of wheeled transportation in early colonial British West Africa”. In: *Africa’s Development in Historical Perspective*. Ed. by Emmanuel Akyeampong et al. Cambridge University Press. Chap. 10.
- Ciccone, Antonio and Robert E Hall (1996). “Productivity and the density of economic activity”. In: *American Economic Review* 86.1, pp. 54–70.
- Cobolli-Gigli, Giuseppe (1937). “Le strade dell’Impero”. In: *Rassegna Economica dell’Africa Italiana* 10, pp. 1479–1481.
- (1938). *Strade Imperiali*. Milan: Mondadori.
- Cogneau, Denis and Alexander Moradi (2014). “Borders that divide: Education and religion in Ghana and Togo since colonial times”. In: *The Journal of Economic History* 74.3, pp. 694–729.
- Comando Militare Italiano (CMI) (1935). *Impero etiopico - Rete delle trasmissioni (Map 1:2,000,000?)*
- Consociazione Turistica Italiana (CTI) (1938?). *Africa Orientale Italiana (6 maps)*.
- Corni, Guido (1937). *Somalia Italiana (2 volumes)*. Milan: Mondadori.
- Dardano, Achille (1925). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico.
- (1935). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico.
- (1939). *Africa Orientale (Map 1:2,000,000)*. Ministero delle Colonie - Ufficio Cartografico.
- Davis, Donald R and David E Weinstein (2002). “Bones, bombs and break points: the geography of economic activity”. In: *American Economic Review* 92.5, pp. 1269–1289.
- Del Boca, Angelo (1986). *Gli italiani in Africa Orientale. La caduta dell’Impero*. Rome: Laterza.
- Dominioni, Matteo (2008). “Il sistema di occupazione politico-militare dell’Etiopia”. In: *Irsifar, Politiche di occupazione dell’Italia Fascista*. IRSIFAR.
- Donaldson, Dave (2018). “Railroads of the Raj: Estimating the impact of transportation infrastructure”. In: *American Economic Review* 108.4-5, pp. 899–934.
- Donaldson, Dave and Richard Hornbeck (2016). “Railroads and American economic growth: a ‘market access approach’”. In: *The Quarterly Journal of Economics*, pp. 799–858.
- Duranton, Gilles (2015). “Growing through cities in developing countries”. In: *World Bank Research Observer* 30.1, pp. 39–73.

- Eaton, Jonathan and Samuel Kortum (2002). “Technology, Geography, and Trade”. In: *Econometrica* 70.5, pp. 1741–1779.
- Emmenegger, Rony (2012). “The Roads of Decentralisation: The History of Rural Road Construction in Ethiopia”. In: *NCCR North-South* 39.
- Ethiopian Mapping Agency (1981). *National Atlas of Ethiopia (preliminary version)*.
- Ethiopian Mapping Authority (1988). *National Atlas of Ethiopia*.
- Faber, Benjamin (2014). “Trade integration, market mize, and industrialization: Evidence from China’s National Trunk Highway System”. In: *Review of Economic Studies* 81.3, pp. 1046–1070.
- Fernald, John G (1999). “Roads to Prosperity? Assessing the Link Between Public Capital and Productivity”. In: *American Economic Review* 89.3, pp. 619–638.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables (1999). *The spatial economy: Cities, regions, and international trade*. Cambridge and London: MIT Press.
- Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr (2016). “Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing”. In: *Economic Journal* 126.591, pp. 317–357.
- Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath (2016). “Urbanization with and without industrialization”. In: *Journal of Economic Growth* 21.1, pp. 35–70.
- Graziani, Rodolfo (1938). *Il fronte sud*. Milan: Mondadori.
- Henderson, J Vernon et al. (2017). “The Global Distribution of Economic Activity: Nature, History, and the Role of Trade”. In: *The Quarterly Journal of Economics* 133.1, pp. 357–406.
- Henderson, Vernon, Adam Storeygard, and David Weil (2012). “Measuring Economic Growth from Outer Space”. In: *The American Economic Review* 102.2, pp. 994–1028.
- Huillery, Elise (2014). “The Black Man’s Burden: The Cost of Colonization of French West Africa”. In: *The Journal of Economic History* 74.1, pp. 1–38.
- Imperial Highway Authority (1955). *Highway Map of Ethiopia (map)*.
- (1962). *Highway Map of Ethiopia (map)*.
- (1973). *Highway Map of Ethiopia (map)*.
- International Travel Maps (2011). *International Travel Maps, Ethiopia*. Vancouver, BC: International Travel Maps.

- Jedwab, Remi, Edward Kerby, and Alexander Moradi (2017). “History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya”. In: *The Economic Journal* 127.603, pp. 1467–1494.
- Jedwab, Remi and Alexander Moradi (2016). “The permanent effects of transportation revolutions in poor countries: Evidence from Africa”. In: *Review of Economics and Statistics* 98.2, pp. 268–284.
- Kiszewski, Anthony et al. (2004). “A global index representing the stability of malaria transmission”. In: *The American journal of tropical medicine and hygiene* 70.5, pp. 486–498.
- Klein Goldewijk, Kees et al. (2011). “The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years”. In: *Global Ecology and Biogeography* 20.1, pp. 73–86.
- Krugman, Paul (1991). “Increasing returns and economic geography”. In: *Journal of Political Economy* 99.3, pp. 483–499.
- Labanca, Nicola (2002). *Oltremare: Storia dell’espansione coloniale italiana*. Bologna: Il Mulino.
- Michalopoulos, Stelios and Elias Papaioannou (2013). “Pre-Colonial Ethnic Institutions and Contemporary African Development”. In: *Econometrica* 81.1, pp. 113–152.
- Murdock, George Peter (1967). *Ethnographic atlas*. Pittsburg: University of Pittsburg Press.
- Nunn, Nathan (2008). “The long-term effects of Africa’s slave trades”. In: *The Quarterly Journal of Economics* 123.1, pp. 139–176.
- Nunn, Nathan and Diego Puga (2012). “Ruggedness: The blessing of bad geography in Africa”. In: *Review of Economics and Statistics* 94.1, pp. 20–36.
- Pankhurst, Richard (1971). “Economic Verdict on the Italian Occupation of Ethiopia (1936–1941)”. In: *Ethiopia Observer* 14.1, pp. 68–82.
- Piccioli (eds.), Angelo (1937–1943). *Gli Annali dell’Africa Italiana*. Milan: Mondadori.
- Podestà, Gian Luca (2004). *Il mito dell’impero: economia, politica e lavoro nelle colonie italiane dell’Africa orientale, 1898-1941*. Torino: Giappichelli.
- Redding, S.J. and M.A. Turner (2015). “Transportation Costs and the Spatial Organization of Economic Activity”. In: *Handbook of Regional and Urban Economics*. Vol. 5. Elsevier B.V., pp. 1339–1398.
- Shiferaw, Bekele (1995). *An economic history of modern Ethiopia*. Dakar: Codesria.

- Shiferaw, Bekele (2019). “Ethiopia’s Transition from a Traditional to a Developing Economy, 1890s–1960s”. In: *The Oxford Handbook of the Ethiopian Economy*. Oxford: Oxford University Press.
- Sillani, Tommaso (1937). *L’Impero (AOI)*. La Rassegna Italiana.
- Stifel, David, Bart Minten, and Bethlehem Koru (2016). “Economic Benefits of Rural Feeder Roads: Evidence from Ethiopia”. In: *The Journal of Development Studies* 52.9, pp. 1335–1356.
- Storeygard, Adam (2016). “Farther on down the road: Transport costs, trade and urban growth in Sub-Saharan Africa”. In: *Review of Economic Studies* 83.3, pp. 1263–1295.
- Union of Soviet Socialist Republics (USSR) (1968). *Ethiopia (Map 1:2,500,000)*. Cartography Department - URSS Ministry Committee.
- United States Department of Agriculture (USDA) (2017). *Cost Estimating Guide for Road Construction*. Tech. rep. URL: http://www.fs.usda.gov/Internet/FSE{_}DOCUMENTS/stelprdb5279284.pdf.
- Wong, Wei-kang (2004). “How good are trade and telephone call traffic in bridging income gaps and TFP gaps ?” In: 64.2, pp. 441–463.
- World Bank (2007). *A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005*. eng. Washington: World Bank.

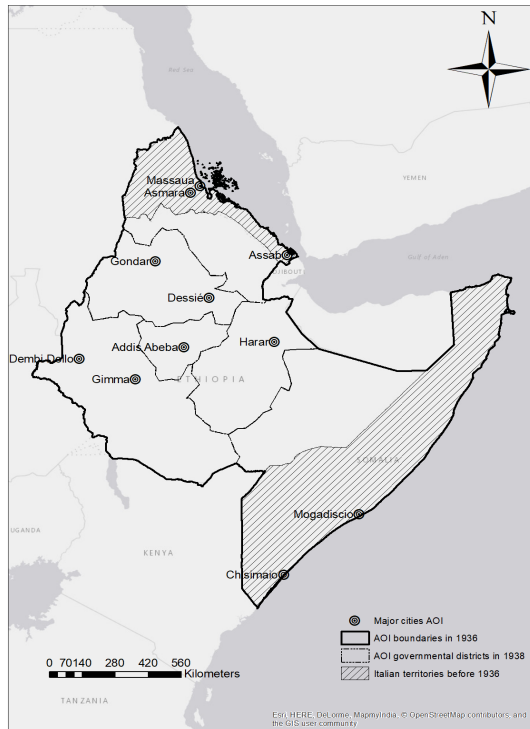
9 Figures and tables

Table 1: Summary Statistics

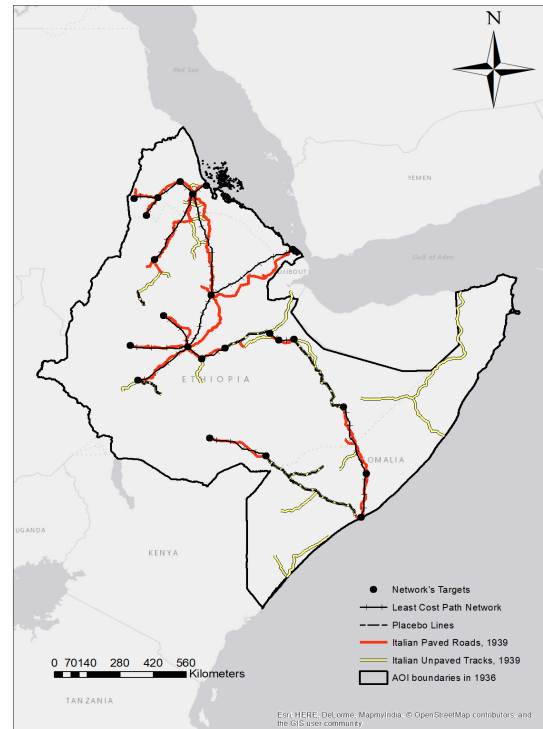
	(1) mean	(2) sd	(3) min	(4) max	(5) N
Population density, 2000	41.06	77.78	0.111	1,623	15,288
Average light density, 2013	0.0414	0.445	0	19.77	15,288
City 2015, 0-20km	0.234	0.423	0	1	15,288
Road 1939, 0-10km	0.0440	0.205	0	1	15,288
Road 1939, 10-20km	0.0451	0.208	0	1	15,288
Track 1935, 0-10km	0.216	0.412	0	1	15,288
Track 1935, 10-20km	0.166	0.372	0	1	15,288
Population density, 1930	10.19	10.23	0	137.7	15,288
Telegraph 1934 CMI, 0-10km	0.0829	0.276	0	1	15,288
Rail 1940, 0-10km	0.00863	0.0925	0	1	15,288
Track 1940, 0-10km	0.0544	0.227	0	1	15,288
District capital 1940, 0-10km	0.0288	0.167	0	1	15,288
Secondary hospital 1940, 0-10km	0.0309	0.173	0	1	15,288
Main hospital 1940, 0-10km	0.00124	0.0352	0	1	15,288
School 1940, 0-10km	0.00922	0.0956	0	1	15,288
Dummy mining 1940	0.142	0.349	0	1	15,288
River, 0-10km	0.318	0.466	0	1	15,288
Coast, 0-10km	0.0117	0.108	0	1	15,288
Rainfall	11.26	11.56	0	95.17	15,288
Malaria index	6.527	5.044	0	31.17	15,288
Altitude	947.5	738.7	-119.5	4,140	15,288
Latitude	8.023	3.821	-0.367	16.67	15,288
Longitude	41.23	4.016	34.53	51.27	15,288
Ruggedness	1,499	1,183	0	6,593	15,288
Temperature	24.45	4.547	0	33.74	15,288
Land Suitability	5.804	1.472	0	9	15,288

Summary statistics of dependent, explanatory and control variables from Equation 1. See Appendix 10.2 for details on individual variables. Similarly to the baseline estimation in Table 2, the sample is restricted to the 15,288 non-nodal locations in Ethiopia, Eritrea and Italian Somalia.

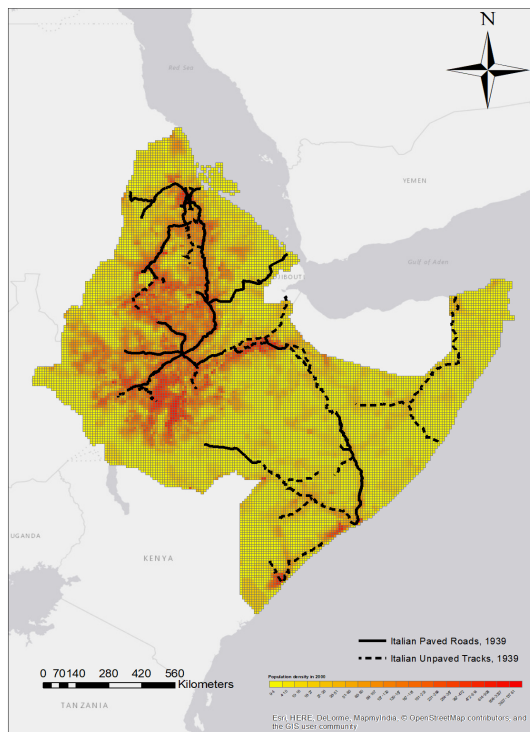
Figure 1: Colonial roads, population density and urbanization



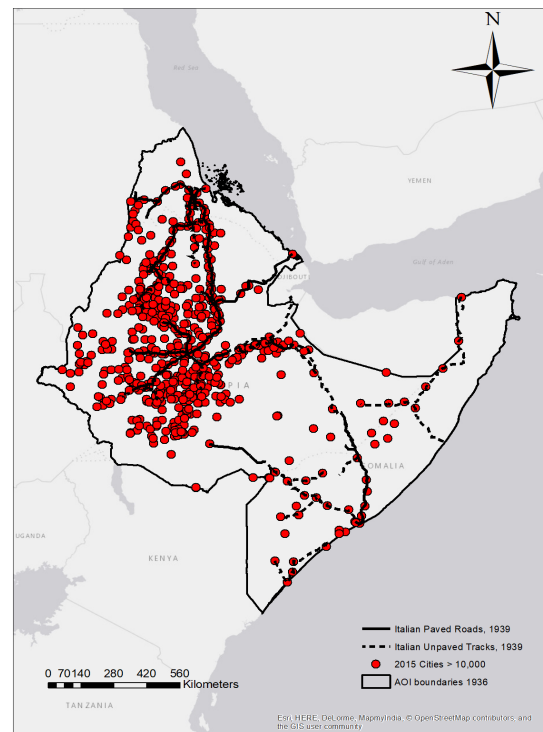
(a) The Horn of Africa in 1934



(b) Italian roads and robustness tests



(c) Population density in 2000



(d) Cities in 2015

Table 2: The long-term effect of colonial roads on economic activity

<i>Dependent variables:</i>	Pop density 2000, z scores				ln luminosity 2013		City 2015, yes/no
	(1)	(2)	(3)	(4)	(5)	(6)	
Road 1939, 0-10km	0.409*** (0.061) [0.067] [0.079]	0.250*** (0.053) [0.062] [0.063]	0.217*** (0.054) [0.065] [0.069]	0.198*** (0.051) [0.063] [0.070]	0.483*** (0.102) [0.136] [0.148]	0.413*** (0.099) [0.126] [0.128]	0.194*** (0.048) [0.053] [0.065]
Road 1939, 10-20km	0.194*** (0.040)	0.090** (0.035)	0.092** (0.035)	0.065** (0.031)	0.006 (0.049)	-0.017 (0.049)	0.107*** (0.031)
Pop density 2000, z scores						0.353*** (0.052)	
Track 1935, 0-10km		0.132*** (0.029)	0.102*** (0.028)	0.101*** (0.026)	0.061** (0.029)	0.026 (0.026)	0.121*** (0.024)
Track 1935, 10-20km		0.075*** (0.018)	0.061*** (0.018)	0.058*** (0.015)	-0.018 (0.015)	-0.039** (0.016)	0.081*** (0.017)
Pop density 1930, z scores		0.456*** (0.052)	0.437*** (0.050)	0.269*** (0.046)	0.206*** (0.068)	0.111* (0.064)	0.120*** (0.031)
Telegraph 1934 CMI, 0-10km		0.083** (0.034)	0.056 (0.035)	0.030 (0.032)	0.001 (0.039)	-0.010 (0.035)	0.035* (0.021)
Rail 1940, 0-10km			-0.066 (0.095)	0.043 (0.086)	0.258 (0.243)	0.243 (0.230)	0.065 (0.059)
Track 1940, 0-10km			0.182*** (0.067)	0.149** (0.060)	0.110** (0.051)	0.058 (0.042)	0.120*** (0.019)
District capital 1940, 0-10km			0.135** (0.058)	0.101* (0.055)	0.624*** (0.097)	0.588*** (0.094)	0.147*** (0.035)
Secondary hospital 1940, 0-10km			0.138** (0.064)	0.126** (0.060)	0.257** (0.100)	0.212** (0.090)	0.116*** (0.038)
Main hospital 1940, 0-10km			0.005 (0.192)	0.061 (0.175)	0.093 (0.448)	0.071 (0.407)	0.118 (0.117)
School 1940, 0-10km			0.189** (0.078)	0.127 (0.078)	0.975*** (0.213)	0.930*** (0.208)	0.195** (0.079)
Observations	15,288	15,288	15,288	15,288	15,288	15,288	14,857
R-squared	0.46	0.50	0.51	0.56	0.19	0.22	
Ethnic FE	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	NO	YES	YES	YES	YES	YES	YES
Ethnic FE	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	NO	YES	YES	YES	YES	YES	YES
Colonial controls	NO	NO	YES	YES	YES	YES	YES
Geographical controls	NO	NO	NO	YES	YES	YES	YES

OLS regression estimates in columns 1 to 6 from the grid dataset (11x11 km) and Equation 1. Column 7 reports marginal effects from a Probit model with the same regressors as in Equation 1. Estimates in column 7 (as in all regressions that employ post-1935 cities as dependent variable) include an additional control for location of cities in 1935 (being within 20 km of city in 1935). All columns exclude the network's targets (cells within a 20 km radius of targets). Dependent variables are population density (*z-scores*) in 2000 (columns 1 to 4), the logarithm of light density at night in 2013 (columns 5 and 6) and a dummy for being located within 20 km of a city with more than 10,000 inhabitants in 2015 (column 7). Robust standard errors, clustered by 202 provincial colonial districts, in parentheses. Squared brackets report Conley standard errors estimated with a 2 decimal degree cut-off (roughly 220 km). Vertical brackets report Conley standard errors estimated with a 5 decimal degree cut-off (roughly 550 km, 1/4 of the sample area). In column 7, Conley errors are estimated with OLS. All columns include fixed effects for 53 ethnic groups from Murdock's map (Nunn, 2008). Column 1 has no controls. Columns 2 and 3 introduce pre-colonial and colonial controls, respectively. Columns 4, 5, 6, and 7 display results with the baseline set of controls (see Equation 1). All pre-colonial and colonial controls are displayed in the table. Geographical controls include altitude, longitude, latitude, average annual rainfall, malaria index, area of mining interest, average annual temperature, crop suitability for low input rain-fed cereals (FAO), ruggedness and dummies for being within 10 km of a perennial river and the coast. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Italian roads and contemporary standards of living

<i>Dependent variables:</i>	Above mean wealth, yes/no			
	(1)	(2)	(3)	(4)
Road 1939, 0-10km	0.139*** (0.040)	0.120*** (0.045)	0.111** (0.047)	0.103** (0.042)
Road 1939, 10-20km	-0.042 (0.061)	-0.042 (0.061)	-0.064* (0.039)	-0.039 (0.033)
Observations	9,669	9,669	9,669	9,669
Ethnic FE	YES	YES	YES	YES
Precolonial controls	NO	YES	YES	YES
Colonial controls	NO	NO	YES	YES
Geographical controls	NO	NO	NO	YES

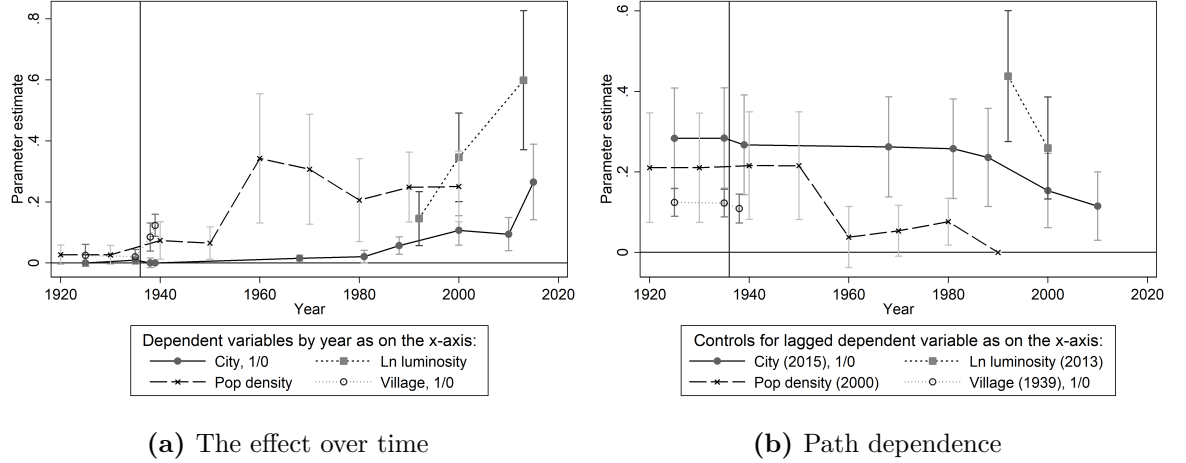
Marginal effects from a Probit model (same specification as Equation 1) employing data from the 2011 male DHS dataset for Ethiopia, in columns 1 to 4. Household's wealth index (mv191) is employed to define the dependent variable. All individuals with a wealth index above the sample's mean (2843.634) are coded as 1. All columns include fixed effects for 53 ethnic groups from Murdock's map (Nunn, 2008). The same set of controls of the baseline estimation (see Table 2 for details) is employed, all columns exclude DHS clusters located within 20km from the network's targets. Robust standard errors, clustered by 202 provincial colonial districts, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Robustness checks: colonial roads on populations density in 2000

<i>Dependent variables:</i>	Pop density 2000, z scores				
	Baseline	First stage	IV LCP 0-10km	Placebo I	Placebo II (No roads 1939)
	(1)	(2)	(3)	(4)	(5)
Road 1939, 0-10km	0.198*** (0.051)		0.233** (0.092)		
Road 1939, 10-20km	0.065** (0.031)	-0.232*** (0.017)	0.072* (0.037)		
0-10km, IV LPC 1940		0.425*** (0.057)			
Proj road 1937 unpav, 0-10km				-0.038 (0.079)	0.015 (0.076)
Proj road 1937 unpav, 10-20km				0.078 (0.064)	0.106 (0.072)
Observations	15,288	15,288	15,288	15,288	13,925
R-squared	0.56	0.38	0.22	0.56	0.54
Cragg-Donald Wald F-Stat	.	.	3141	.	.
Kleibergen-Paap F-Stat	.	.	55.15	.	.
Ethnic FE	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES

All columns include fixed effects for 53 ethnic groups from Murdock's map (Nunn, 2008). Robust standard errors, clustered by 202 provincial colonial districts, are shown in parentheses. The baseline set of controls (see Table 2) is included in all columns. Column 1 reports the baseline estimates from column 4 of Table 2. Column 2 shows the first stage regression, whereas column 3 reports the second stage. A dummy for being between 0 and 10 km from least cost paths connecting the road network's targets is employed as the instrumental variable (columns 2 and 3). Column 4 reports the placebo exercise with projected paved roads that were under construction but were not completed due to the outbreak of WWII. Column 5 is the same as column 4, but excludes those cells that were located within 20 km of Italian paved roads (treatment). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 2: The effect of Italian roads over time



Coefficients for the 0-10 km cut-off distance dummy from colonial paved roads (as in Equation 1) are reported with 95% confidence intervals. Dependent variables as shown in the legend. All estimations include the baseline set of controls and fixed effects^a (see Table 2 for details). When pre-1935 outcomes are employed, HYDE population density in 1900 is used as a control instead of the 1930 one. The employed sample includes all non-nodal cells located in contemporary Ethiopia.^b Vertical lines correspond to 1936. *Panel A*: Reported coefficients for population density and luminosity (Log light density) are OLS estimates. Results for villages are marginal effects from a Probit model, similar to Equation 1, but with dummies for presence of villages in 1925, 1935, 1938 and 1939 as dependent variables. Results for cities above 10,000 inhabitants are marginal effects from a Probit model with the dependent variable being a dummy for being located within a 20 km radius of a city with more than 10,000 inhabitants. *Panel B*: Lagged controls of the dependent variable are included separately as indicated on the x-axis. The employed dependent variables (last year available) are reported in the legend.

^aFor the set of probabilistic estimations with presence of cities as dependent variable, the set of fixed effects is different from Equation 1 and corresponds to the four regions in which the Italian government divided Ethiopia in 1936. For the early years (1925, 1935, 1938 and 1939), in fact, there is not enough variation in the dependent variable to keep ethnic fixed effects in the model.

^bEstimates from HYDE population density for 1920, 1930, 1940 and 1950 exclude the region of Shewa (the surroundings of Addis Abeba) due to missing data.

Table 5: Relocation of economic activity

<i>Panel A</i>									
<i>Dependent variables:</i>	Pop density 2000 z scores			ln luminosity 2013			City 2015, yes/no		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Road 1939, 0-10km	0.251*** (0.059)	0.259*** (0.069)	0.260*** (0.067)	0.599*** (0.115)	0.552*** (0.109)	0.558*** (0.110)	0.271*** (0.068)	0.244*** (0.073)	0.246*** (0.072)
Road 1939, 10-20km	0.081** (0.038)	0.089* (0.052)	0.090* (0.050)	0.027 (0.061)	-0.018 (0.072)	-0.013 (0.069)	0.142*** (0.044)	0.113** (0.053)	0.115** (0.051)
Road 1940, 20-30km		-0.006 (0.045)			-0.121** (0.052)			-0.059 (0.048)	
Road 1940, 30-40km		0.039 (0.044)			-0.049 (0.043)			-0.046 (0.046)	
Road 1940, 40-50km		0.018 (0.033)			-0.055 (0.041)			-0.055 (0.040)	
Road 1940, 50-60km		0.010 (0.027)			-0.049* (0.026)			-0.019 (0.036)	
Road 1940, 20-60km			0.015 (0.032)			-0.067** (0.030)			-0.044 (0.035)
Observations	10,088	10,088	10,088	10,088	10,088	10,088	9,896	9,896	9,896
R-squared	0.56	0.56	0.56	0.19	0.19	0.19	.	.	.
<i>Panel B</i>									
<i>Dependent variables:</i>	Agriculture, yes/no			Manufacturing, yes/no			Services, yes/no		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Road 1939, 0-10km	-0.097* (0.053)	-0.118** (0.056)	-0.105* (0.055)	0.015* (0.008)	0.012 (0.010)	0.011 (0.010)	0.060* (0.031)	0.087*** (0.033)	0.080** (0.032)
Road 1939, 10-20km	0.067* (0.040)	0.050 (0.042)	0.059 (0.043)	-0.023*** (0.008)	-0.024*** (0.009)	-0.027*** (0.009)	-0.030 (0.022)	-0.007 (0.023)	-0.011 (0.024)
Road 1940, 20-30km		-0.023 (0.050)			-0.006 (0.010)			0.029 (0.032)	
Road 1940, 30-40km		-0.055 (0.053)			0.001 (0.010)			0.049 (0.032)	
Road 1940, 40-50km		0.008 (0.059)			0.012 (0.014)			0.007 (0.034)	
Road 1940, 50-60km		0.032 (0.049)			-0.029* (0.016)			0.010 (0.027)	
Road 1940, 20-60km			-0.010 (0.036)			-0.005 (0.007)			0.025 (0.020)
Observations	10,172	10,172	10,172	9,820	9,820	9,820	10,033	10,033	10,033
Ethnic FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES	YES

All estimations include the baseline set of controls and fixed effects as in Table 2. All columns exclude the network's targets and the sample is restricted to Ethiopia. *Panel A*: OLS estimates from the 11x11 km grid dataset with all three dependent variables from Table 2. *Panel B*: Probit estimates reporting marginal effects using the 2011 male DHS dataset. Dummy variables indicate the sector of employment. Robust standard errors, clustered by 202 provincial colonial districts in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Population relocation vs population growth in 1939

<i>Dependent variables:</i>	Village 1939, yes/no			Village 1938, yes/no			Fascist branch (district) yes/no	Italian School (district) yes/no
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road 1939, 0-10km	0.123*** (0.019)	0.117*** (0.022)	0.118*** (0.021)	0.085*** (0.024)	0.073*** (0.026)	0.078*** (0.026)	0.099*** (0.015)	0.144*** (0.052)
Road 1939, 10-20km	0.015 (0.019)	0.009 (0.023)	0.010 (0.023)	-0.005 (0.024)	-0.017 (0.027)	-0.012 (0.026)	0.075*** (0.011)	0.109*** (0.037)
Road 1940, 20-30km		-0.011 (0.023)			-0.041 (0.026)			
Road 1940, 30-40km		-0.011 (0.020)			-0.027 (0.019)			
Road 1940, 40-50km		-0.013 (0.021)			-0.002 (0.023)			
Road 1940, 50-60km		-0.000 (0.017)			0.015 (0.021)			
Road 1940, 20-60km			-0.009 (0.015)			-0.012 (0.016)		
Observations	10,088	10,088	10,088	10,088	10,088	10,088	6,983	10,088
Ethnic FE	YES	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES

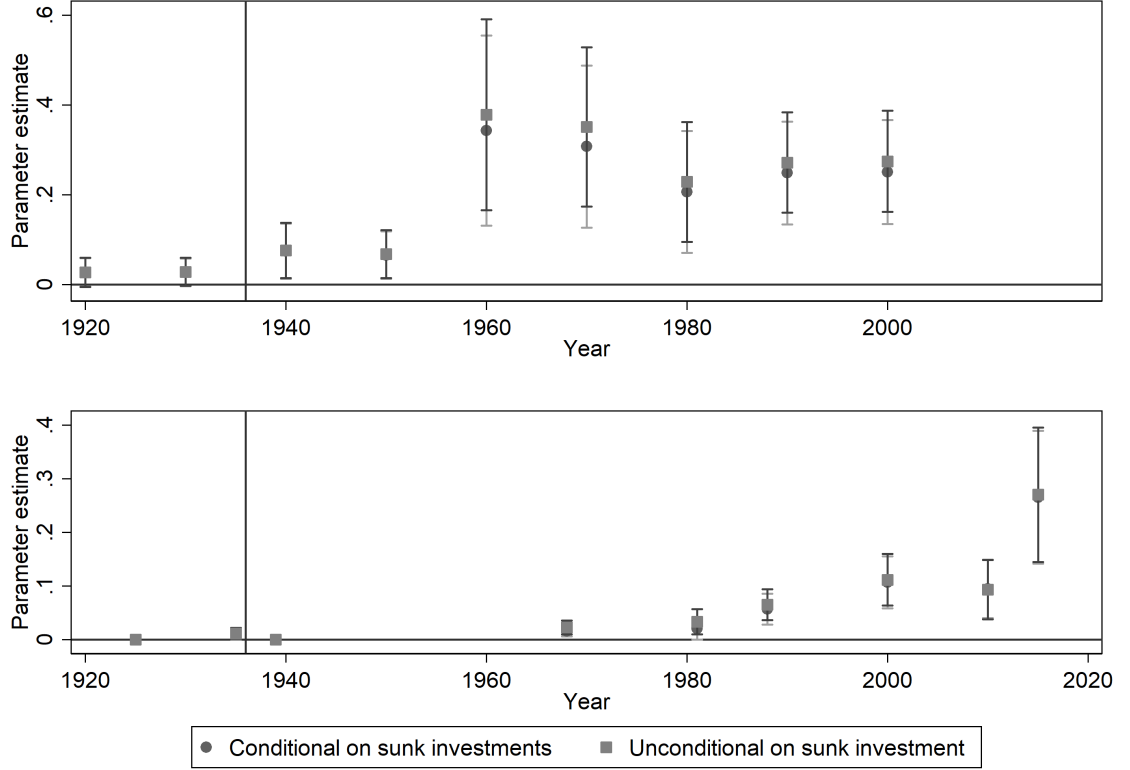
Marginal effects from a Probit model based on Equation 1, from the 11x11 grid dataset. The baseline set of controls is employed in all columns (see Table 2). The sample is restricted to Ethiopia only. The dependent variables are dummies for village presence in 1939 (columns 1 to 3), dummies for village presence in 1938 (columns 4 to 6) and dummies for whether the cell's colonial district (residenza) had an office of the fascist party or a school for Italian settlers in column 7 and 8, respectively. Robust standard errors, clustered by 202 provincial colonial districts, in parentheses. All columns include fixed effects for 53 ethnic groups. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Post-colonial road building

<i>Dependent variables:</i>	Pop density 2000, z scores		ln luminosity 2013		City 2015, yes/no	
	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.251*** (0.059)	0.237*** (0.065)	0.599*** (0.115)	0.177* (0.100)	0.271*** (0.068)	0.237*** (0.064)
Road 1939, 0-10km * Never collapsed, 1940-2011		0.160 (0.117)		1.061*** (0.176)		0.167 (0.140)
Road 1939, 0-10km * Never collapsed * Extended, 2011		-0.184* (0.102)		-0.507** (0.205)		-0.135 (0.137)
Observations	10,088	10,088	10,088	10,088	9,896	9,896
R-squared	0.56	0.56	0.19	0.20	.	.
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES

OLS estimates with population density z-scores in 2000 (columns 1 and 2), the logarithm of light density in 2013 (columns 3 and 4) and a dummy variable for being located within 20 km of a city with more than 10,000 inhabitants in 2015 (columns 5 and 6) as dependent variables. All columns include the baseline set of controls and fixed effects as in Table 2 and the sample is restricted to Ethiopia only. "Never collapsed" is a dummy that equals one if a cell is within 10 km of a trunk of the Italian road network that never lost its pavement between 1940 and 2011. "Extended" is a dummy that takes the value of 1 if a cell is located within 10 km of a trunk of the Italian road network that was extended (either prolonged or intersected) by a new asphalt motorway between 1940 and 2011 and which is still operational in 2011 (see Figure A.4). See Figure A.1 for a description of the evolution of the road network. *** p<0.01, ** p<0.05, * p<0.1

Figure 3: The effect of sunk investments over time



Coefficients for the 0-10 km cut-off distance dummy from colonial paved roads (as in Equation 1) are reported with 95% confidence intervals. Sunk investments are colonial unpaved tracks, railways, schools, hospitals and administrative capitals. Both panels report estimates for the sub-sample of non-nodal cells located in contemporary Ethiopia.^a *Top panel:* OLS estimates (Equation 1) with population density over time as dependent variable. Baseline ethnic fixed effects are employed. *Bottom panel:* marginal effects from a probabilistic model (similar to Equation 1) with a dummy for being located within a 20 km radius of cities (more than 10,000 inhabitants) as dependent variable. Fixed effects for the 4 regional administrative areas in which Italy split Ethiopia during the occupation are employed.

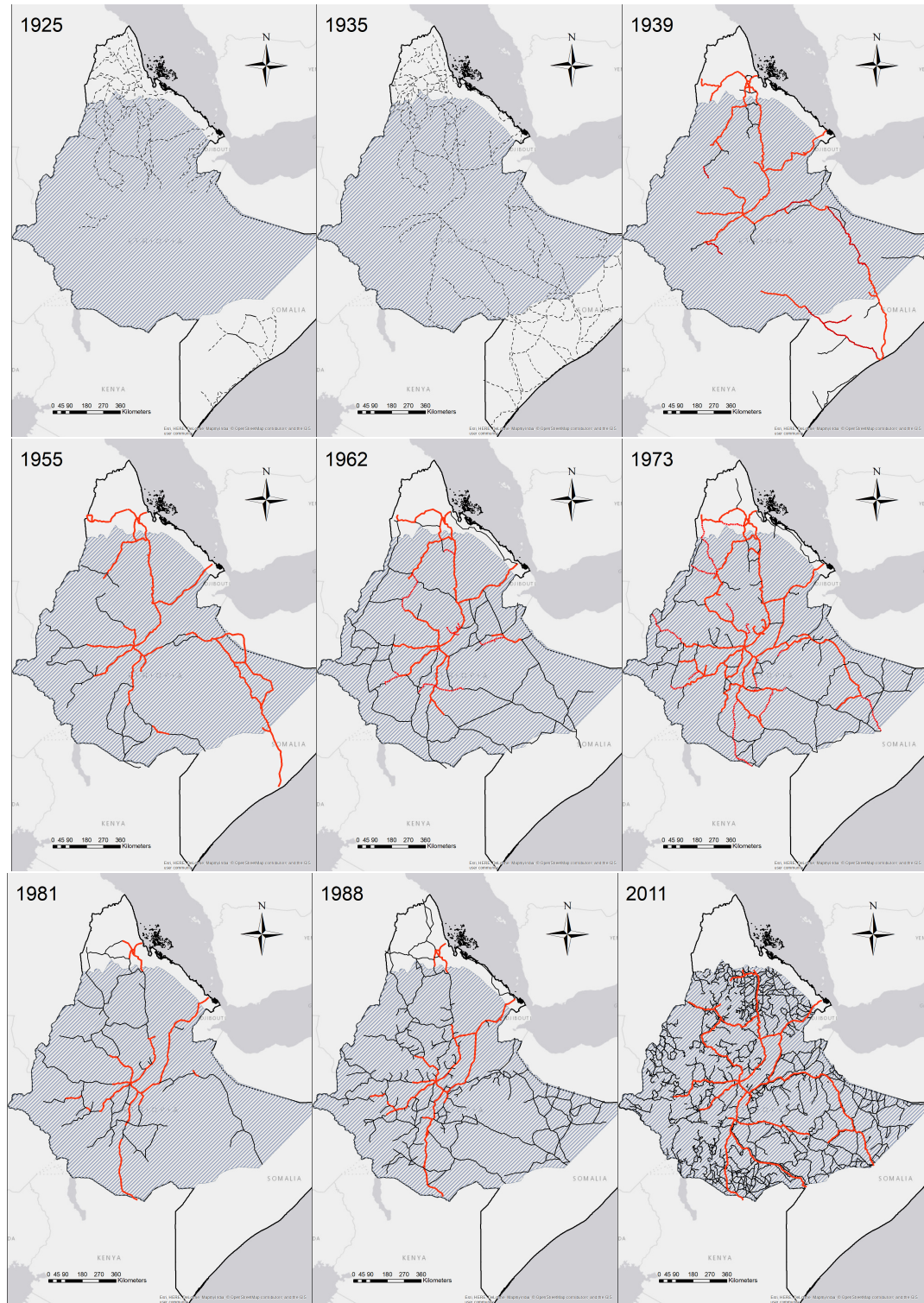
^aEstimates from HYDE population density for 1920, 1930, 1940 and 1950 exclude the region of Shewa (the surroundings of Addis Abeba) due to missing data.

Table 8: Coordination of post-colonial investments

<i>Dependent variables:</i>	School 2010, 0-10km		Hospital 2010, 0-10km		Feeder road 2016, 0-10km	
<i>Panel A</i>	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.021*** (0.006)	0.016*** (0.005)	0.058*** (0.018)	0.035* (0.019)	0.196** (0.087)	0.148 (0.102)
Pop density 1960, z scores		0.013*** (0.002)		0.062*** (0.009)		0.269*** (0.045)
<i>Panel B</i>	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.021*** (0.006)	0.021*** (0.006)	0.058*** (0.018)	0.044*** (0.017)	0.196** (0.087)	0.137 (0.103)
City 2010, 0-20km		-0.001 (0.004)		0.065*** (0.010)		0.300*** (0.043)
<i>Panel C</i>	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.021*** (0.006)	0.008* (0.005)	0.058*** (0.018)	0.022 (0.021)	0.196** (0.087)	0.168* (0.090)
ln luminosity 2010		0.011*** (0.002)		0.055*** (0.006)		0.079*** (0.025)
Observations	7,007	7,007	9,284	9,284	9,464	9,464
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial Controls	YES	YES	YES	YES	YES	YES
Colonial Controls	YES	YES	YES	YES	YES	YES
Geographical Controls	YES	YES	YES	YES	YES	YES

Marginal effects from Probit estimates with dummies for being located between 0 and 10 km from schools, hospitals (around 2010) and feeder roads in 2016 as dependent variables, from the 11x11 km grid dataset, in columns 1 to 6. The full set of controls is included in columns 1 to 6 (see Table 2) and the sample is restricted to Ethiopia only. Robust standard errors, clustered by 202 provincial colonial districts, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A.1: The evolution of the road network in Ethiopia (1925-2011)



All-weather motorways (asphalt) are in red. All-weather motorways (asphalt) under construction are in dashed red. Gravel tracks are in solid black. Caravan tracks (dirt roads) are in dashed black and are only reported for 1925 and 1935. Data coverage is as follows: 1925, 1935 and 1939 maps are complete for Eritrea, Ethiopia and Italian Somalia. 1955, 1962, 1973, 1981 and 1988 maps are complete for Eritrea and Ethiopia. The 2011 map is complete for Ethiopia only. The year of each map corresponds to the date of publication of the respective source. Details on the sources of each period can be found in Appendix [10.2](#)

10 Appendices

10.1 Additional estimations and figures

Table A.1: Additional robustness tests for Table 2

<i>Specification:</i>							
	Baseline	No FE	Full sample	No targets 30km	<10km Roads 2010	Distance from targets	Non-arid cells
<i>Panel A</i>	Pop density 2000, z score						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road 1939, 0-10km	0.198*** (0.051)	0.170*** (0.058)	0.255*** (0.071)	0.191*** (0.052)	0.163*** (0.052)	0.189*** (0.050)	0.200** (0.078)
Road 1939, 10-20km	0.065** (0.031)	0.045 (0.038)	0.058 (0.064)	0.051 (0.031)	0.070* (0.036)	0.055* (0.031)	0.019 (0.042)
Log dist targets, km						-0.026 (0.018)	
Observations	15,288	15,288	15,550	14,965	8,594	15,288	7,513
R-squared	0.56	0.41	0.25	0.56	0.63	0.56	0.59
<i>Panel B</i>	ln luminosity 2013						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road 1939, 0-10km	0.483*** (0.102)	0.454*** (0.101)	0.561*** (0.112)	0.445*** (0.086)	0.444*** (0.103)	0.485*** (0.102)	0.558*** (0.130)
Road 1939, 10-20km	0.006 (0.049)	-0.012 (0.047)	0.019 (0.052)	0.009 (0.051)	0.010 (0.075)	0.008 (0.050)	-0.019 (0.044)
Log dist targets, km						0.007 (0.021)	
Observations	15,288	15,288	15,550	14,965	8,594	15,288	7,513
R-squared	0.19	0.17	0.27	0.18	0.22	0.19	0.20
<i>Panel C</i>	City 2015, yes/no						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road 1939, 0-10km	0.194*** (0.048)	0.193*** (0.043)	0.219*** (0.051)	0.201*** (0.049)	0.219*** (0.064)	0.174*** (0.048)	0.141* (0.074)
Road 1939, 10-20km	0.107*** (0.031)	0.109*** (0.028)	0.139*** (0.032)	0.108*** (0.032)	0.155*** (0.043)	0.089*** (0.031)	0.082** (0.042)
Log dist targets, km						-0.048*** (0.018)	
Observations	14,857	15,288	15,119	14,534	8,431	14,857	7,179
Ethnic FE	YES	NO	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES

The baseline set of controls and fixed effects are included in columns 1 to 7 (see Table 2 for a description). Column 1 of panels A, B and C reports the baseline estimation from columns 4, 5 and 7 of Table 2, respectively. Columns 2 and 3 drop fixed effects and include nodal locations, respectively. Column 4 drops cells located within 30 km of targets. Column 5 limits the sample to cells located within 10 km from roads in 2011 (both asphalt and gravel) as in Figure A.1. Column 6 controls for logarithmic distance from the targets. Column 7 drops cells below the median annual rainfall. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Additional robustness tests for Table 2

<i>Specification:</i>	Baseline	No low-lands	< 200km	<100km	<50km	Province FE	Regional FE	State FE
<i>Panel A</i>	Pop density 2000, z score							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road 1939, 0-10km	0.198*** (0.051)	0.321*** (0.079)	0.195*** (0.051)	0.205*** (0.050)	0.220*** (0.046)	0.231*** (0.052)	0.194*** (0.053)	0.164*** (0.062)
Road 1939, 10-20km	0.065** (0.031)	0.120** (0.047)	0.063** (0.032)	0.064** (0.031)	0.073*** (0.028)	0.089*** (0.032)	0.075** (0.033)	0.042 (0.040)
Observations	15,288	5,713	11,143	6,553	3,486	15,288	14,955	15,288
R-squared	0.56	0.54	0.55	0.55	0.57	0.59	0.52	0.41
<i>Panel B</i>	ln luminosity 2013							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road 1939, 0-10km	0.483*** (0.102)	0.369*** (0.088)	0.471*** (0.101)	0.441*** (0.099)	0.469*** (0.102)	0.441*** (0.093)	0.442*** (0.087)	0.439*** (0.097)
Road 1939, 10-20km	0.006 (0.049)	0.075 (0.079)	0.003 (0.049)	-0.004 (0.048)	0.032 (0.046)	-0.033 (0.049)	-0.030 (0.048)	-0.032 (0.046)
Observations	15,288	5,713	11,143	6,553	3,486	15,288	14,955	15,288
R-squared	0.19	0.18	0.19	0.21	0.27	0.24	0.22	0.18
<i>Panel C</i>	City 2015, yes/no							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road 1939, 0-10km	0.194*** (0.048)	0.356*** (0.073)	0.256*** (0.060)	0.295*** (0.066)	0.423*** (0.057)	0.233*** (0.044)	0.213*** (0.039)	0.192*** (0.044)
Road 1939, 10-20km	0.107*** (0.031)	0.214*** (0.047)	0.142*** (0.039)	0.165*** (0.043)	0.241*** (0.037)	0.138*** (0.030)	0.136*** (0.026)	0.111*** (0.028)
Observations	14,857	5,190	10,755	6,477	3,244	14,286	14,797	15,288
Ethnic FE	YES	YES	YES	YES	YES	NO	NO	NO
Precolonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES
State FE	NO	NO	NO	NO	NO	NO	NO	YES
Regional FE	NO	NO	NO	NO	NO	NO	YES	NO
Provincial FE	NO	NO	NO	NO	NO	YES	NO	NO

The baseline set of controls and fixed effects are included in columns 1 to 8 (see Table 2 for a description). Robust standard errors, clustered by 202 colonial provincial districts, in parentheses. Column 1 of panels A, B and C reports the baseline estimation from columns 4, 5 and 7 of Table 2, respectively. Column 2 restricts the sample to those cells with an average altitude between 800 and 2,000 meters above sea level. Columns 3, 4, and 5 restrict the sample to cells located within 200, 100, and 50 km from Italian paved roads, respectively. Columns 6 to 8 employ different sets of fixed effects, using different administrative colonial units: 202 provincial boundaries in column 6 (“residenze”), 73 regional boundaries (“commissariati”) in column 7 and 6 state boundaries (“governi”) in column 8. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Robustness for light density and cities

<i>Specification:</i>	IV				IV			
	Baseline	First stage	LCP 0-10km	Placebo	Baseline	First stage	LCP 0-10km	Placebo
	ln luminosity 2013				City 2015, yes/no			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road 1939, 0-10km	0.483*** (0.102)		0.446** (0.192)		0.194*** (0.048)		0.145* (0.084)	
Road 1939, 10-20km	0.006 (0.049)	-0.232*** (0.017)	-0.001 (0.056)		0.107*** (0.031)	-0.233*** (0.017)	0.108*** (0.037)	
0-10km, IV LPC 1940		0.425*** (0.057)				0.424*** (0.057)		
Proj road 1937 unpav, 0-10km				-0.118* (0.060)				0.086 (0.063)
Proj road 1937 unpav, 10-20km				-0.061* (0.034)				0.078*** (0.030)
Observations	15,288	15,288	15,288	15,288	14,857	15,288	15,288	14,857
R-squared	0.19	0.38	0.14	0.18	.	0.38	0.20	.
Cragg-Donald Wald F-Stat	.	.	3141	.	.	.	3125	.
Kleibergen-Paap F-Stat	.	.	55.15	.	.	.	55.58	.
Ethnic FE	YES	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES

The baseline set of controls and fixed effects are included in columns 1 to 8 (see Table 2 for a description). Robust standard errors, clustered by 202 colonial provincial districts, in parentheses. Columns 1 and 5 report the baseline estimations for light density and location of cities, as in columns 5 and 7 from Table 2. Columns 2, 3 4, 6, 7 and 8 replicate the robustness checks from Table 4 (columns 2 3 and 4) with light density and city location as dependent variables. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Replication of Table 2 with 30x30 km grid cells

<i>Dependent variables:</i>	Pop density 2000, z scores				ln luminosity 2013		City 2015, yes/no
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road 1939, 0-10km	0.747*** (0.191)	0.414** (0.166)	0.314** (0.142)	0.314** (0.137)	1.496*** (0.354)	1.266*** (0.329)	0.146** (0.058)
Road 1939, 10-20km	0.582*** (0.117)	0.265** (0.104)	0.233** (0.106)	0.225** (0.106)	0.771*** (0.254)	0.606** (0.248)	0.068* (0.038)
Pop density 2000 z scores						0.735*** (0.092)	
Observations	1,697	1,697	1,697	1,697	1,697	1,697	1,625
R-squared	0.52	0.60	0.61	0.63	0.37	0.43	.
Ethnic FE	YES	YES	YES	YES	YES	YES	YES
Precolonial controls	NO	YES	YES	YES	YES	YES	YES
Colonial controls	NO	NO	YES	YES	YES	YES	YES
Geographical controls	NO	NO	NO	YES	YES	YES	YES

The baseline set of controls and fixed effects are included in columns 1 to 7 (see Table 2). Robust standard errors, clustered by 202 colonial provincial districts, in parentheses. The table replicates Table 2 using larger grid cells of about 30x30 km (0.3x0.3 degrees). *** p<0.01, ** p<0.05, * p<0.1

Table A.5: T-test between placebo and treated observations

	Mean	t-stat
	(1)	(2)
Rainfall	-0.360	(-0.53)
Land Suitability	0.367***	(4.50)
Temperature	3.209***	(8.96)
Agglomeration 1935, 1-0	-0.0281	(-0.99)
City 1935, 1-0	0.0750**	(2.95)
Town 1935, 1-0	-0.0259	(-1.22)
Track 1935, 0-10km	0.0135	(0.37)
Telegraph 1934 CMI, 0-10km	-0.00858	(-0.25)
Observations	918	

t statistics in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Alternative estimates for relocation

<i>Panel A</i>						
<i>Dependent variables:</i>	Village Rank, 1939			Village Rank, 1938		
	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.485*** (0.066)	0.463*** (0.080)	0.462*** (0.078)	0.302*** (0.084)	0.268*** (0.095)	0.283*** (0.094)
Road 1939, 10-20km	0.067 (0.072)	0.046 (0.089)	0.045 (0.087)	-0.029 (0.079)	-0.062 (0.091)	-0.047 (0.090)
Road 1940, 20-30km		-0.029 (0.088)			-0.118 (0.088)	
Road 1940, 30-40km		-0.030 (0.075)			-0.076 (0.070)	
Road 1940, 40-50km		-0.063 (0.076)			-0.017 (0.078)	
Road 1940, 50-60km		-0.024 (0.064)			0.067 (0.074)	
Road 1940, 20-60km			-0.037 (0.055)			-0.029 (0.056)
<i>Panel B</i>						
<i>Dependent variables:</i>	Villages 1939, 0-10km			Villages 1938, 0-10km		
	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.122*** (0.017)	0.118*** (0.020)	0.118*** (0.020)	0.081*** (0.024)	0.071*** (0.026)	0.076*** (0.026)
Road 1939, 10-20km	0.015 (0.018)	0.011 (0.023)	0.011 (0.022)	-0.005 (0.023)	-0.015 (0.025)	-0.011 (0.025)
Road 1940, 20-30km		-0.008 (0.021)			-0.038 (0.025)	
Road 1940, 30-40km		-0.008 (0.019)			-0.025 (0.019)	
Road 1940, 40-50km		-0.010 (0.019)			0.002 (0.022)	
Road 1940, 50-60km		-0.001 (0.016)			0.016 (0.020)	
Road 1940, 20-60km			-0.006 (0.014)			-0.009 (0.015)
Agglomeration 1935, 1-0	0.227*** (0.014)	0.227*** (0.014)	0.227*** (0.014)	0.180*** (0.018)	0.179*** (0.018)	0.179*** (0.018)
Observations	10,088	10,088	10,088	10,088	10,088	10,088
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES

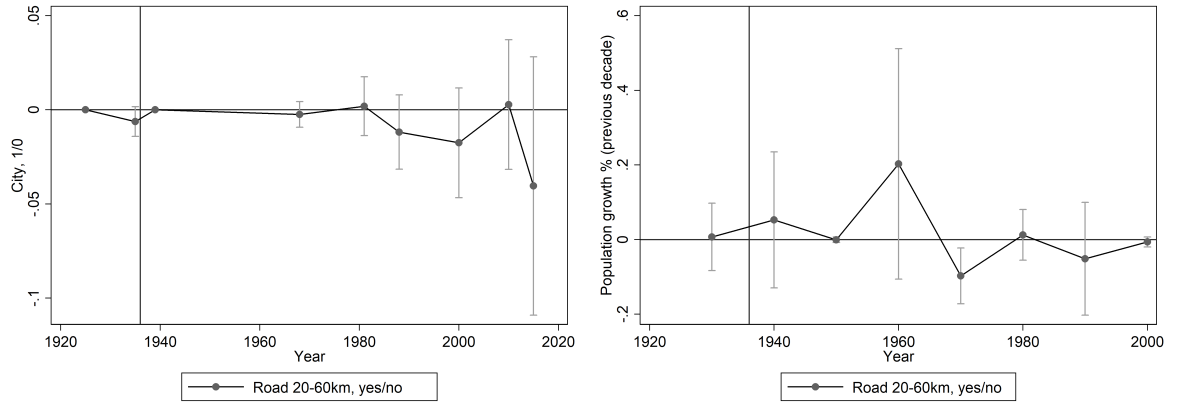
Panel A: marginal effects from an ordinal probabilistic regression with village ranks as dependent variable. *Panel B:* marginal effects from a Probit model (same as Table 6), but controlling for settlement in 1935. The baseline set of controls and fixed effects are included in columns 1 to 7, in both panels (see Table 2). Robust standard errors, clustered by 202 provincial colonial districts, in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Transformation of the road network in Ethiopia, 1955-2011

Year	Deterioration, %	N
	(1)	(2)
1955	0.172	605
1962	0.283	605
1973	0.174	605
1981	0.526	605
1988	0.486	605
2011	0.0926	605
	Expansion, %	N
	(1)	(2)
1955	0.0997	9,655
1962	0.0669	9,655
1973	0.167	9,655
1981	0.0643	9,655
1988	0.0732	9,655
2011	0.169	9,655

The table refers to contemporary Ethiopia only and includes both nodal and non-nodal locations. “Deterioration” is a dummy that equals one if a cell’s centroid was located within 10 km from an Italian paved road, but has lost accessibility (not within a 10 km radius from a paved road) in one of the reported years. “Expansion” equals one if a cell of the control group in 1940 (farther than 10 km away from an Italian paved road) gains access to an asphalt road. Targets are included.

Figure A.2: Relocation over time



Coefficients for the 20-60 km cut-off distance dummy from colonial paved roads (same model as in column 3 of Table 5) are reported with 95% confidence intervals. Dependent variables as shown on the Y-axis. Results for cities above 10,000 inhabitants are marginal effects from a Probit model with the dependent variable being a dummy indicating a 20 km cut-off radius from a city with more than 10,000 inhabitants's location. Results for population growth are OLS estimates. Errors are clustered at provincial colonial district level. All estimations include the baseline set of controls (see Table 2 for details) and the employed sample includes all non-nodal cells located in contemporary Ethiopia.^a Vertical lines correspond to 1936. For the set of probabilistic estimations with presence of cities as dependent variable, the set of fixed effects is different from Equation 1 and corresponds to the four regions in which the Italian government divided Ethiopia in 1936.^b

^aEstimates for population growth in 1930, 1940, 1950 and 1960 exclude the region of Shewa (the surroundings of Addis Abeba) due to missing data.

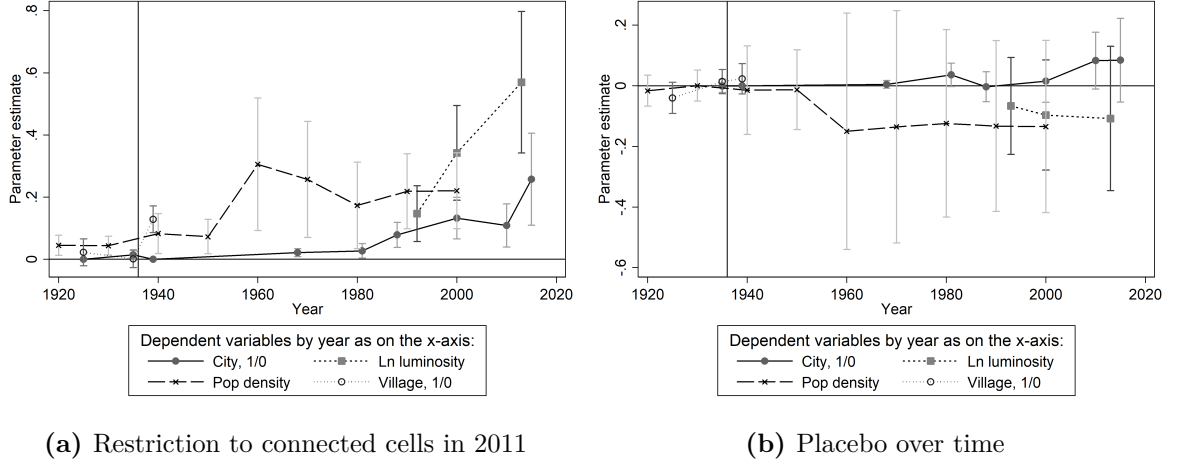
^bFor the early period there is not enough variation in the dependent variable to keep ethnic fixed effects in the model.

Table A.8: Core-periphery structure

<i>Dependent variables:</i>	ln luminosity 2013			Manufacturing, yes/no		
	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.599*** (0.115)	0.665** (0.282)	0.597*** (0.115)	0.015* (0.008)	0.009 (0.016)	0.015* (0.008)
Road 1939, 0-10km*Distance from 1935 targets (km)		-0.001 (0.002)			0.000 (0.000)	
Road 1939, 0-10km*Town 1935 (yes/no)			0.691 (1.368)			0.018 (0.012)
Observations	10,088	10,088	10,088	9,820	9,820	9,820
R-squared	0.19	0.19	0.19	.	.	.
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES

OLS estimates with the logarithm of light density in 2013 (grid dataset) in columns 1 to 3. Marginal effects from a Probit model with dummies for being employed in manufacturing (DHS individual dataset) as dependent variable, in columns 4 to 6. All columns include the baseline set of controls and fixed effects as in Table 2 and the sample is restricted to Ethiopia only (non-nodal observations). Robust standard errors, clustered by 202 provincial colonial districts, in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A.3: Robustness for the effect of Italian roads over time



Panel A: Replication of Figure 2a, but restricting to the sub-sample of cells that are within 10 km of a road in 2011 (All gravel roads as reported in Figure A.1). Coefficients for 10 km cut-off distance dummy of colonial paved roads are reported with 95% confidence intervals. The applied restriction yields a sample of 5,595 non-nodal cells, from the original 10,088 for Ethiopia. *Panel B:* Coefficients for 10 km cut-off distance dummy of placebo roads are reported with 95% confidence intervals. Treated locations are included in the sample. *Both panels:* Dependent variables as shown in the legend. All estimations include the baseline set of controls and fixed effects as in Equation 1^a (see Table 2 for details). The employed sample includes all non-nodal cells located in contemporary Ethiopia.^b Vertical lines correspond to 1936. Reported coefficients for population density and luminosity (Log light density) are OLS estimates. Results for villages are marginal effects from a Probit model (same as Equation 1) with binary variables for the presence of villages in cell i in 1925, 1935 and 1939. Results for cities above 10,000 inhabitants are marginal effects from a Probit model with the dependent variable being a dummy indicating a 20 km cut-off radius from a city with more than 10,000 inhabitants.

^aFor the set of probabilistic estimations with presence of cities as dependent variable, the set of fixed effects is different from Equation 1 and corresponds to the four regions in which the Italian government divided Ethiopia in 1936. For the early years (1925, 1935, 1938 and 1939), in fact, there is not enough variation in the dependent variable to keep ethnic fixed effects in the model.

^bEstimates from HYDE population density for 1920, 1930, 1940 and 1950 exclude the region of Shewa (the surroundings of Addis Abeba) due to missing data.

Figure A.4: Extended colonial roads in 2011

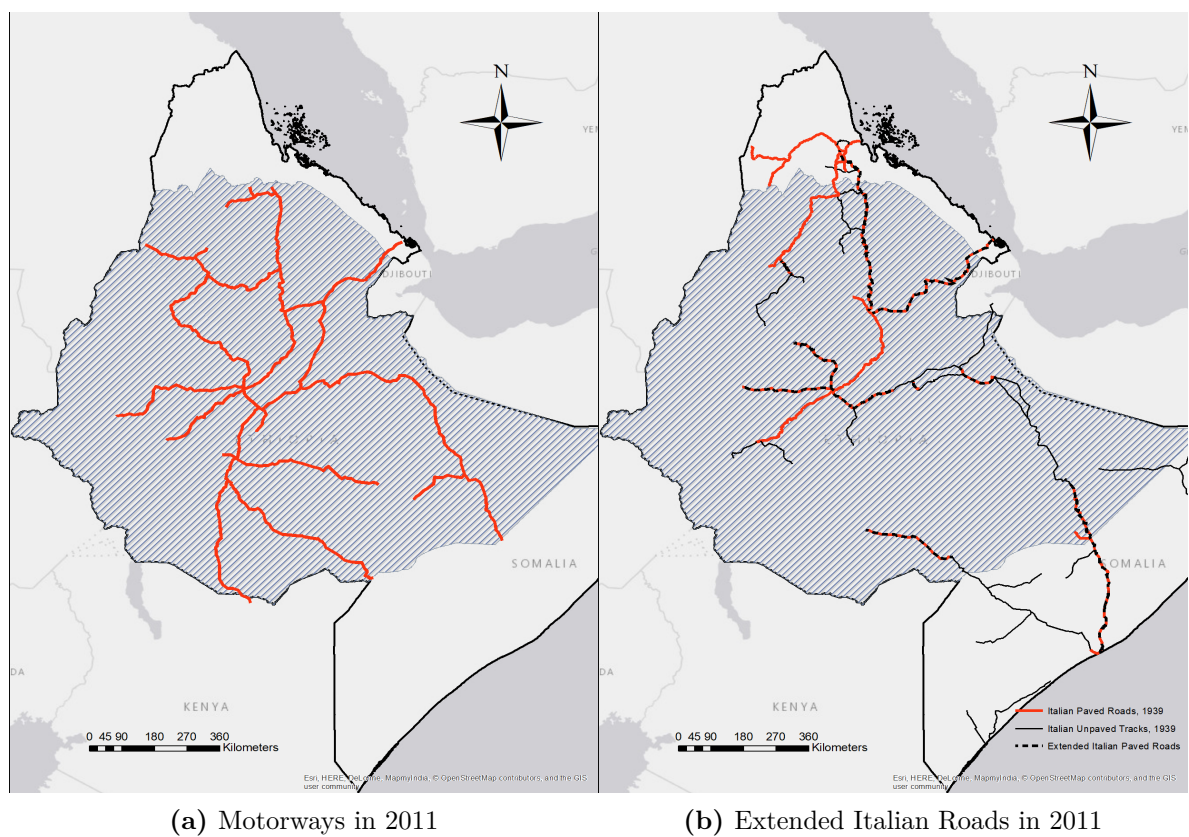


Table A.9: First and second-wave effect

<i>Dependent variables:</i>	Pop density 2000, z scores					
	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.251*** (0.059)					
Pop density 1930, z scores	0.238*** (0.051)					
New road 1955, 0-10km		0.058 (0.049)				
Pop density 1950, z scores		0.104*** (0.026)				
New road 1962, 0-10km			0.030 (0.066)			
Pop density 1960, z scores			0.655*** (0.038)			
New road 1973, 0-10km				0.034 (0.024)		
Pop density 1970, z scores				0.657*** (0.023)		
New road 1981, 0-10km					0.015 (0.019)	
Pop density 1980, z scores					0.856*** (0.033)	0.855*** (0.033)
New road 1988, 0-10km						0.065* (0.034)
Observations	10,088	9,575	9,575	9,575	9,575	9,575
R-squared	0.56	0.55	0.93	0.92	0.96	0.96
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial controls	YES	YES	YES	YES	YES	YES
Colonial controls	YES	YES	YES	YES	YES	YES
Geographical controls	YES	YES	YES	YES	YES	YES

OLS estimates with z-scores of population density in 2000 (grid dataset) in columns 1 to 6. All columns include the baseline set of controls and fixed effects as in Table 2 and the sample is restricted to Ethiopia only (non-nodal observations). New roads by year are shown in Figure A.1. For each year, a cell is coded as being next to a new road if it is within 10 km of a paved road that was not extant in the previous road cross-section. In each column, the control for population density corresponds to the first year available before the date of the road cross-section. The sample in columns 2 to 6 excludes treated cells, thus providing upper bound estimates of the effect of second-wave infrastructure. Robust standard errors, clustered by 202 provincial colonial districts, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.10: Coordination of post-colonial investments

<i>Dependent variables:</i>	School 1981, 0-10km			School 1988, 0-10km		
<i>Panel A</i>	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.049*** (0.013)	0.037*** (0.013)	0.042*** (0.014)	0.063*** (0.018)	0.052*** (0.017)	0.057*** (0.017)
Pop density 1960, z scores		0.027*** (0.004)			0.029*** (0.006)	
City 1981, 0-20km			0.078*** (0.012)			0.091*** (0.014)
Observations	8,895	8,895	8,895	9,196	9,196	9,196
<i>Dependent variables:</i>	Pop density 2000, z scores		ln luminosity 2013		City 2015, yes/no	
<i>Panel B</i>	(1)	(2)	(3)	(4)	(5)	(6)
Road 1939, 0-10km	0.251*** (0.059)	0.182*** (0.058)	0.599*** (0.115)	0.463*** (0.102)	0.271*** (0.068)	0.221*** (0.069)
School 2010, 0-10km		0.449*** (0.075)		1.192*** (0.147)		0.092** (0.046)
Hospital 2010, 0-10km		0.283*** (0.033)		0.462*** (0.057)		0.177*** (0.023)
Feeder road 2016, 0-10km		0.093*** (0.017)		0.036 (0.024)		0.252*** (0.024)
Observations	10,088	10,088	10,088	10,088	9,896	9,896
R-squared	0.56	0.61	0.19	0.29	.	.
Ethnic FE	YES	YES	YES	YES	YES	YES
Precolonial Controls	YES	YES	YES	YES	YES	YES
Colonial Controls	YES	YES	YES	YES	YES	YES
Geographical Controls	YES	YES	YES	YES	YES	YES

Panel A: marginal effects from Probit estimates with dummies for being located between 0 and 10 km from schools in 1981 and 1988 as dependent variables, from the 11x11 km grid dataset. The full set of controls is included in columns 1 to 6 (see Table 2) and the sample is restricted to Ethiopia only. Panel B: see Table 2 for details. The sample is restricted to Ethiopia. Robust standard errors, clustered by 202 provincial colonial districts, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

10.2 Data appendix

10.2.1 Dependent variables

1. **Population density, 1960-2000:** from the Unep/Grid Sioux Falls database²⁷, raster format (0.04 x 0.04 decimal degrees). It is combined with the grid dataset by taking the average for all raster cells, whose centroid falls within a 11x11 km grid's cells. The raster dataset is constructed from district-level population statistics, which are distributed on the raster surface through an accessibility index that incorporates agglomerations and infrastructure. Population densities are calculated by dividing population count per cell by the actual area of the cell (different latitudes can cause severe distortions in the grid). This source covers the period between 1960 and 2000 with decadal frequency.
2. **Population density, 1900-1950:** reconstructed population density from the History Database of Global Environment (HYDE)²⁸. This database is created using the available historical data which is distributed across the earth's surface based on cropland and pasture statistics and is available in raster format (0.08x0.08 decimal degrees). The reconstruction is accurate for Eritrea and Italian Somalia, whereas the Ethiopian territory shows significant spatial correlation.
3. **Villages, 1925-1939:** data on location and relative size of villages (agglomerations between 100 and 10,000 inhabitants) have been digitized from different cartographic sources. Data on villages' locations for Ethiopia, Eritrea and Italian Somalia come from a set of Maps by the Cartographic office of the Ministry of the colonies for the years 1925, 1935 and 1939 (Dardano, 1925; Dardano, 1935; Dardano, 1939). An additional year could be added by using a very detailed tourist map of the Consociazione Turistica Italiana from 1938 (Consociazione Turistica Italiana (CTI), 1938?).
4. **Cities, 1925-2015:** information on the location of cities with more than 10,000 inhabitants come from various cartographic sources. For the years 1925, 1935 and 1939 these are retrieved from Achille Dardano's maps (Dardano, 1925; Dardano, 1935; Dardano, 1939). All agglomerations marked with a symbol of relative rank 4 and above are coded as cities. For 1968 data come from a Russian intelligence map that covers Ethiopia

²⁷ Available at <http://na.unep.net/siouxfalls/datasets/datalist.php>, last accessed on 20th of April 2016

²⁸ Available at <http://themasites.pbl.nl/tridion/en/themasites/hyde/download/index-2.html>, last accessed on 20th of April 2016

and Eritrea and reports exact figure on city’s population (Union of Soviet Socialist Republics (USSR), 1968). Location of cities in 1981 and 1988 come from two geographical atlases of Ethiopia (Ethiopian Mapping Agency, 1981; Ethiopian Mapping Authority, 1988) and report information on urban population accurately. Finally, information on the location of cities in 2000, 2010 and 2015 comes from the OECD/SWAC Africapolis Database (2020).²⁹ Cells are coded as being part of a city if their centroid falls within a 20 km radius of them.

5. **Satellite Light Density, 1992-2013:** the employed data come from Version 4 of DMSP-OLS Nighttime Lights Time Series, which is accessible from the National Geophysical Data Center (NOAA) in a raster format (0.008x0.008 decimal degrees)³⁰. The years 1992 (first available), 2000, 2010 and 2013 (last available) are selected. Each 11x11 km grid cell is given the average value of night density index across all raster cells that fall within the square. This raster data provides the measured average nightly luminosity produced by humans, and it is corrected for potential natural lights and atmospheric phenomena. In the regressions, the transformation $\log(\text{Min}(x)/2 + x)$ is employed.
6. **Individual data, 2011:** from the 2011 Ethiopian USAID/DHS (Demographic and Health Survey) individual dataset for men.³¹ For each individual, this source provides categorical information about their profession and a wealth index that combines all goods and assets owned by the individual’s household. Each individual is linked to one of the 649 geo-coded Ethiopian villages. When residents living within 20 km of the Italian road network’s targets are excluded, the sample shrinks from a total of 14,008 to 10,772 individuals. Dummies for sectoral occupation are created from the DHS occupation variable mv716. Individuals are coded as being employed in the tertiary sector if their code is 11, 12, 13, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 51 and 52. Codes for the secondary sector are 71, 72, 73, 74, 81, 82, 83 and 93. Codes for the primary sector are 61, 62 and 92. An individual is coded as being above the mean wealth index if the value of variable mv190 is above the sample’s mean of 2843.634 (including nodal cells).

²⁹ Available at www.africapolis.org last accessed on 11 February 2020.

³⁰ Available at <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>, last accessed on 20th April 2020

³¹ Available at <http://www.dhsprogram.com/Data/>, last accessed on 20th of April 2016

10.2.2 Historical data

1. **Colonial road network, 1939:** Italian paved motorways and unpaved tracks completed between 1936 and 1939 were digitized from the map *Carta progressiva delle costruzioni stradali* (outside text) contained in the 4th issue of 1939 of Piccioli (eds.) (1937–1943). The Euclidean distance of each cell’s centroid from paved and unpaved roads was calculated.
2. **Roads under construction, 1939–40:** roads that were projected in 1936–7 are digitized from the map “Strade in AOI” (outside text) in Sillani (1937). To define placebo lines, I select unpaved tracks from the map *Carta progressiva delle costruzioni stradali* (outside text) contained in the 4th issue of 1939 of Piccioli (eds.) (1937–1943) that were also included in the initial construction plan as reported by Sillani (1937).
3. **Colonial railroads, 1939:** the location of colonial railroads for Eritrea and Ethiopia comes in shapefile format from DIVA GIS³² as functioning railway lines have not changed since 1941. For Italian Somalia, where colonial trunks were dismissed, I have geo-referenced the map *Carta dei commissariati, delle residenze, delle vie di comunicazione e delle popolazioni* (outside text) in Corni (1937).
4. **Colonial facilities, 1940:** colonial primary hospital, secondary hospitals and colonial schools for indigenous people were digitized from a series of maps (one for each regional government and type of facility) contained in the first issue of 1940 of Piccioli (eds.) (1937–1943).
5. **Administrative boundaries, 1940:** colonial district capitals (Residenze), along with provincial (Residenze), regional (Commissariati) and governmental (Governi) boundaries come from the map “Circoscrizioni amministrative dell’Arca Orientale Italiana” (outside text), contained in the 1939 3rd issue of Piccioli (eds.) (1937–1943).
6. **Areas of mining interest, 1940:** information on whether a cell’s centroid falls within a zone of “mining interest” come from the map *Carta mineraria dell’AOI* (outside text) from the 2nd issue of 1940 of Piccioli (eds.) (1937–1943). Areas of “mining interest” refer to areas where excavations were taking place between 1936 and 1940, or were significant

³²Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

mineral deposit had been identified. All cell's centroids falling within these areas are assigned to this category.

7. **Pre-1935 communication infrastructure:** telegraph and telephone lines were digitized from the map “Impero etiopico - Rete delle trasmissioni” (Comando Militare Italiano (CMI), 1935).
8. **Pre-colonial tracks, 1925-1935:** the location of pre-colonial tracks employed in the baseline were digitized from Dardano's maps from 1925 (Dardano, 1925) and 1935 (Dardano, 1935). Both major (“piste camionabili”) and secondary (“piste secondarie”) tracks were digitized.
9. **Post-colonial transportation network, 1955-2011:** post-colonial data on the road network come from various maps and atlases published by the Ethiopian government, which I have geo-coded and digitized. The road network in 1955, 1962 and 1973 has been digitized from a series of maps from the Ethiopian Imperial Highway Authority (Imperial Highway Authority, 1955; Imperial Highway Authority, 1962; Imperial Highway Authority, 1973), while maps from the “National Atlases of Ethiopia” were employed to reconstruct the transportation network in 1981 and 1988 (Ethiopian Mapping Agency, 1981; Ethiopian Mapping Authority, 1988). Asphalt motorways in 2011 are from International Travel Maps (2011), while secondary roads around 2010 were downloaded from DIVA GIS.³³ Feeder roads in 2016 are from a survey carried out by the Ethiopian Road Authority (ERA).³⁴
10. **Post-colonial Hospitals and schools:** the location of contemporary (2010?) hospitals and schools was retrieved from online sources: location of contemporary hospitals in Ethiopia is from the map *Ethiopia health facilities*, available on the GIS online platform. For contemporary schools, I rely on a list of Ethiopian schools taken from an Ethiopian social network that is used by former students to keep in contact after graduation.³⁵ The location of Ethiopian schools in 1981 and 1988 was obtained by digitizing two maps contained in Ethiopian Mapping Agency (1981) and Ethiopian Mapping Authority (1988).

³³Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

³⁴<http://www.era.gov.et/web/guest/about-us>

³⁵Available on the website [graduates.com](http://www.graduates.com/), <http://www.graduates.com/Schools/Ethiopia>, last accessed on 21st of April 2016

10.2.3 Geographical and environmental controls

1. **Altitude:** altitude in meters was downloaded, in raster format (0.008x0.008 decimal degrees), from DIVA-GIS.³⁶ Average raster value by 11x11 km grid cell is calculated.
2. **Rainfall:** annual rainfall average (centimetres/year) is from the “Worldclimate” project, available in raster format (0.04x0.04 decimal degrees) on DIVA GIS.³⁷ Average raster value by 11x11 km grid cell is calculated.
3. **Malaria:** malaria transmission index is available, in raster format (0.5x0.5 decimal degrees) from Kiszewski’s dataset (Kiszewski et al., 2004). This source informs on the rapidity and likelihood of malaria transmission in each cell. Each cell’s centroid is given the value of the raster cell it falls into.
4. **Land suitability:** agricultural suitability for rain-fed low input cereals for each cell was calculated from the GAEZ-FAO land fertility database³⁸, available in raster format (0.5x0.5 decimal degrees). This index ranges from 0 to 9 (0 highest, 9 lowest) and factors in temperature, soil moisture, and rainfall patterns. Average raster value by 11x11 km grid cell is calculated.
5. **Ruggedness:** terrain ruggedness in thousands of percentage points comes, in raster format (0.08x0.08 decimal degrees) from Nunn and Puga (2012).³⁹ Average raster value by 11x11 km grid cell is calculated.
6. **Temperature:** monthly (September) average temperature data between 1960 and 1990 are in raster format (0.008x0.008 decimal degrees) and from the “Worldclimate” project.⁴⁰ Average raster value by 11x11 km grid cell is calculated.
7. **Perennial rivers:** perennial waterways come in shapefile format by country from DIVA GIS.⁴¹ The Euclidean distance from each cell’s centroid to the closest river is calculated in kilometers.

³⁶ Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

³⁷ Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

³⁸ Available at <http://www.fao.org/nr/gaez/about-data-portal/en/>, last accessed on 21st of April 2016

³⁹ Available at <http://diegopuga.org/data/rugged/>, last accessed on 3rd of October 2017

⁴⁰ Available at <http://www.worldclim.org/version1>, last accessed 3rd October 2017

⁴¹ Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

8. **Coastline:** the coastline is derived from shapefiles of contemporary, downloaded from DIVA GIS.⁴² The Euclidean distance from each cell’s centroid to the closest point of the coastline is calculated in kilometers.

10.2.4 Least Cost Path Network (LCP)

1. Summary of the procedure

The LCP network was created starting from the target locations reported in Figure 1b. These are selected based on the targeted towns as reported in the Italian construction plans (Cobolli-Gigli, 1937). In case a trunk was not completed, the two ending points of the segment are selected instead. The network’s termini thus tend to coincide with pre-1935 large cities and ports, although a few discrepancies exist and not all major cities were targeted by the network. Least cost paths were only calculated between pairs that had an actual road built in between by the end of the Italian occupation in 1940. The outcome of the calculation is plotted in Figure 1b.

The calculation of building costs is based on the two main geographical variables that, from historical and contemporary sources, emerge as the most significant factors in determining the cost of road-building in the context of the Horn of Africa (Cobolli-Gigli, 1938). These are land cover, which is created by merging a land use raster with both major and seasonal rivers and terrain gradient (calculated in percentage points from elevation data).

I first create a weighted cost raster based on building costs differentials and a weighting procedures (described in points 1 and 2). I then use this to calculate a cost distance raster for each target and, finally, the least cost paths between pairs of targets using the “Cost path” tool in Arcgis.

2. Cost differentials

Cost differentials between raster cells with different characteristics are determined by combining the land cover and the gradient data through a weighting matrix. The cost

⁴²Available at <http://www.diva-gis.org/>, last accessed on 20th of April 2016

differentials for building on different terrains are retrieved from contemporary construction manuals (Anas, 2018; United States Department of Agriculture (USDA), 2017) and are reported in Table 3.B.1. For instance, building on an empty cell with a gradient between 1 and 4% would be 10% more expensive than the base cost for building on a flat and empty cell.

Table 3.B.1: Cost matrix from contemporary sources

<i>Slope</i>	0-1%	1-4%	4-8%	8-12%	>12%
<i>Land Cover</i>					
Empty	100	110	115	120	125
Forest	110	120	125	130	135
S Rivers	115	125	130	135	140
L Rivers	120	130	135	140	145
Lakes	125	135	140	145	150

As these gaps reflect the costs of building a paved road on the described surface at current prices and with the technology available today, these are unlikely to reflect the actual costs faced by Italian builders in the Horn of Africa during the 1930s. The available technology, in fact, was rudimentary and materials costly as they had to be largely imported from Italy or transported at high cost from coastal areas (Cobolli-Gigli, 1938). Therefore, I multiply cost differential from Table 3.B.1 by a factor of 4, following the suggested ratio between cost per kilometer in Italian East Africa and mainland in Italy in 1936 as reported by contemporary sources (Cecini, 2007, pp. 134-5). The adjusted weights are reported in Table 3.B.2.

Table 3.B.2: Cost matrix adjusted by 1936 costs

<i>Slope</i>	0-1%	1-4%	4-8%	8-12%	>12%
<i>Land Cover</i>					
Empty	100	140	160	180	200
Forest	140	180	200	220	240
S Rivers	160	200	220	240	260
L Rivers	180	220	240	260	280
Lakes	200	240	260	280	300

In the final cost matrix, therefore, the cost of building a paved road in a certain cell, ranges from the basic unitary cost of 100 (for a flat and empty cell) to 300 (for a hypothetical cell which has a gradient of more than 12% and it is covered by a lake.)

3. Weighting procedure:

First, I re-classify the land cover and the gradient rasters into the categories reported in Tables 3.B.1 and 3.B.2. Second, I use the weighted values from cost matrix in Table 3.B.2 to weight the land cover and the gradient rasters's categories using the “weighted overlay” tool in ArcMap. For the final version, I assign an overall weight of 40% to the land cover layer and 60% to the gradient one as, from the available historical sources (Cobolli-Gigli, 1938), slope seems to have been the single most important factor in determining building costs, given the significant variation in altitude across the area.

As the assigned weights are, to a certain extent, arbitrary, I run several robustness checks, both changing the overall weight assigned to each raster (using 40-60%, 50-50% and 60-40% relative overall weights) and by replicating all calculations using the unadjusted cost differentials (as in Table 3.B.1) and different adjustments for local costs (namely, x2, x8 and x12). Small changes to the LCP network occur when these different robustness checks are implemented, but they do not change the IV calculations significantly. Consistently with the secondary evidence, the selected weighting procedure has the best fit.