# Spatial Inefficiencies in Africa's Trade Network

Tilman Graff*

May 30, 2024

**Abstract**

I assess the efficiency of transport networks for every country in Africa. Using spatial data from various sources, I simulate trade flows over more than 70,000 links covering the entire continent. I maximise over the space of networks and find the optimal road system for every African state. My simulations predict that Africa would gain 1.3% of total welfare from reorganising its national road systems, and 0.8% from optimally expanding it by a tenth. I then construct a dataset of local network inefficiency and find that colonial infrastructure projects significantly skew trade networks towards a sub-optimal equilibrium today. I find suggestive evidence that regional favouritism played a role sustaining these imbalances.

## 1 Introduction

Trade costs in Africa are the highest in the world, severely inhibiting interregional trade (Limao and Venables, 2001; The Economist, 2015; Nugent and Lamarque, 2022). Sub-Saharan Africa's coverage with paved roads is by far the lowest of any world region, with only 31 total paved road kilometres per 100 square kilometres of land, compared to 134 in other low-income countries (Foster and Briceño-Garmendia, 2010). The World Bank has identified an annual infrastructure gap amounting to 93 billion US dollars and urges countries in Sub-Saharan Africa to spend almost one per cent of GDP on building new roads (Foster and Briceño-Garmendia, 2010; Nugent and Lamarque, 2022). This reasoning is also reflected in the composition of development aid – in 2017, by far the largest share of World Bank lending to African countries was allocated to transport infrastructure projects (The World Bank, 2017). There appears to be a clear consensus that Africa needs more roads.

In this paper, I investigate a neglected, yet powerful additional drag on Africa's transport system. I don't ask if the continent has too *few* roads, but rather analyse whether the current infrastructure is *in the wrong place*. Do Africa's roads connect the right areas to promote trade? How would a social planner design a perfect transport network which optimises welfare in a given country? And what can colonial history tell us about why some countries are far behind their hypothetical optimum?

I derive the unique optimal trade network for every country in Africa. Using data from satellites and online routing services, I first construct an interconnected economic topography of more than 10,000 square grid cells covering the entire continent. I then employ a simple network trade model

*Harvard University. tgraff@g.harvard.edu. This paper originated as my master's thesis under the invaluable guidance by my advisor Ferdinand Rauch. I thank the editor Samuel Bazzi, as well as two anonymous referees for insightful comments that have greatly improved this paper. I also thank Pol Antras, Antonia Delius, Dave Donaldson, Lukas Freund, Doug Gollin, Johannes Haushofer, Gabriel Kreindler, Ameet Morjaria, Felix Samy Soliman, Fabio Schmidt-Fischbach, Evan Soltas, Tony Venables, and seminar participants at Oxford, LSE, NBER, and the World Bank for helpful feedback and discussion.

to simulate trade flows through more than 70,000 links spanning all of Africa. In a second step, I use a variant of a recently established framework by Fajgelbaum and Schaal (2020) to optimise over the space of networks and find the optimally redesigned transport system given the underlying economic fundamentals for every African country. An intuitive thought experiment demonstrates this process: suppose the social planner were to observe the spatial distribution of roads, people, and economic activity in a given country before being allowed to lift all roads from their current location, freely shuffle them around, and then reorganise them in the most efficient way for mutual trade. The planner is not allowed to build completely new roads, but is only allowed to move infrastructure from one part of the country to another. In this exercise, she takes into account local incentives for trade between all sets of neighbours on a complex network graph, regional differences in trade costs caused by geographical and network characteristics, and heterogeneous costs to constructing new roads depending on the underlying terrain.

I then compare these optimal networks to the current system. I argue that the degree to which the optimum differs from the status quo can be interpreted as an intuitive measure for the inefficiency of a country's current road network. I show that potential welfare gains from reshuffling roads would improve overall welfare on the continent by about 1.3%. I use my simulations to identify Somalia and Sudan as the countries with the most inefficient transport network in Africa. I also compute returns to a hypothetical large-scale expansion of the trade network on the continent, amounting to additional investments worth 10% of Africa's current road stock. I find a return on investment of around 8% in welfare terms to such a program. Trade networks in the US and China, by comparison, are much closer to their hypothetical optimum.

On the regional level, this scenario creates winners and losers. The model identifies some areas as having *too* many roads and decides to put them to better use somewhere else. These areas were inefficiently overendowed with transportation infrastructure before the reshuffling exercise. Other regions, however, did not have enough infrastructure given their relative position in the network and are now awarded additional roads by the social planner. I identify these areas as discriminated against by the current transportation network design. By comparing welfare levels before and after the hypothetical intervention, I create a novel dataset of local infrastructure discrimination for more than 10,000 cells covering the entire African continent.

Why are some regions systematically cut off from the benefits of efficient trade? I investigate the long-run effects of large infrastructure investments from the colonial area. Similarly to Jedwab and Moradi (2016), I find a persistent impact of railway lines constructed by the colonial powers over a century ago. Plausibly exogeneous variation in the number of kilometres crossing a given area significantly skews the current trade network towards a suboptimal state today. Even though many of the railway lines have fallen into disarray since independence, regions close to colonial railroads still have too much road infrastructure given their relative position in the network. In contrast, railway lines that were planned, but by historical accident never built, do not predict any significant departure from the optimal spatial distribution.

Using historical roads data from Kenya, I investigate mechanisms for this spatial persistence and find that policymakers often failed to close the infrastructure imbalances left behind by the colonial powers (Burgess et al., 2015). In times of non-democracy, Kenyan road development projects did not go to the areas most in need, but rather to places sharing ethnicity with the leader at the time.

The main caveat to my findings is that my spatial model does not feature a time dimension. My exercise identifies the optimally reallocated or expanded network at the current point in time, and hence might mis-characterise forward-looking infrastructure investments into high-potential areas as inefficient.

My study contributes to several strands of literature. I analyse the lasting impact of transport revolutions and thereby add to the large body of work devoted to identifying the economic returns to improving infrastructure systems (Donaldson and Hornbeck, 2016; Swisher, 2017; Donaldson, 2018; Burgess and Donaldson, 2012; Asher and Novosad, 2020; Faber, 2014; Baum-Snow et al., 2017). In contrast to these studies, I do not analyse the impact of existing transport revolutions, but rather measure how much a hypothetical first-best transport system would improve welfare (see also Santamaria, 2022; Kreindler et al., 2023).

Methodologically, I optimise over the space of networks in order to find the globally efficient transport system, harnessing the framework by Fajgelbaum and Schaal (2020) (see also Bernot et al., 2009; Galichon, 2016). Previous studies in economics relied on stepwise heuristics to eliminate suboptimal counterfactual networks but did not include a derivation of the globally optimal network design (Burgess et al., 2015; Alder, 2022).

I also contribute to the literature employing regional trade models to explain subnational welfare disparities caused by internal transport geography in a development context (Atkin and Donaldson, 2015; Storeygard, 2016; Coşar and Fajgelbaum, 2016; Fiorini et al., 2021; Gorton and Ianchovichina, 2022) as well as the literature of optimal spatial policies (Fajgelbaum and Gaubert, 2020).

In my empirical inquiry, I add to the literature examining long-run persistence of colonial transportation revolutions in Africa (Jedwab and Moradi, 2016; Jedwab et al., 2017), as well as add to our understanding of how regional and ethnic favouritism can explain such persistence (Burgess et al., 2015; Michalopoulos and Papaioannou, 2013, 2014, 2016; De Luca et al., 2018; Hodler and Raschky, 2014).

## 2 A model of optimal transport networks

In this paper, I derive the optimal goods trade network for every country in Africa. To be able to maximise over the space of networks, I harness a version of the framework proposed by Fajgelbaum and Schaal (2020), which I outline below. My main departures from their framework are the way I calibrate the model to be amenable to the African continent, which I explain in the section 3.

**Geography** Following the set-up and notation of Fajgelbaum and Schaal (2020), I consider a set of locations $\mathcal{I}$. Each location $i$ is endowed with a fixed total amount of (rival) non-tradeables $H_i$, such as housing, and a number of homogeneous consumers $L_i$. Each consumer has an identical set of preferences characterised by

$$u = c_i^{\alpha} h_i^{1-\alpha}$$

where $h_i = H_i/L_i$ denotes per capita housing. $c_i = C_i/L_i$ denotes per capita consumption of a CES aggregate over $N$ goods:

$$C_i = \left( \sum_{n=1}^{N} (C_i^n)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

where $\sigma$ denotes the standard elasticity of substitution and $C_i^n$ denotes the consumption of good $n$ in location $i$.

Every location produces one variety $n$, yet multiple locations might produce the same variety. I assume a linear production function with labor as the only input:

$$Y_i^n = z_i^n L_i$$

where $z_i^n$ is a location-specific productivity term for good $n$. In the calibration outlined below, I generally assume that only a few locations produce their own unique specialised good, while the majority of locations produce a homogeneous "agricultural good".

**Network topography** Locations represent nodes of an undirected network graph. Each location $i$ is directly connected to a set of neighbours $N(i)$. I consider locations to be arranged on a two-dimensional square lattice where each node is connected to its eight surrounding nodes to the north, north-east, and so on.

All goods except housing can be traded within the network. Let $Q_{i,k}^n$ denote the total flow of good $n$ travelling between nodes $i$ and $k \in N(i)$. While goods can only be traded between neighbouring nodes, nothing prevents them from travelling long distances through the network by passing multiple locations after each other. Sending goods from location $i$ to location $k \in N(i)$ incurs trade costs, which are modelled in the canonical iceberg form. I follow Fajgelbaum and Schaal and model iceberg trade costs for trading good $n$ between neighbouring locations $i$ and $k$ as

$$\tau_{i,k}^n(Q_{i,k}^n, I_{i,k}) = \delta_{i,k}^\tau \frac{(Q_{i,k}^n)^\beta}{I_{i,k}^\gamma} \tag{1}$$

where $I_{i,k}$ is defined as the level of *infrastructure* on the edge between nodes $i$ and $k$. More infrastructure on a given link decreases the cost of trading between them. Trade costs also depend on $Q_{i,k}^n$, the total flow of goods on the link. Higher existing trade volumes on a given edge make sending an additional good more costly, a dynamic Fajgelbaum and Schaal refer to as *congestion externality*.[1] The social planner realises this and takes congestion into account when determining optimal trade flows. $\delta_{i,k}^\tau$ is a scaling parameter allowing trade costs to be flexibly adjusted based on inherent characteristics of an $i, k$ location pair, such as distance.

In equilibrium, each location cannot consume and export more than it produced and imported. More formally

$$C_i^n + \sum_{k \in N(i)} Q_{i,k}^n(1 + \tau_{i,k}^n(Q_{i,k}^n, I_{i,k})) \leq Y_i^n + \sum_{j \in N(i)} Q_{j,i}^n \tag{2}$$

must hold for every $n$ and $i$.

I follow the contribution of the Fajgelbaum and Schaal (2020) framework and proceed to endogenise infrastructure provision $I_{i,k}$ in order to facilitate optimal trade flows. Analytically, this problem nests the static trade flow exercise outlined above. The social planner chooses an infrastructure network, and given the network proceeds to compute optimal trade flows subject to (2). To make the problem more interesting, I introduce a constraint on infrastructure. This is specified in fairly straightforward manner as the *Network Building Constraint*

$$\sum_i \sum_{k \in N(i)} \delta_{i,k}^I I_{i,k} \leq K \tag{3}$$

where $\delta_{i,k}^I$ denotes the cost of building infrastructure on the edge between nodes $i$ and $k$. Total spending on infrastructure is constrained by an exogeneous value $K$, representing the total budget of infrastructure spending available to the social planner.

[1]There are multiple ways to conceivably micro-found such an externality. For example, in traffic or queueing theory, inverse speed is a convex function of volume processed (see eg. Brancaccio et al., 2024, for a recent treatment). On the other hand, indivisibilities in transport infrastructure, such as a fixed container size, could also lead to opposite effects (see eg. Kreindler et al., 2023). A theory of optimal trade networks in the presence of such indivisibilities represents an exciting avenue for future work.

**Planner's problem and equilibrium** I consider two versions of spatial equilibrium in my model: with and without labor mobility. In the case without labor mobility, the social planner observes geography, population, endowments, technologies, and preferences and solves solves for the optimal transport network which induces trade flows leading to welfare-maximising consumption levels while respecting the *Network Building Constraint* (3). In the case with mobile labor, the planner additionally allows for free migration between locations until utility levels are perfectly equalised across space. The full planner's problem can hence be stated as

$$
\begin{aligned}
&\max_{\left\{\left\{C_i^n,\{Q_{i,k}^n\}_{k\in N(i)}\right\}_n,\; c_i,\{I_{i,k}\}_{k\in N(i)},L_i\right\}_i} && \sum_i L_i u(c_i,h_i) \\
&\text{subject to} && L_i c_i \le \left(\sum_{n=1}^{N}(C_i^n)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \forall i \in \mathcal{I} && \textsc{CES Consumption} \\
& && C_i^n + \sum_{k\in N(i)} Q_{i,k}^n(1+\tau_{i,k}^n(Q_{i,k}^n, I_{i,k})) \\
& && \qquad \le z_i^n L_i + \sum_{j\in N(i)} Q_{j,i}^n, \forall i \in \mathcal{I}, n \in \mathcal{N} && \textsc{Balanced Flows Constraint} \\
& && \sum_i \sum_{k\in N(i)} \delta_{i,k}^I I_{i,k} \le K && \textsc{Network Building Constraint} \\
& && I_{i,k} = I_{k,i}, \forall i \in \mathcal{I}, k \in N(i) && \textsc{Infrastructure Symmetry} \\
& && u(c_i,h_i) \ge u \quad \forall i \in \mathcal{I}. && \textsc{mobile labor} \\
& && C_i^n, c_i, Q_{i,k}^n, L_i \ge 0, I_{i,k}^u \ge I_{i,k} \ge I_{i,k}^\ell, \forall i \in \mathcal{I}, n \in \mathcal{N}, k \in N(i). && \textsc{Non-Negativity}^{2}
\end{aligned}
$$

In the case of mobile labor, this setup corresponds to standard notions of spatial or urban equilibrium, in which agents move to wherever their utility is highest, until there are no differentials left to exploit (Roback, 1982). The case with immobile labor is less standard in spatial models. There are many possible microfoundations rationalising utility differences across space as equilibrium outcomes. There might be hidden migration costs large enough to fully offset any gains from moving (Allen et al., 2019; Porcher, 2021), borrowing constraints preventing otherwise beneficial moves (Bilal and Rossi-Hansberg, 2021), or idiosyncratic location-preference shocks powerful enough to sustain utility differentials (Gaubert et al., 2020; Fajgelbaum and Gaubert, 2020). In this case, the social planner can identify depressed areas and implement optimal networks as a place based policy.

While optimising over the space of networks might appear daunting, Fajgelbaum and Schaal (2020) provide conditions under which deriving the unique spatial optimum is both ensured and feasible. Instead of solving for every single infrastructure link, I follow the authors and recast the problem in its dual representation, which greatly reduces the dimensionality of the problem (see appendix section B for technical details).

**Optimal reallocation and expansion** Similar to Fajgelbaum and Schaal (2020), I choose two different values for the total infrastructure budget available to the social planner $K$, aimed at identifying two different sources of spatial improvements to the trade network.

In the first exercise, which I call *optimal reallocation*, I set $K$ equal to the cost of the *current network*,

[2] $I_{i,k}^\ell$ represents a lower-bound for infrastructure. As will be discussed in chapter 3, I calibrate the reallocation version of the model with $I_{i,k}^\ell = 4$, and the expansion version of the model with $I_{i,k}^\ell = I_{i,k}^c$. Also note that in the model without mobile labor, the planner cannot optimise over the vector $L_i$ by definition. In this case, I also drop the utility equalisation condition $u(c_i,h_i) \ge u \quad \forall i \in \mathcal{I}$.

ie. $K = \sum_i \sum_{k \in N(i)} \delta^I_{i,k} I^c_{i,k}$, where $I^c_{i,k}$ is the *current* level of infrastructure on a link. I observe the current network, infer how much it must have cost to build it, and bind total infrastructure spending by that amount. In this exercise, the planner is otherwise free to choose her optimal level of $I_{i,k}$ on each link, as long as it complies with the budget constraint. In essence, this scenario amounts to a reallocation exercise: the planner gets to redesign the entire network, freely lifting infrastructure off the ground, shuffling it around, and reallocating it in an optimal way.

I argue that the amount of welfare an economy can gain from this reallocation exercise serves as an intuitive measure for how inefficient the current infrastructure network was to begin with. If the current network already connects the most important trade routes with fast roads without much congestion, then welfare gains from reallocating roads optimally will be small. On the other hand, if the infrastructure network connects only secondary locations while leaving major trade hubs unconnected, then an optimal reallocation could lead to large welfare benefits.

This reallocation exercise also comes with two clear drawbacks. For one, many roads in Africa were built decades ago and financed first by colonial powers and later by international organisations. Hence, fixing $K$ at the value of the current road network will likely overestimate the cost of building the network at the time. Second, allowing the planner to lift roads off the ground and move them seamlessly throughout the country is practically infeasible. In reality, demolishing a road in one part of the country does not free up resources to build a road of equal quality somewhere else (even worse, there are likely labor, capital, and political costs associated with this). I hence see this reallocation scenario as a hypothetical exercise aimed at quantifying which level of welfare could be achieved if the same raw quantity of roads were allocated optimally. To address these drawbacks, I also conduct a second exercise which I call *optimal expansion*. In this scenario, $K$ is set to 10% more than the current level of infrastructure ($K = 1.1 \times \sum_i \sum_{k \in N(i)} \delta^I_{i,k} I^c_{i,k}$), while the current network serves as a lower bound to infrastructure investment $I^\ell_{i,k} = I_{i,k}$. In other words, this simulation allows the planner to make additional investments worth 10% of the current road stock in the country, but she cannot reallocate existing infrastructure from one place to the next. This is an arguably much more policy-relevant exercise in the spirit of a "big push" infrastructure building campaign (Murphy et al., 1989; Kline and Moretti, 2014; Buera et al., 2023), and aims to quantify how high the returns to optimally placed additional infrastructure investments are.

# 3 Calibration of current and optimal trade network designs

To calibrate the topography of economic activity and trade in all African countries, I construct a novel network representation covering the entire continent using a variety of data sources.

## 3.1 Network calibration

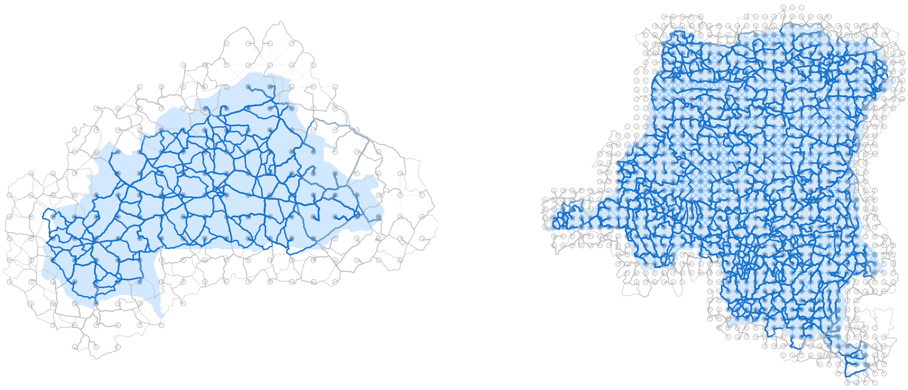
I first divide the entire continent into grid cells of 0.5 degrees latitude by 0.5 degrees longitude (roughly 55 by 55 kilometres at the equator). For all of Africa, this amounts to 10,167 cells. I calibrate population $L_i$ with data on 2015 population totals from the *Gridded Population of the World* dataset (2016).[3] To proxy for heterogeneity in economic activity over space, I rely on the established practise

[3]This NASA-funded project gathers data from hundreds of local census bureaus and statistical agencies in order to construct a consistent high-resolution spatial dataset of the world's population. When a datasource only reports population totals for large, higher-level administrative districts, the dataset smoothes population uniformly over the entire area. GPW does not employ any auxiliary data sources – like satellite data – to weight-adjust population totals over space (Doxsey-Whitfield et al., 2015). Africa is the continent with the coarsest resolution of administrative input data. However, the average coverage of $(57\text{KM})^2$ neatly matches the grid cell resolution of my study. On average, a cell is home to 110,000 people (median 25,000). The most populous cell contains Cairo and inhabits almost 18 million people. 212 cells are uninhabited.

Figure 1: Road networks for different countries as scanned off OSM

(a) Burkina Faso

(b) DRC



*Road networks as scanned off Open Street Map (OSM). Blue lines represent routes from each grid cell centroid to each of its eight surrounding neighbours. These routes may include a portion walked by foot in order to get to the nearest street. Grey lines are connections abroad. Connections in which walking the entire distance is faster are printed as thin lines. Data scanned in July 2023.*

of using satellite imagery of light intensity at night (Henderson et al., 2012). Data on 2010 night luminosity come from Henderson et al. (2018) and are also aggregated onto my study's 0.5 × 0.5 degree grid resolution to form output $Y_i$, and together with population, an implied productivity $z_i$.

**Infrastructure.** To quantify the degree to which network nodes are connected to each other, I make use of the open source routing service OPEN STREET MAP (OSM). The OSM routing algorithm is specified for cars and takes into account differential speeds attainable on different types of roads. For every centroid location, I scan OSM for the optimal route to each of their respective eight surrounding neighbours (or less for coastal grid cells). For all of the resulting almost 75,000 routes, I gather distance travelled, average speed, and step-by-step coordinates of the travel path.[4,5] For some particularly remote areas, the nearest street is very far away, such that the car routing provided by OSM is not sensible. To counter these cases, I also calculate the outside option of walking the entire link in a straight line at 4 km/h and replace OSM's route with walking in cases where this is faster.

My study is concerned with optimally reallocated or expanded domestic road networks for each country in Africa. I hence divide the entire grid along national borders. Reoptimising roads solely within a country's borders comes at the risk of undervaluing roads built primarily for international trade. A highway connecting to an important trading post or port just beyond a country's borders might look inefficient to a social planner who is only given national data. I hence create a buffer of 120km around each country and allow the social planner to take these border regions into account when computing the optimal network.

Figure 1 presents the resulting road networks for two countries and their border buffer (in grey). Figure 1a displays every OSM connection for Burkina Faso, which appears overall fairly well connected. Connections in which walking were the preferred alternative are displayed in thin straight lines and fairly rare. Figure 1b presents the case for the DRC, which displays a clear lack of infrastructure in the middle of the country.

[4] Scans of OSM were conducted in July 2023. The service does not allow a retrospective scan over past road databases, so a time difference between lights (2010), population (2015), and roads (2023) can not be overcome.

[5] If either start or destination location do not directly fall onto a street, the optimal route jumps to the nearest road and goes from there. To take this into account, I add a walking distance to the travel path. Agents are assumed to walk in straight lines to the nearest street at a fixed speed of 4 km/h. They then take the car and drive the route with average speed as specified by OSM, before they potentially have to walk the last stretch again to their exact centroid destination.

I use the average attainable speed between locations according to the OSM algorithm as a proxy for the quality of current infrastructure on the edge between them. If two locations are linked by a faster connection, I assume this to be the result of higher infrastructure $I_{i,k}$ on this edge. I hence set $I_{i,k} = \text{Average Speed}_{i,k}$. This measure is naturally bound from below at 4 km/h, as walking is always available as a backup. Empirically, average speeds range between 6 km/h (Mauritania, where most routes go through the desert and have to be covered by walking) and 33 km/h (Swaziland).

Relying on the open source community of OSM does come with some drawbacks. The most pressing concern is that data on the position and quality of roads are user-generated and hence subject to reporting bias. Richer areas may appear to be equipped with more roads if local residents have the time and necessary access to a computer to enter their neighbourhoods into the database. While this is certainly troubling, I believe this bias to be much more important on finer resolutions than the operating one in this study. Start and destination of the elicited routes are on average more than 55 kilometres apart and travel will hence take place mostly on larger roads and national highways. It is unlikely that these major streets are systematically underreported in OSM, the primary open source routing platform on the internet. To nevertheless get a sense of the magnitude of this potential bias, for a random 1% subset of connections I additionally scrape GOOGLE MAPS (GM). While GM might in principle also suffer from the above bias, its status as the biggest commercial provider of routing services in the world makes it a useful benchmark. Figure A.1 in the appendix compares the two providers. As can be seen in panel (a), the two services generally agree on speed magnitudes, with GM offering slightly faster connections. Panel (b) reveals that this difference is particularly pronounced in low population areas, confirming the above suspicions – connections between the lowest density areas are about 50% faster in GM than in OSM, potentially due to GM knowing backroads that haven't been user-reported to OSM. Above a grid cell population of about 50,000, the difference between both providers becomes statistically indistinguishable. In my empirical investigation below, I try to account for this bias by controlling for grid cell population levels.

**Trade costs.** To parameterise iceberg trade costs defined in equation (1), I use a variety of sources. As documented in more detail in Appendix section C, I leverage estimates of the costs of road delays to African truckers by Teravaninthorn and Raballand (2009) to set the infrastructure elasticity $\gamma$ to 0.946. I also use data on the cost of traffic congestion by Wang et al. (2011) to calibrate the congestion elasticity to $\beta = 1.1774$.

To calibrate the inherent component of trade costs $\delta^{\tau}_{i,k}$, I incorporate existing evidence on the impact of distance on price dispersion in Africa. Specifically, I follow Atkin and Donaldson (2015), who posit that price differences of the same good between locations $i, k$ are a composite of transport costs and markups charged by intermediaries with market power. Market power is a first-order determinant of trade costs in Africa, where entry barriers and burdensome regulation enable the formation of large cartels operating along the major transport corridors of the continent and distorting prices (Teravaninthorn and Raballand, 2009). In light of this, I transform the basic Fajgelbaum and Schaal (2020) framework to be able to incorporate the impacts of market power on prices across space.

In particular, Atkin and Donaldson (2015) estimate that markups charged by intermediaries *decline* with distance, a counter-intuitive fact explained by consumers in remote areas tending to be poorer and hence moving along a more elastic part of their demand curve. At the same time, transport costs intuitively increase with distance travelled. As the Fajgelbaum and Schaal framework I build includes additional, endogeneous components of travel costs (namely infrastructure and endogeneous congestion), I treat the Atkin and Donaldson estimates as *equilibrium outcomes* and run a fixed-point routine which induces their reduced-form relationship *in equilibrium*. As explained in more detail in Appendix C, I arrive at the calibration $\delta^{\tau}_{i,k} = 0.1159 \times \ln(\text{Distance in miles}_{i,k})$ for the trade cost elasticity to distance travelled.[6]

$\delta^{I}_{i,k}$ from equation (3) denotes the constant cost of increasing the average speed on a given link by one. I follow Fajgelbaum and Schaal who in turn make use of a recent study by Collier et al. (2015), which estimates infrastructure building costs in developing countries. Readily applying their specification, I calculate $\ln(\delta^{I}_{i,k}) = 0.12 \times \ln(\text{Ruggedness}_{i,k}) + \ln(\text{Distance}_{i,k})$ as the cost of increasing infrastructure $I_{i,k}$ on the link between $i$ and $k$.[7]

**Varieties.** To build incentives for trade, I introduce $N = 6$ different varieties. The four most populous grid cells of a given country are assumed to be producing their own variety, which creates incentives for trade between major cities. Another "international" variety is supplied by the three most populous grid cells within each country's border buffer. This ensures incentives for international trade beyond the border. I also collect data from Lloyd's List on grid cells which are home to a major international port.[8] Every port location (within a country's borders or within its international buffer) not covered by the previous varieties is assumed to also produce the "international" variety. Lastly, every other location is assumed to produce a sixth, "agricultural" variety.

Finally, as explained in more detail in Appendix section C, I calibrate the budget share of tradeables as $\alpha = 0.7$ following Porteous (2022) who use data on consumption spending from Angola and Nigeria, and the trade elasticity of substitution as $\sigma = 5$ following Atkin and Donaldson (2022).

**Local non-tradeables.** Local non-tradeables such as housing in each location are unobserved and so need to be calibrated. I distinguish between the cases of mobile and immobile labor. In the case of immobile labor, housing levels make no difference (other than monotonically shifting residents' utility levels) and so I set $H_i = L_i$ in each location.

In the case of mobile labor, however, local non-tradeable stocks codetermine spatial equilibrium. In this case, I invert the model to find the vector of housing $H$ that rationalises the *observed* distribution of population as spatial equilibrium, given the current road network $I^c$. Column 5 of Table A.1 in the Appendix correlates the level of (per-capita) housing with a host of geographic measures and finds few significant correlations. Only the amount of night lights (ie. production) is negatively associated with local amenities. This could be because such places have more direct access to consumption, and to restore spatial equilibrium, these regions have lower housing per capita.

Local non-tradeables also introduce a notion of congestion externalities *within locations*, which are distinct from the previously introduced trade-congestion on edges. Since the total stock of housing is fixed and rival, migration to a given region (eg. induced by an improved road network) leads to competition for those scarce resources (Redding, 2016). The social planner takes this into account, and the population distribution induced by an optimal network with mobile labor will balance those disagglomeration forces with local network improvements.

[6] Allowing both transport costs $t$ as well as markups $\mu$ to vary with distance $d_{i,k}$, the authors write: $P_k - P_i = t_{j,k}(d_{i,k}) + \mu_{i,k}(d_{i,k}) = \widehat{\xi} \log(d_{i,k})$ and estimate $\widehat{\xi} = 0.0251$. This corresponds to equation (2) of Atkin and Donaldson (2015), who have barcode-level data on identical goods for Ethiopia and Nigeria. They estimate $\widehat{\xi} = 0.0248$ for Ethiopia and $\widehat{\xi} = 0.0254$ for Nigeria. As explained in more detail in appendix section C, my fixed-point routine translates these parameters into 0.2258 for Ethiopia and 0.006 for Nigeria. I take the average of those two.

[7] $\text{Distance}_{i,k}$ denotes the road distance travelled between nodes and enters positively, implying that longer roads are costlier to develop as every single road kilometre will have to be improved. The term $\text{Ruggedness}_{i,k,c}$ denotes the average ruggedness between grid cells $i$ and $k$ and enters positively, highlighting the additional expenses accompanied with building on uneven terrain. Data on local ruggedness come from Henderson et al. (2018) and is described in more detail with other geographical covariates below.

[8] I use the open-access portal at https://directories.lloydslist.com/port and hand-code the locations of the 90 biggest ports in Africa. Figure A.6a prints the resulting locations.

### 3.2 Trade network optimisation

For each country, I conduct four simulations: optimal reallocation and expansion, with and without labor mobility. As mentioned above, the reallocation scenario binds total infrastructure spending $K$ at the level of the *current* road network, with the social planner free to reshuffle roads within the country in order to improve connections as she chooses. This reallocation exercise does not seek to identify where to place the optimal next investment, but rather represents a purely hypothetical scenario in which every roads can be freely reshuffled.[9]

For the expansion scenario, the planner treats the current network as fixed, but gets to make new investments worth 10% of the current infrastructure stock. This scenario is closer to the policy-relevant question of identifying the highest-return investments. While both exercises aim to quantify the extend of infrastructure misallocation in a country, the expansion scenario is more directly addressing the "too few roads" hypothesis, while the reallocation scenario speaks to claim of roads being "in the wrong place".

When conducting these optimisations, I fix infrastructure in the international buffer around each country at their current level. While the planner takes into account the economic geography of a country and its surroundings, she can only reallocate infrastructure within the country in question.[10] I also impose a maximum speed of 120 km/h on any given link.[11]

For a small number of links (less that 0.1% of all connections), the OSM algorithm is unable to find any available route between locations. This is mostly due to some obvious geographical barrier, like an island separated from the mainland without a bridge, or a non-navigable jungle or mountain range. In these rare cases, I treat these locations as if they were not connected at all and also don't allow the reallocation scenario to connect these previously separated regions. This is meant to forbid the social planner from building infrastructure on obviously impassable terrain.

Should we generally expect the social planner to see the need to redistribute infrastructure across space? An influential series of papers argues that in standard spatial equilibrium models, the observed market equilibrium is efficient even in the presence of agglomeration externalities, and hence any reallocation of productive resources (eg. through place-based policy or local hiring subsidies) is zero-sum (Glaeser and Gottlieb, 2008; Kline and Moretti, 2013, 2014). However, the problem of allocating road infrastructure is different from the problem studied in those papers in two fundamental aspects. First, infrastructure is generally not competitively provided (while a competitive labor market will induce workers to relocate to the place where their expected value is highest, units of infrastructure do not freely move to the link where they can be most impactful). As such, infrastructure investments (which are notoriously sticky) might be the result of past optimisation problems, regional favouritism, or other imperfect spatial policies. Second, my model features con-

[9]Note that equation (3) only fixes $\sum_i \sum_{k \in N(i)} \delta^I_{i,k} I_{i,k} = K$. Hence, not the overall sum of infrastructure is fixed, but more precisely the overall cost of infrastructure. This still allows the social planner to take away one unit of infrastructure on a very expensive (high $\delta^I_{i,k}$) link and exchange it for much more than one unit on a cheaper (low $\delta^I_{i,k}$) link.

[10]There are two reasons why I conduct the simulation procedure within countries and not over the entire African continent. One is computational; the requirements for numerically solving the model increase quadratically in the number of locations $\mathcal{I}$. The largest country in Africa (Algeria) is made up of almost 900 locations and already strains computing power quite heavily. Simulating all of Africa's 10,000+ locations at once is then almost unattainable with available technology. The second reason is interpretational; while lifting a country's roads from the ground and flexibly reshuffling them across the nation is already a fictitious scenario, it still operates within a government transport authority's locus of control. Regions disadvantaged by their own government can reasonably be considered discriminated against. This is less the case if one were to optimise over the entire continent. Without a central planning body for all of Africa, it is hard to interpret why a road in e.g. Tunisia should rather be moved into Namibia.

[11]In the reallocation exercise, I also bind the social planner's set of permissible roads from below at 4 km/h. This is motivated by the assumption at the beginning that walking straight lines at this speed is an outside option and always available to any traveler. As discussed above, in the expansion exercise, the current observed network $I^c$ serves as the lower bound, as only new roads can be constructed in this scenario.

gestion externalities in trade costs, so that even in the case with mobile labor, people endogeneously form a spatial population distribution that might put heavy strain on the network. As Fajgelbaum and Gaubert (2020) show, the presence of congestion opens the door to welfare-improving transfers, such as the infrastructure reallocation considered here. With this in mind, it should not be surprising if the reallocation scenario uncovers substantial hypothetical welfare gains.

I conduct the reallocation and expansion scenarios for every African country separately.[12] Optimisations are performed using the optimisation toolkit provided by Fajgelbaum and Schaal (2020). Figure 2 visualises this reallocation exercise for the two countries from above. Subfigure 2a displays the discretised network representation of Burkina Faso. The edges to this network are printed almost evenly thick, implying that infrastructure is fairly evenly distributed across the country. Subfigure 2b then displays the country after the network reshuffling exercise with immobile labor. Three patterns stand out. First, the social planner sees a clear need to connect the populous areas in the center of the country with each other. For that, the social planner is willing to salvage some of the apparently less important infrastructure in the north of the country. Second, there still is a benefit to having a few trails connecting the center with a regional hub in the south-west producing it's own variety. Third, nodes are printed in a colour scale corresponding to individual welfare gains and losses for each location. As can be seen from first-glance, most southern regions (brighter colors) stand to gain from this scenario, while the big cities on average seem to lose (darker colors). Subfigure 2c prints results from the optimal expansion scenario. After optimally investing a budget worth 10% of its current infrastructure stock, the country's network gets improved along similar edges as under reallocation.

The simulation for the Democratic Republic of Congo in Figures 2d – 2f are made under the assumption of mobile labor. The social planner sees a need to better connect the center of the country to its surroundings and the populous border regions. This is in line with the common perception of DRC's periphery being notoriously poorly connected to the centers of power and commerce. In these Figures, nodes are colored based on their population changes compared to baseline (as welfare changes by definition are the same in all locations). Large parts in the southern center of the country gain population in this scenario, while many border regions lose out.

## 4 A measure of spatial transport network inefficiency

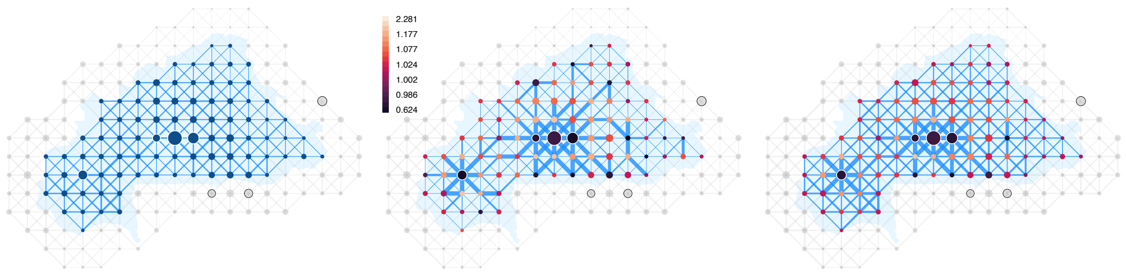
After successfully reshuffling a country's transport network, national welfare will by construction (weakly) increase. Such welfare gains are mainly caused by enabling mutual benefits from trade through connecting the right locations. In the case of mobile labor, there is a potentially added effect due to increased productivity, as better transportation networks might induce people to migrate to more productive locations. With immobile labor, overall production is fixed by construction. Nevertheless, welfare gains from reallocation are not negligible. Burkina Faso, for instance, stands to gain 3.4% from optimal reallocation and 2.5% from optimal expansion under immobile labor. Under mobile labor, these gains would be even higher (3.6% and 2.8%).

Figure 3b reports hypothetical welfare gains for all African countries under immobile labor. Bars represent gains from optimal reallocation, crossed lines print gains from optimal expansion. Some nations like South Africa (0.5% welfare gains from reallocation) or Egypt (0.6%) perform relatively well and don't leave much room for improvement through reallocation. Many countries are leaving much more on the table, like Somalia (6.6%) or Madagascar (6.3%). South Sudan has the third-most

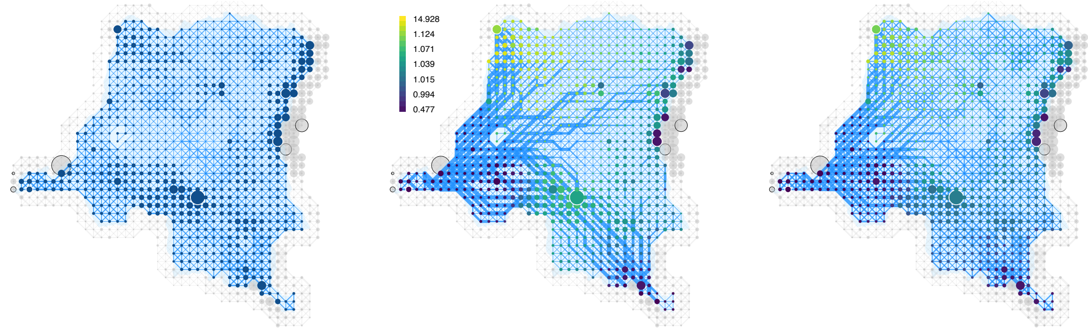
[12] Six small countries (Cape Verde, Comoros, The Gambia, Mauritius, São Tomé and Príncipe, and Reunion) are too small to form a sensible network as they only show up as a single location in the dataset and are henceforth no longer considered.

Figure 2: Reallocation scenario for different countries

(a) Burkina Faso ex-ante (b) Reallocation (immobile labor) (c) 10% expansion (immobile labor)




(d) DRC ex-ante (e) Reallocation (mobile labor) (f) 10% expansion (mobile labor)




*Results from optimally reshuffling roads in three African countries. In each network graph, every node represents a grid cell centroid location with radius proportional to the size of its local population. Edges are drawn thicker depending on their allotted infrastructure $I_{i,k}$. In the optimal networks on the right, nodes are coloured based on their relative welfare gains and losses (population gains and losses for the case with mobile labor), with more light areas gaining more. Color scheme is the same as in Figure 3a and 3c.*

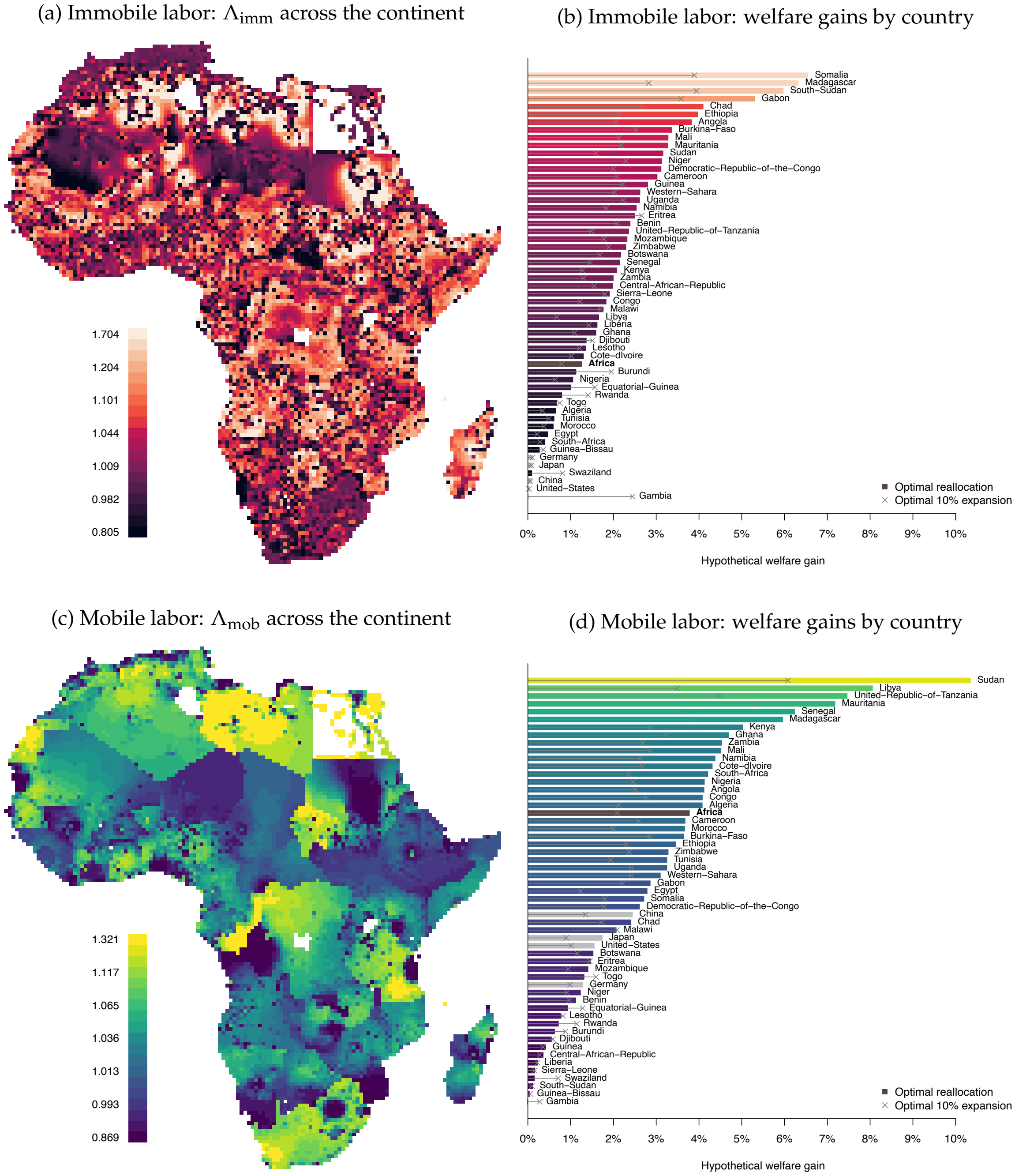
inefficient network and represents a constructive case. Its citizens stand to gain 6% of overall welfare if just their roads were better placed. This may come as no surprise, as the world's newest country has largely inherited a road network that was not conceived to sustain an independent nation, but rather connect it to its former capital up north. For the entire continent, optimal reallocation of national road systems under immobile labor would improve overall welfare by 1.3%.

Results are similar for optimal expansion. Eyeballing Figure 3b confirms that countries that would gain a lot from optimal reallocation would also gain from an ideally placed new investment. For all of Africa, welfare gains from an optimal infrastructure program equivalent to 10% of its current road system would yield 0.8% of welfare, implying a return on investment of around 8%. Comparing this to the 1.3% welfare gains from optimal reallocation, it also implies that the drag on welfare through spatial misallocation is even bigger than what could be achieved under an optimally targeted, large-scale infrastructure program worth 10% of Africa's total road stock. According to my model, spatial inefficiencies in Africa's trade network are hence a major source of welfare losses on the continent.

Figure 3d repeats the same exercise in the case of mobile labor and finds on average 2.5-3.0× larger welfare gains than with no labor mobility. Sudan would gain over 10% of welfare from reallocation (6% from expansion), all of Africa combined stands to gain 3.8% (2.1%). Scenarios with mobile and immobile labor also produce roughly similar rankings across countries.

To gain a sense of how African countries compare to the rest of the world, I also compute my measure for four non-African countries: China, Japan, Germany, and the US (printed in grey). In the case of immobile labor, these countries all are found to have an order of magnitude more efficient

Figure 3: Africa by network inefficiency




*Discrimination index at the grid cell level (a) and welfare gain under the optimal reallocation counterfactual for each country (b). Country gains are computed by comparing the population-weighted mean of the discrimination index of all cells in a country. (c-d) repeat this exercise in the case with mobile labor. Grid cells with zero population are printed blank.*

road networks than most African countries. The United States, for example, only stands to gain between 0.02-0.03% from optimal expansion and reallocation under immobile labor. Estimates with mobile labor are slightly bigger, yet all four non-African countries still consistently rank among the countries with the most efficient networks.[13]

Forgone welfare gains can be conceived as an intuitive measure for overall network inefficiency. While each country only stands to gain overall welfare from the reallocation procedure, individual

[13] I calibrate these four non-African countries with the same data sources described above. Computations for large China and the US are performed on a 4× coarser sample (1×1 degree grid cells) for computational reasons.

locations might very well lose in the process. Intuitively, some regions might be equipped with far too many good roads such that the social planner takes these roads away to use someplace else. Comparing each grid cell's welfare before and after the major reshuffling or expansion can help to identify regions which are currently over or underprovided for. More formally, I define

$$\Lambda_{\text{imm},i} = \frac{\text{Welfare under the optimal infrastructure}_i}{\text{Welfare under the current infrastructure}_i} \tag{5}$$

as the *Local Infrastructure Discrimination Index* with immobile labor for grid cell $i$. Areas with high $\Lambda_{\text{imm}}$ scores would be gaining under the optimal reallocation scenario and are hence underprovided for in the network's current state. A low score of $\Lambda_{\text{imm}}$ on the other hand, implies that a region is *too* well off given its position in the network today and hence should be stripped off some of its infrastructure to increase overall welfare.

Figure 3a displays the spatial distribution of $\Lambda_{\text{imm}}$ over all 10,000+ grid cells of the entire continent. The darker a grid cell's shade, the more it is advantaged by the inefficiencies of the current network. When interpreting this map, note that grid cells are undergoing the reshuffling scenario solely within their respective country (while taking into account respective international buffer regions). National borders hence play a role and can at times even clearly be inferred from the printed map. Keeping this in mind, the map reveals substantial spatial variation in the index across the African continent. The luckiest region stands to loose almost 70% of total welfare if the fictitious social planner intervened and reshuffled roads away from it. On the other extreme of the spectrum, the residents of the most discriminated grid cell are missing out on a welfare hike of almost 65%.

The definition of $\Lambda_{\text{imm}}$ naturally translates to the case of optimal expansion. I denote the ratio of welfare before and after the expansion scenario by $\Lambda^{10\%}_{\text{imm}}$. Empirically, this measure ranges from 0.82 to 1.52. Note that it is possible for locations to lose welfare (ie. have $\Lambda^{10\%}_{\text{imm}} < 1$), even under this scenario where only new routes are added. Figure A.2 in the Appendix presents an intuitive example for how this might happen because of optimal re-routing of previous bottlenecks. Figure A.3 in the Appendix, furthermore, shows that the social planner generally targets the same regions with both expansion and reallocation scenarios. $\Lambda_{\text{imm}}$ and $\Lambda^{10\%}_{\text{imm}}$ are highly correlated ($\rho \approx 0.93$) across locations.

I extend my definition of local infrastructure discrimination to the case of mobile labor. Since welfare gains are by definition equal for every location, I instead focus on population movements:

$$\Lambda_{\text{mob},i} = \frac{\text{Population under the optimal infrastructure}_i}{\text{Population under the current infrastructure}_i}$$

Intuitively, locations which would see large population increases in response to infrastructure reallocation are underprovided for by the current network. Figure 3c presents a map of $\Lambda_{\text{mob}}$. This definition also extends naturally to the case of optimal expansion, $\Lambda^{10\%}_{\text{mob}}$.

Table A.1 in the appendix presents correlations of the various $\Lambda$ indices with features of Africa's geography. My reallocation or expansion simulations do not systematically harm or benefit more hot, agriculturally suitable, or rugged locations, or grid cells containing a national capital or with higher prevalence for malaria. Without labor mobility, they also don't differentially affect more populous areas. With labor mobility, there is slight evidence (albeit small) of people leaving more crowded areas. Across all simulations, more productive areas (proxied by night lights) are associated with welfare and population drops, as the social planner finds it optimal to benefit economically less productive regions. Places close to national borders or navigable rivers and lakes are hurt by this exercise, as the model is not focusing on long-haul international or multi-modal trade.

In interpreting $\Lambda_i$, keep in mind that this is a measure of differences in welfare or population. It need not be a direct mapping of changes in actual infrastructure provision. Figure A.4 in the Appendix plots the various $\Lambda$ measures against other network measures. All $\Lambda$ measures are strongly related to a notion of market access (Donaldson and Hornbeck, 2016), highlighting that more efficient networks are succeeding in connecting populous regions with each other.[14] Furthermore, the measures only barely (and often negatively) correlate with the change in actual roads in a given location, reflecting the non-linear nature of optimal network design. A region might substantially profit from the optimal policy, even though it is not directly granted additional roads.

Lastly, the measures I propose also need not correlate with existing infrastructure stocks. Optimal networks might instead very well feature an unequal distribution of roads across space, if places with lots of infrastructure represent important regional bottlenecks. Across my sample the various measures of $\Lambda$ do negatively vary with existing infrastructure levels (implying that the planner does tend to help places with currently low levels of infrastructure), but counter examples in both directions exist. My measure hence looks beyond the question of *how many* roads there are, and instead asks whether roads are *in the right place*.

# 5 Persistent effects of colonial railways on Africa's trade network

Why are some African roads not in the right place to promote beneficial trade? A key candidate explanation is lasting impacts of infrastructure investments from colonial times. Between 1890 and 1960, British, French, Belgian, German, Italian, and Portuguese administrations undertook efforts to permeate their colonies with more or less expansive railway networks (Jedwab and Moradi, 2016).

There were two main motivations for this: supporting extractive economies and ensuring military domination (Jedwab et al., 2017). These colonial railroads have been found to have a persistent impact on the spatial organisation of economic activity today. Jedwab and Moradi (2016) show how urbanisation started to center around railway tracks in the decades following their construction. Even as most railway lines have fallen into disarray and road traffic has replaced trains as the most important means of transportation, economic activity today still clusters in places close to the former rail lines. Figure 4 confirms these findings on the grid cell level: in pink, it prints how *current* levels of infrastructure are associated with distance to the colonial railroads.[15] This hints at a form of *complementarity* of railroads and roads: grid cells within 10 kilometres to the nearest colonial railroad have about 0.5 standard deviations more infrastructure than those further away. There is evidence of a declining gradient: the further a grid cell is away from colonial railroads, the less roads it still has today. Figure 4 also gives suggestive evidence that there would be welfare gains from restructuring this imbalance: in blue, it prints coefficients of an equivalent regression, yet putting the infrastructure discrimination index $\Lambda_{\text{imm}}$ on the left-hand side. As can be seen, the closer a grid cell to a century-old railway line, the *smaller* the index, implying the social planner seeks to reallocate infrastructure in a way to hurt these locations.

[14] This is defined as $MA_i = \sum_j (1+\tau_{ij})^{-\sigma} z_j L_j$ where $\tau_{ij}$ corresponds to the cost of shipping a quantity of $Q = 1$ over the network. I compute the change in $MA$ from the static network before re-allocation to the optimal one post re-allocation.

[15] In particular, I run a regression on the grid-cell level $i$:

$$I_i = \beta_0 + \sum_{r=10}^{100} \beta_r \text{Distance}(r \text{ KM}; (r-10) \text{ KM})_i + \epsilon_i$$

where $I_i$ is the (z-scored) total stock of infrastructure (average OSM speed) in a given place, and $\text{Distance}(r \text{ KM}; (r-10) \text{ KM})_i$ is an indicator for a colonial railroad to pass within $r$ to $r-10$ kilometres of a grid cell centroid. The ommited category are grid cells further than 100KM from the nearest colonial railroad. Data on the spatial outlay of colonial railroads come from Jedwab and Moradi (2016).

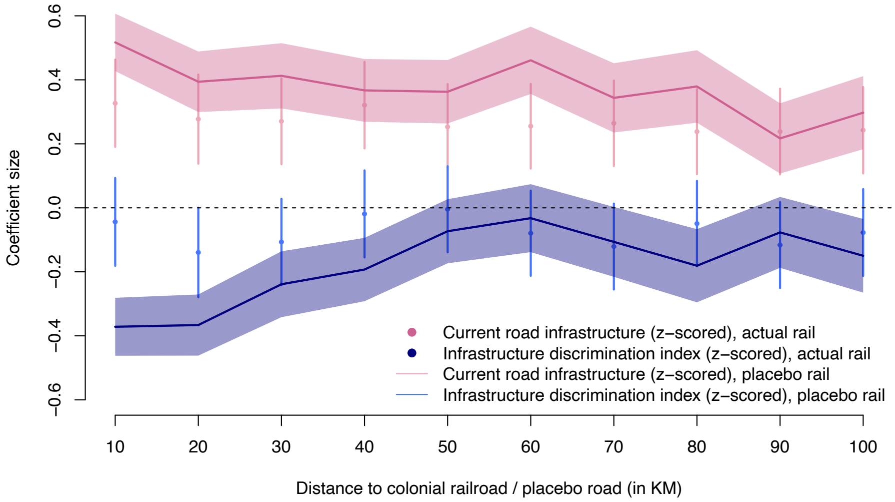



Figure 4: Distance to railroad and current infrastructure and $\Lambda$

*Estimates of a regression of the type $Y_i = \beta_0 + \sum_{r=10}^{100} \beta_r Distance(r\ KM; (r-10)\ KM)_i + \epsilon_i$ where $\beta_r$ represents the effect of being between r and $(r-10)$ kilometres away from either a realised colonial railroad (line and shaded area) or a placebo railroad (points and vertical bars). $Y_i$ represents either current road infrastructure density of a cell i (in pink, this is the average of all $I_{i,j}$ values across neighbors $j \in N(i)$, or infrastructure discrimination $\Lambda_{imm}$ (in blue). Both dependent variables are z-scored before the analysis. 95% confidence intervals plotted around coefficients.*

**Econometric strategy.** I complement the above suggestive evidence with a more formal empirical specification. In particular, I employ the various definitions of $\Lambda_i$ as dependent variable in a standard OLS regression set-up. In the base specification, I estimate

$$\tilde{\Lambda}_{i,c} = \beta \text{Rail}_{i,c} + \mathbf{X}_{i,c}\gamma + \delta_c + \epsilon_{i,c} \tag{6}$$

for grid cell $i$ in country $c$. $\text{Rail}_i$ denotes different indicators of the presence of a colonial railroads in a grid-cell, using data compiled by Jedwab and Moradi (2016). I also add information on colonial roads in South Africa which I digitised myself using maps from Herranz-Loncán and Fourie (2017). I compute the total number of kilometres of different colonial railroad classes crossing a grid cell, as well as the distance of a cell's centroid to the nearest railroad line. $\delta_c$ denotes country fixed effects, $\mathbf{X}'_{i,c}$ is a vector of controls, and $\beta$ is the coefficient of interest. The dependent variable $\tilde{\Lambda}_{i,c}$ corresponds to the z-scored transformation of $\Lambda_{ic}$ (ie. $\tilde{\Lambda}_{i,c} = (\Lambda_{ic} - \mu_\Lambda)/\sigma_\Lambda$). As a baseline outcome, I focus on $\Lambda_{\text{imm}}$ stemming from the optimal reallocation scenario with immobile labor. Conclusions are very similar when focusing on the case with mobile labor or the optimal expansion scenarios (see Tables A.3 and A.4 in the Appendix). To account for spatial autocorrelation, I follow Bester et al. (2011) and construct a higher-level spatial grid of 3 degrees latitude by 3 degrees longitude and cluster standard errors within each of these higher-level grid cells. Country fixed effects make observations comparable internationally by accounting for the fact that reallocation is performed for each country individually. The vector of controls $\mathbf{X}'_{i,c}$ contains a series of observable geographic, economic, and political characteristics.[16]

The colonial authorities didn't place railroads randomly across space. To gain more confidence

[16]Data on controls come from Henderson et al. (2018). In particular, I include in $\mathbf{X}$ each grid cell's average altitude, average temperature and precipitation, land suitability for agriculture, length of the annual growing period, an index for the stability of malaria transmission, indicators for the twelve predominant vegetation regions, and indicators for whether a grid cell's centroid is within 25 kilometres of a natural harbour, big lake, or navigable river respectively. I also add fourth-order polynomials of a cells latitude and longitude. Lastly, I add information on a cell's population, night lights, and ruggedness. Note that these latter three variables were already used in a non-linear way to compute a country's optimal trade network above. Results are robust to excluding those "simulation controls" (available upon request).

in the causal nature of my estimates, I follow Donaldson (2018); Jedwab and Moradi (2016) and also run all my estimations on a sample of railroad lines that the colonial authorities planned, but never built. These "placebo railroads" serve as a plausible control group if they were selected based on similar characteristics at the time (such as growth potential or strategic importance), but ended up not being realised because of a variety of quasi-random events exogeneous to my model. As Jedwab and Moradi explain, reasons why such railroads were never realised include a colonial officer who had been a proponent of a line between Accra and Kumasi dying before making a trip to London to secure final funding for the project, the Anglo-Ashanti wars ending before development of a military feeder railroad to Cape Coast could begin, or a support for a planned line between Kericho and Sotik in Kenya being scrapped because building it would have required viaduct technology and was hence deemed too expensive (see the online appendix of Jedwab et al., 2017). In total, there are 61 such placebo lines (34,000 kilometres long) and 237 realised railroads (47,000km). South Africa has the highest amount of total realised rail kilometres, as well as the densest network (measured in miles per square kilometre). The DRC and Cameroon have the longest and densest placebo network, respectively. Figure A.5 in the Appendix prints a map.

Figure 4 also prints (in bars) the relevant reduced-form gradient of contemporary road density and infrastructure discrimination with regards to distance to the placebo railroads. As can be seen, regions close to a placebo railroads also have more contemporary road infrastructure today, hinting at a *substitutability* between rail and road: these connections where evidently deemed important enough for road infrastructure to step in when rail infrastructure didn't materialise. Interestingly, the social planner sees no need to significantly reorganise infrastructure away from locations close to the placebo rail (the blue bars often cross zero) – potentially because these roads over time formed more organically and are still important to the functioning of the trade network today.

**Results.** Table 1 displays results from OLS estimation of equation (6) with $\tilde{\Lambda}_{\text{imm}}$ on the left hand side and total rail kilometres as explanatory variable $\text{Rail}_i$. Column (1) reports the baseline association between colonial railways and present-day infrastructure discrimination: every 50 kilometres of colonial railway construction are associated with grid cells losing 0.07 standard deviations of welfare at the hand of areas without any investment. Column (2) repeats the exercise but using "placebo" railroads. Reassuringly, the coefficient is slightly less than half as big (in absolute terms) and not significantly different from zero. Columns (3) and (4) distinguish between railroads built for different purposes: while both military and mining purpose railroads are associated with too much infrastructure today, it is in particular areas with military railroads that lose out under the reallocation scenario.

Columns (5) and (6) trace out the spatial gradient of infrastructure discrimination close to colonial railroad lines. These coefficients are analoguous to those printed in Figure 4, yet with a full set of controls and clustered errors. We again observe a substantial gradient, with grid cells within 10KM of a realised railroad being 0.18 standard deviations too well off. Magnitudes decline for cells up to 20KM away, after which the association reverses and even becomes positive beyond 30KM. No comparable associations can be found for placebo railroads in column (6). This is suggestive evidence that the confounding effect of colonial infrastructure policies is locally contained. Areas blessed with a close-by railway line are still too well off today, which comes at the expense of their neighbouring regions just a few kilometres away.

Do placebo railroads represent a plausible control group? Table A.2 in the Appendix reports results of balance tests between placebo and realised railroads across the entire sample. The table prints baseline means of various geographic and economic characteristics of cells (most of which are

Table 1: Colonial railroads and local infrastructure discrimination index

| | Infrastructure discrimination $\Lambda_{\text{imm}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.0677*** | | | | | |
| | (0.0249) | | | | | |
| 50 KM of Colonial Placebo Railroads | | -0.0401 | | | | |
| | | (0.0580) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.0939** | | | |
| | | | (0.0390) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.0448 | | |
| | | | | (0.0284) | | |
| <10KM to railroad | | | | | -0.0926*** | -0.0393 |
| | | | | | (0.0334) | (0.0241) |
| 10-20KM to railroad | | | | | -0.117*** | -0.0128 |
| | | | | | (0.0331) | (0.0267) |
| 20-30KM to railroad | | | | | -0.0278 | -0.0141 |
| | | | | | (0.0363) | (0.0380) |
| 30-40KM to railroad | | | | | 0.0468* | 0.0334 |
| | | | | | (0.0279) | (0.0280) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.764 | 0.764 | 0.764 | 0.764 | 0.765 | 0.764 |

*Results of estimation of equation* (6) *on the sample grid cells for the entire African continent. Dependent variable is the z-scored Local Infrastructure Discrimination Index $\tilde{\Lambda}_{imm}$. Columns (1)-(4) estimate the effect of colonial infrastructure investments as measured by the total number of colonial railroad kilometres of different types crossing a cell. Geography controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, indicators for the 12 predominant agricultural biomes, indicators for whether a cell is a capital, within 25 KM of a natural harbour, navigable river, or lake, the fourth-order polynomial of latitude and longitude, and an indicator of whether the grid cell lies on the border of a country's network. They also include population, night lights, and ruggedness. These indicators went into the original infrastructure reallocation simulation. Standard errors are clustered on the 3 $\times$ 3 degree level and are shown in parentheses.*

in the control vector $\mathbf{X}'_{i,c}$ from above) for cells crossed by a realised railroad (column 1) and a placebo railroad (column 2), plus the p-value of the null hypothesis of no plain difference-in-means between the two (column 3). Placebo cells are significantly different from cells with an actual railroad in a variety of ways, such as more agricultural suitability, precipitation, and more malaria prevalence. Similarly, they have less population today (which is one of the headline results of the paper by Jedwab and Moradi), and are less likely to cross the national capital. To account for the fact that many of these differences are spatially autocorrelated, column 4 presents p-values of a joint regression of these geographic indicators on presence of a colonial rail, which shows fewer significant differences. To assess in what direction these differences bias the main result empirical result that the social planner would like to reorganise infrastructure away from places with railroads, but not from places with placebo railroads, these differences need to be combined with how much each covariate impacts infrastructure discrimination $\Lambda$. For example, treated cells have a lower agricultural suitability (conceivably a proxy for productive places in colonial times) yet higher night luminosity (a proxy for productivity today) than placebo cells. If the social planner systematically reallocated roads towards places with such characteristics, this would result in bias. However, looking at column 1 of Table A.1, neither of these proxies significantly correlates with $\Lambda_{\text{imm}}$ today, making any such bias less of a concern. Column 5 combines the sign of any remaining significant difference with the sign and significance of the relevant connection between all covariates and $\Lambda_{\text{imm}}$. The two main potential avenues for bias in favour of the result are road density and whether a cell is at a country's border (places with realised railroads have more current roads and are less likely to be at the border, both of which are associated with lower levels of discrimination, leading to opposite signs of the potential boas). Since current road density can be seen to *itself* be outcomes of colonial railroad investments, one plausible interpretation of this is that railroads had an effect on infrastructure density, and through this are now affecting infrastructure disparities across space.[17,18]

The effects described in Table 1 are small, yet remarkable. Across the African continent, areas that received large infrastructure investments a century ago are still *too* well off given their position in the national trade network. In contrast, areas that were not crossed by tracks are inefficiently short on infrastructure today. To see that this is a non-trivial finding, note that, firstly, most of the colonial railway lines have been in disrepair for decades and thus do not immediately dictate trade flows today. Secondly, recall that the optimal network reallocation and construction of $\Lambda_i$ was based on roads and cars, not rails and trains. The implication is hence *not* that colonial railway systems themselves are inadequate to efficiently sustain inter-regional trade today. Rather the transport revolution a century ago coordinated the entire economy into a certain spatial equilibrium, which persists even though it has become inefficient. African nations would benefit from moving to a better equilibrium, but are locked in the current state. The social planner identifies this, seeks to overcome these misallocations, and moves infrastructure away from regions once considered important by the colonisers. The even stronger findings for military railroads reinforces that point: while mining railroads arguably still dictate trade flows today, military railroads have become completely obsolete and are much more clearly associated with imbalanced infrastructure stocks across space. In short, Jedwab and Moradi show that colonial investments helped the economy to coordinate on one of many spatial equilibria – my findings suggest that this is not the optimal one.

These results are largely unchanged in the case of mobile labor (though estimates are slightly larger, see upper panel of Table A.3 in the Appendix), and for optimal expansion instead of reallocation (with similar magnitudes, see Table A.4)

**Persistence: suggestive mechanisms.** Why would colonial investments a century ago still play a role in the spatial configuration of Africa's trade network today? The literature has documented such spatial persistence across contexts and time periods (Davis and Weinstein, 2002; Miguel and Roland, 2011; Bleakley and Lin, 2012; Dell, 2010; Lowes et al., 2017; Bertazzini, 2022). Infrastructure is a prime candidate for spatial persistence, as it is long-lasting and immobile. However, if infrastructure investments by colonial authorities skewed African trade networks into an inefficient configuration, how well could African governments have been able to counter these imbalances post independence?

Existing studies have shown that infrastructure and other government investments are often motivated in part by political economy concerns and regional favouritism (Hodler and Raschky, 2014; Burgess et al., 2015; De Luca et al., 2018; Dreher et al., 2019). To investigate whether such favouritism is associated with trade network imbalances today, I make use of a dataset of African national leaders provided by Dreher et al. (2019). The data entail information about the birth region and time in office of 117 heads of state holding power in 44 African countries dating back to 1969

[17] There are other covariates that might account for bias against the main estimate of Table 1. For example, the fact that treated cells have higher population today does not immediately bias the main specification, as population is not a significant determinate of infrastructure provision $\Lambda_{\text{imm}}$ (if anything, places with more population have slightly higher discrimination measures, which would go against the direction of the main estimate, see column 1 of Table A.1). Column 5 of Table A.2 reports in which direction significant differences between rail and placebo cells could plausibly produce concerns of bias, based on whether a covariate is a significant predictor of infrastructure discrimination.

[18] The finding that the social planner finds it optimal to reallocate benefit economically less productive regions at the expense of regions with high night luminosity could hint at previous over-agglomeration on the African continent. Investigating this hypothesis could be an exciting avenue for future work.

Table 2: Regional favoritism

| | Discrimination $\Lambda_{\text{imm}}$ | | Relative road expenditure | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A: entire sample* | | | | |
| Ever in power dummy | -0.0988* | | | |
| | (0.0525) | | | |
| log(1 + Total years in power) | | -0.0324 | | |
| | | (0.0214) | | |
| *Panel B: Kenyan road expenditure over time* | | | | |
| $\Lambda_{\text{imm}}$ (non-democracy) | | | -2.550** | |
| | | | (1.114) | |
| $\Lambda_{\text{imm}}$ (democracy) | | | | 0.503 |
| | | | | (0.756) |
| Year and District FE | | | Yes | Yes |
| (Demographic, political, economic, geographic) $\times$ trend | | | Yes | Yes |
| Country FE | Yes | Yes | | |
| Geography Controls | Yes | Yes | | |
| N | 10158 | 10158 | 451 | 410 |
| R2 | 0.764 | 0.764 | 0.362 | 0.215 |

*Columns 1-2 print coefficient from a regression similar to equation (6): $\tilde{\Lambda}_{i,c} = \beta Power_{i,c} + \boldsymbol{X}_{i,c}\gamma + \delta_c + \epsilon_{i,c}$ where $Power_{i,c}$ is either a dummy equaling one if anybody from grid cell i ever rose to power in country c (column 1), or the log total amount of years such people have spent in power. Geography controls as i Table 1. Columns 3-4 present estimates from a panel regression of equation (7) in the spirit of Burgess et al. (2015). Demographic, political, economic, and geographic controls include a district's 1962 population, urbanisation rate, wage earnings, wage-employment, cash crop production value, as well as its area, distance to Nairobi, and whether it is at Kenya's border or on the Mombasa-Nairobi corridor, all interacted with a time trend.*

and spatially merge them onto my grid dataset.

Column 1 of Table 2 present results of a specification similar to (6), yet including on the right hand side a dummy for whether somebody born in a given grid cell ever rose to power of their nation (in place of $\text{Rail}_{i,c}$). Indeed, birthplaces of African leaders would lose about 0.1 standard deviations worth of welfare compared to other cells if the social planner had her way. Column 2 investigates the intensive margin by regressing $\tilde{\Lambda}_{\text{imm}}$ on the log of total years that a son or daughter of a given grid cell was in power[19] and finds suggestive (albeit insignificant) evidence that more years in power are associated with even more favourable networks.

To get a better sense of how political economy concerns might impede government's efforts to rein in on inefficient trade networks in real time, I revisit an analysis by Burgess et al. (2015). The authors focus on Kenya and find that government spending on road developments were more likely to go to regions that are coethnic with the country's leader at the time, but that this association goes away in times of democracy. A natural question of interest is whether this spending on favoured regions also represents suboptimal investments from the social planner's view.

To speak to this, I recompute my infrastructure discrimination measure at different points in time across Kenya's history. I use historical data on Kenya's road network from 21 different years between 1963 and 2007 coming from digitised Michelin maps provided by Jedwab and Storeygard (2022). I aggregate these infrastructure stocks onto the level of Kenyan districts, provided by Burgess et al. (2015). At each point in time, I then use my model to compute the optimally reallocated and expanded counterfactual road network, as well as discrimination measures $\Lambda$. Figure A.7 print maps of optimal networks for two different years. More details on this procedure are described in

[19]To account for the many zeroes in the data, I add 1 to each entry, noting that this is not innocuous (see Chen and Roth, 2023).

appendix C.3. I then run an analogue of the main regression in Burgess et al. (2015)

$$\text{Road expenditure share}_{dt} = \gamma_d + \alpha_t + \beta(\text{coethnic district}_{dt}) + \theta(\mathbf{X}_d \times [t - 1962]) + \Lambda_{dt} + \epsilon_{dt} \quad (7)$$

across districts $d$ and years $t$. Following Burgess et al., Road expenditure share$_{dt}$ is the ratio of a districts road expenditure share, divided by its population share. $(\mathbf{X}_d \times [t - 1962])$ is a vector of controls interacted with a time trend, and (coethnic district$_{dt}$) is a dummy of whether the country's president in year $t$ belongs to the same ethnic group as the majority of district $d$. $\Lambda_{dt}$ is my model-generated discrimination measure.

Table 2, columns 3-4 present results of estimating equation (7) separately for periods in which Kenya had no multiparty democracy (1970-1992) and in which it did (1963-1969, and again 1993-2007). Column 3 shows that in times of non-democracy, road expenditure significantly skews towards regions the social planner identifies as already being too well off, relative to the social optimum. The negative coefficient implies that a one standard deviation lower welfare improvement from optimal reallocation is associated with about 2.5 *more* relative spending per capita. The trend somewhat reverses in times of democracy, even though the point estimate is noisy: column 4 yields suggestive, yet not statistically significant evidence that during multiparty rule, investments tend to go towards districts that the social planner identifies as being in need of more infrastructure.[20,21]

**Discussion.** African countries were left with an economic network that wasn't designed to sustain them as independent countries. Through badly drawn national borders and railroads built for extractive economies, the colonial authorities coordinated countries onto a certain spatial equilibrium, characterised fundamentally by substantial geographic inequalities (Alesina et al., 2011, 2016; Jedwab and Moradi, 2016; Michalopoulos and Papaioannou, 2014, 2016, 2020). Subsequent investments into road infrastructure were constrained by this violent history. They were also often less effective because of regional and ethnic favouritism. With this legacy in mind, my analysis asks what could have been achieved if, over the long run, road networks in each African country had been designed optimally, without being constrained by colonial boundaries or outdated infrastructure projects.

My analysis shows that the shadow of these transport network inefficiencies is still looming large: African countries would gain between 1-6% of welfare if they could reorganise their networks from scratch, representing substantially higher gains than in other world regions. Yet at the same time, my findings also imply that similar welfare gains can be achieved by a large one-time infrastructure investment program. While this would undoubtedly represent an ambitious endeavor, it might not be out of reach.

[20] Results are similar, albeit even noisier and insignificant, in the case of immobile labor, or optimal expansion (see Table A.9 in the Appendix). I also run a version of equation (7) with added interaction effects (see Table A.8 in the Appendix). I find suggestive evidence that non-democracies target districts that are even less in need of added infrastructure, if that district shares ethnicity with the president at the time (column 3). This effect also doesn't fully disappear in democratic times: while road expenditure does go towards places the planner also favours during those times, this effect is strongest for coethnic districts. This could hint at a "residual favouritism" effect, ie. a democratic leadership makes welfare-improving investments, yet in particular if those also serve their core constituency.

[21] Next to local governments, the other large source of infrastructure spending in Africa is foreign aid. In Appendix section D, I investigate whether aid by the World Bank and China help in overcoming the infrastructure imbalances I identify. Using data from AidData (2017) and Strange et al. (2017) on the precise destination of thousands of lending lines from the World Bank and China to Africa, I find that such foreign-sponsored projects by either source are also not more likely to go to cells my model predicts as being particularly in need. Indeed, total $-value of World Bank or China disbursements, as well as the number of transport-sector projects, are significantly correlated with locations deemed too well off, relative to the global optimum. Foreign aid hence presents a second powerful candidate to explain persistence of infrastructure imbalance across Africa.

# 6 Conclusion

In this paper, I identify spatial inefficiencies in Africa's trade network. I first construct a comprehensive economic topography of the entire continent, bringing together data from a variety of sources. I then present a simple network trade model and simulate the flow of goods through the internal geography formed by 10,000 African regions and almost 75,000 network connections. Harnessing the theoretical contribution by Fajgelbaum and Schaal (2020), I endogenise the transport network in order to derive the unique optimally reorganised as well as expanded road network for every country in Africa.

I then compare each country's current network to its hypothetically optimal one. I rank countries by overall network efficiency and presented a fine-resolution spatial dataset quantifying which sub-national areas are disadvantaged by the status quo. I empirically investigate patterns of trade network imbalances over space and link inefficiencies to persistent lock-in effects caused by colonial infrastructure investments and differential treatment on the basis of regional favouritism.

The main theoretical shortcoming of my model is its lack of dynamics, especially when combined with agglomeration forces. Indeed, infrastructure policy in the real world is often an attempt to steer an economy onto a persistent new growth path by making people meet that otherwise wouldn't have met (Michaels et al., 2021). This could threaten the results by making dynamically efficient infrastructure choices look statically inefficient. Recent theoretical advances such as Allen and Donaldson (2022) open the door for potentially exciting research on quantifying not just spatial, but also temporal inefficiencies in trade networks across the world.

In contributing a comprehensive spatial measure on the differential provision of a primary public good covering an entire continent, my study provides the quantitative foundation for further research questions pertaining to inequality over space. Future research designs could employ my dataset to analyse regional roots of conflict, political activism, social mobility, or subjective overall wellbeing. Another interesting avenue for inquiry could be to investigate whether infrastructure inefficiency spatially covaries with the provision of other public goods like education or health. My findings can also be benchmarked against other programs aimed at overcoming spatial inefficiencies, such as the relaxation of labor and capital frictions. Lastly, extending my analysis to larger geographical units could shed light on returns to continent-wide international infrastructure projects.

Identifying spatial inefficiencies and understanding their historical and political roots can be the first step in outlining effective place-based policies. Equipped with an unparalleled availability of spatial data and computing power, policymakers in Africa and around the world should feel empowered to combat local imbalances and design powerful interventions to better connect millions.

# References

**AidData (2017).** World Bank Geocoded Research Release, Version 1.4.2. http://aiddata.org/data/world-bank-geocoded-research-release-level-1-v1-4-2 (Accessed: 2018-04-06). Williamsburg, VA and Washington, DC

**Alder, Simon (2022).** Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. Working Paper

**Alesina, Alberto, William Easterly, and Janina Matuszeski (2011).** Artificial States. *Journal of the European Economic Association* 9(2), pp. 246–277

**Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou (2016).** Ethnic inequality. *Journal of Political Economy* 124(2), pp. 428–488

**Allen, Treb, Cauê Dobbin, and Melanie Morten (2019).** Border Walls. Technical report

**Allen, Treb and Dave Donaldson (2022).** Persistence and Path Dependence in the Spatial Economy. Working Paper

**Asher, Sam and Paul Novosad (2020).** Rural Roads and Local Economic Development. *American Economic Review*

**Atkin, David and Dave Donaldson (2015).** Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs. Working Paper, National Bureau of Economic Research

——— **(2022).** The Role of Trade in Economic Development. *Handbook of International Economics* 5

**Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang (2017).** Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics* 99(3), pp. 435–448

**Bernot, Marc, Vicent Caselles, and Jean-Michel Morel (2009).** *Optimal Transportation Networks*, volume 1955 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg

**Bertazzini, Mattia C (2022).** The long-term impact of Italian colonial roads in the Horn of Africa, 1935–2015. *Journal of Economic Geography* 22(1), pp. 181–214

**Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen (2011).** Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165(2), pp. 137–151

**Bilal, Adrien and Esteban Rossi-Hansberg (2021).** Location as an Asset. *Econometrica* 89(5), pp. 2459–2495

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

**Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou (2024).** Investment in Infrastructure and Trade: The Case of Ports

**Buera, Francisco, Yongseok Shin, Hugo Hopenhayn, and Nicholas Trachter (2023).** Big Push in Distorted Economies. Working Paper

**Burgess, Robin and Dave Donaldson (2012).** Railroads and the Demise of Famine in Colonial India. Working Paper, LSE and NBER

**Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel (2015).** The Value of Democracy: Evidence from Road Building in Kenya. *The American Economic Review* 105(6), pp. 1817–1851

**Center, Socioeconomic Data and Applications (2016).** Gridded Population of the World (GPW), v4 | SEDAC. http://sedac.ciesin.columbia.edu/data/collection/gpw-v4 (Accessed: 2017-11-26)

**Chen, Jiafeng and Jonathan Roth (2023).** Logs with Zeros? Some Problems and Solutions. *The Quarterly Journal of Economics* p. qjad054

**Collier, Paul, Martina Kirchberger, and Mans Söderbom** (**2015**). The cost of road infrastructure in low-and middle-income countries. *The World Bank Economic Review* 30(3), pp. 522–548

**Coşar, A. Kerem and Pablo D. Fajgelbaum** (**2016**). Internal Geography, International Trade, and Regional Specialization. *American Economic Journal: Microeconomics* 8(1), pp. 24–56

**Davis, Donald R. and David E. Weinstein** (**2002**). Bones, bombs, and break points: the geography of economic activity. *The American Economic Review* 92(5), pp. 1269–1289

**De Luca, Giacomo, Roland Hodler, Paul A. Raschky, and Michele Valsecchi** (**2018**). Ethnic favoritism: An axiom of politics? *Journal of Development Economics* 132, pp. 115–129

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

**Donaldson, Dave** (**2018**). Railroads of the Raj: Estimating the impact of transportation infrastructure. *The American Economic Review* 108(4), pp. 899–934

**Donaldson, Dave and Richard Hornbeck** (**2016**). Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics* 131(2), pp. 799–858

**Doxsey-Whitfield, Erin, Kytt MacManus, Susana B. Adamo, Linda Pistolesi, John Squires, Olena Borkovska, and Sandra R. Baptista** (**2015**). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography* 1(3), pp. 226–234

**Dreher, Axel, Andreas Fuchs, Roland Hodler, Bradley Parks, Paul Raschky, and Michael Tierney** (**2019**). Aid on demand: African leaders and the geography of China's foreign assistance. *Journal of Development Economics* 140, pp. 44–71

**Duranton, Gilles and Anthony J Venables** (**2017**). Place-Based Policies for Development. Working Paper, The World Bank

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

**Fajgelbaum, Pablo D and Cecile Gaubert** (**2020**). Optimal Spatial Policies, Geography, and Sorting*. *The Quarterly Journal of Economics* 135(2), pp. 959–1036

**Fajgelbaum, Pablo D. and Edouard Schaal** (**2017**). Optimal transport networks in spatial equilibrium. Working Paper, National Bureau of Economic Research

——— (**2020**). Optimal Transport Networks in Spatial Equilibrium. *Econometrica* 88(4), pp. 1411–1452

**Fiorini, Matteo, Marco Sanfilippo, and Asha Sundaram** (**2021**). Trade liberalization, roads and firm productivity. *Journal of Development Economics* 153, p. 102712

**Foster, Vivien and Cecilia Briceño-Garmendia** (**2010**). *Africa's Infrastructure: A Time for Transformation*. Africa Development Forum. World Bank Publications, Washington DC

**Galichon, Alfred** (**2016**). *Optimal Transport Methods in Economics*. Princeton University Press

**Gaubert, Cecile, Patrick Kline, Damian Vergara, and Danny Yagan** (**2020**). Place-Based Redistribution. Working Paper

**Glaeser, Edward and Joshua Gottlieb** (**2008**). The Economics of Place-Making Policies. Brookings Papers on Economic Activity 2

**Gorton, Nicole and Elena Ianchovichina** (**2022**). Trade Networks in Latin America: Spatial Inefficiencies and Optimal Expansions. Technical report, The World Bank

**Head, Keith and Thierry Mayer** (**2014**). Gravity Equations: Workhorse,Toolkit, and Cookbook. In: *Handbook of International Economics*, volume 4, pp. 131–195. Elsevier

**Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David N Weil** (**2018**). The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade. *The Quarterly Journal of Economics* 133(1), pp. 357–406

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

**Herranz-Loncán, Alfonso and Johan Fourie** (**2017**). "For the public benefit"? Railways in the British Cape Colony. *European Review of Economic History* pp. 1–28

**Hodler, Roland and Paul A. Raschky** (**2014**). Regional Favoritism. *The Quarterly Journal of Economics* 129(2), pp. 995–1033

**Jedwab, Remi, Edward Kerby, and Alexander Moradi** (**2017**). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya. *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. *The Review of Economics and Statistics* 98(2), pp. 268–284

**Jedwab, Remi and Adam Storeygard** (**2022**). The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010. *Journal of the European Economic Association* 20(1), pp. 1–38

**Kline, Patrick and Enrico Moretti** (**2013**). Place Based Policies with Unemployment. *American Economic Review* 103(3), pp. 238–243

——— (**2014**). Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics* 129(1), pp. 275–331

**Kreindler, Gabriel E, Arya Gaduh, Tilman Graff, Rema Hanna, and Benjamin Olken** (**2023**). Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta. NBER Working Paper 31369

**Krugman, Paul** (**1991**). History Versus Expectations. *The Quarterly Journal of Economics* 106(2), pp. 651–667

**Limao, Nuno and Anthony J. Venables** (**2001**). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review* 15(3), pp. 451–479

**Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel** (**2017**). The Evolution of Culture and Institutions: Evidence From the Kuba Kingdom. *Econometrica* 85(4), pp. 1065–1091

**Michaels, Guy, Dzhamilya Nigmatulina, Ferdinand Rauch, Tanner Regan, Neeraj Baruah, and Amanda Dahlstrand** (**2021**). Planning Ahead for Better Neighborhoods: Long-Run Evidence from Tanzania. *Journal of Political Economy* 129(7), pp. 2112–2156

**Michalopoulos, Stelios and Elias Papaioannou** (**2013**). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica* 81(1), pp. 113–152

——— (**2014**). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics* 129(1), pp. 151–213

——— (**2016**). The Long-Run Effects of the Scramble for Africa. *The American Economic Review* 106(7), pp. 1802–1848

——— (**2020**). Historical Legacies and African Development. *Journal of Economic Literature* 58(1), pp. 53–128

**Miguel, Edward and Gérard Roland** (**2011**). The long-run impact of bombing Vietnam. *Journal of Development Economics* 96(1), pp. 1–15

**Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny** (**1989**). Industrialization and the Big Push. *Journal of Political Economy* 97(5), pp. 1003–1026. Place: Chicago Publisher: The University of Chicago Press

**Nugent, Paul and Hugh Lamarque** (**2022**). *Transport Corridors in Africa*. James Currey, Rochester, NY. Publisher: Boydell and Brewer

**Porcher, Charly** (**2021**). Migration with Costly Information. Working Paper

**Porteous, Obie** (**2022**). Reverse Dutch disease with trade costs: Prospects for agriculture in Africa's oil-rich economies. *Journal of International Economics* 138, p. 103651

**Redding, Stephen J.** (**2016**). Goods trade, factor mobility and welfare. *Journal of International Economics* 101, pp. 148–167

**Roback, Jennifer** (**1982**). Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90(6), pp. 1257–1278

**Santamaria, Marta** (**2022**). Reshaping Infrastructure: Evidence from the division of Germany. Working Paper

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

**Strandow, Daniel, Michael Findley, Daniel Nielson, and Josh Powell** (**2011**). The UCDP and AidData codebook on georeferencing aid Version 1.1. Technical report, Department of Peace and Conflict Research, Uppsala University, Uppsala

**Strange, Austin M., Axel Dreher, Andreas Fuchs, Bradley Parks, and Michael J. Tierney** (**2017**). Tracking Underreported Financial Flows: China's Development Finance and the Aid–Conflict Nexus Revisited. *Journal of Conflict Resolution* 61(5), pp. 935–963

**Swisher, I. V.** (**2017**). Reassessing railroads and growth: Accounting for transport network endogeneity. Working Paper, Faculty of Economics, University of Cambridge

**Teravaninthorn, Supee and Gaël Raballand** (**2009**). *Transport Prices and Costs in Africa: A Review of the Main International Corridors*. Directions in Development; Infrastructure. World Bank Publications, Washington DC

**The Economist** (**2015**). All aboard. *The Economist* February 26, 2015

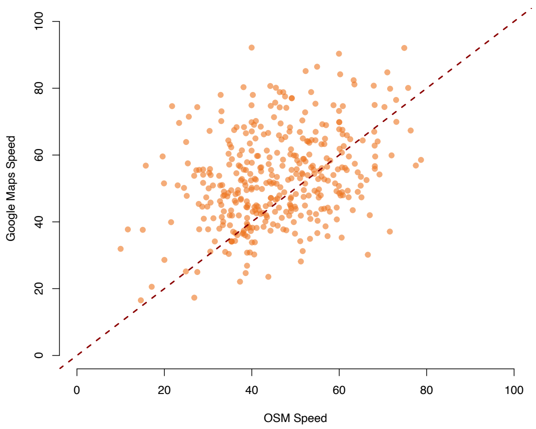
**The World Bank** (**2017**). The World Bank Annual Report 2017. Technical report, The World Bank, Washington DC

**Wang, Haizhong, Jia Li, Qian-Yong Chen, and Daiheng Ni** (**2011**). Logistic modeling of the equilibrium speed–density relationship. *Transportation Research Part A: Policy and Practice* 45(6), pp. 554–566

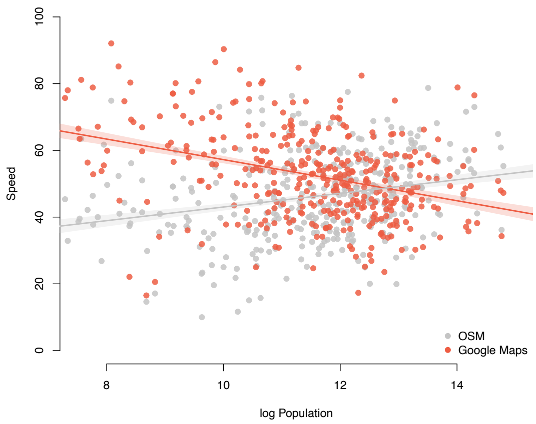
# Appendix

# A Additional figures and tables

Figure A.1: Cross validation of OSM roads data



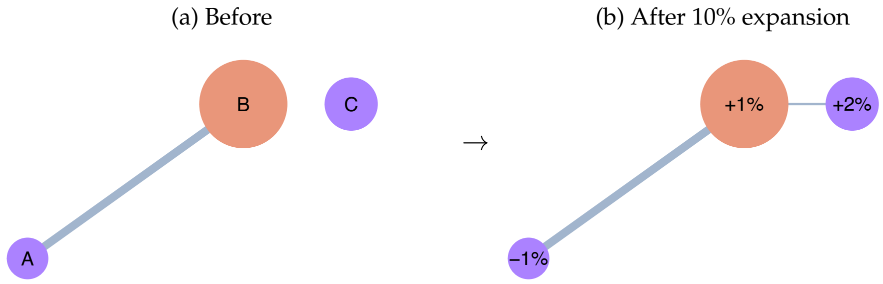

(a) Direct speed comparison




(b) Speed comparison by log population

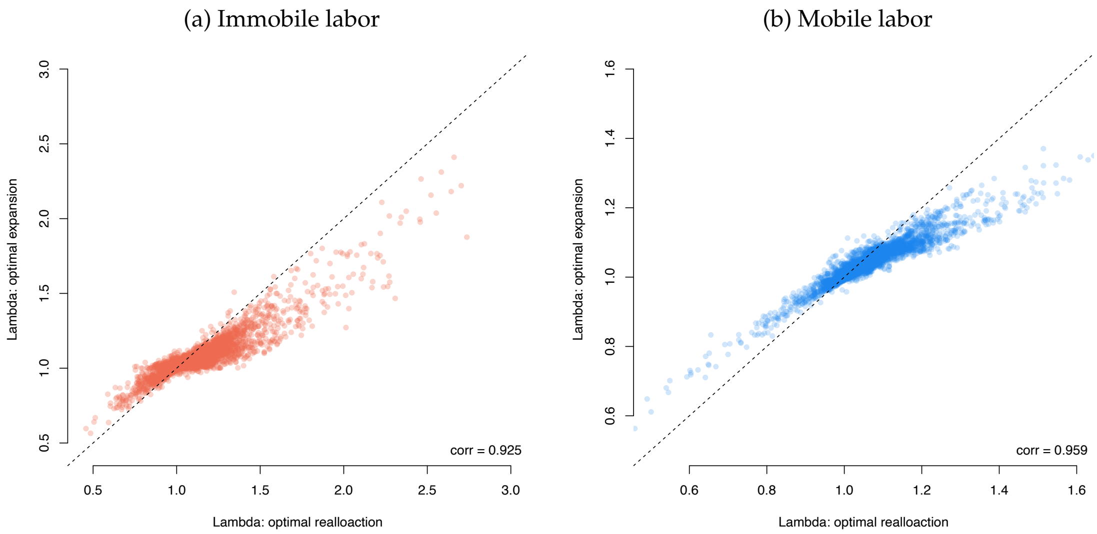
*Cross validation of Open Street Maps (OSM) speed data with Google Maps (GM). I scrape routing information from GM for a random 1% subset of connections. Panel (a) plots the distributions of the resulting speed measures from both providers against each other. Panel (b) plots both speed distributions against the log of the average population between origin and destination grid cells. Regression lines with 95% confidence intervals are overlaid. Note, these speeds are faster than the average speeds reported in the main text of the paper, as they do not include the significant amount of time spent walking in many parts of the network.*

Figure A.2: Example where optimal expansion can still lead to local losses



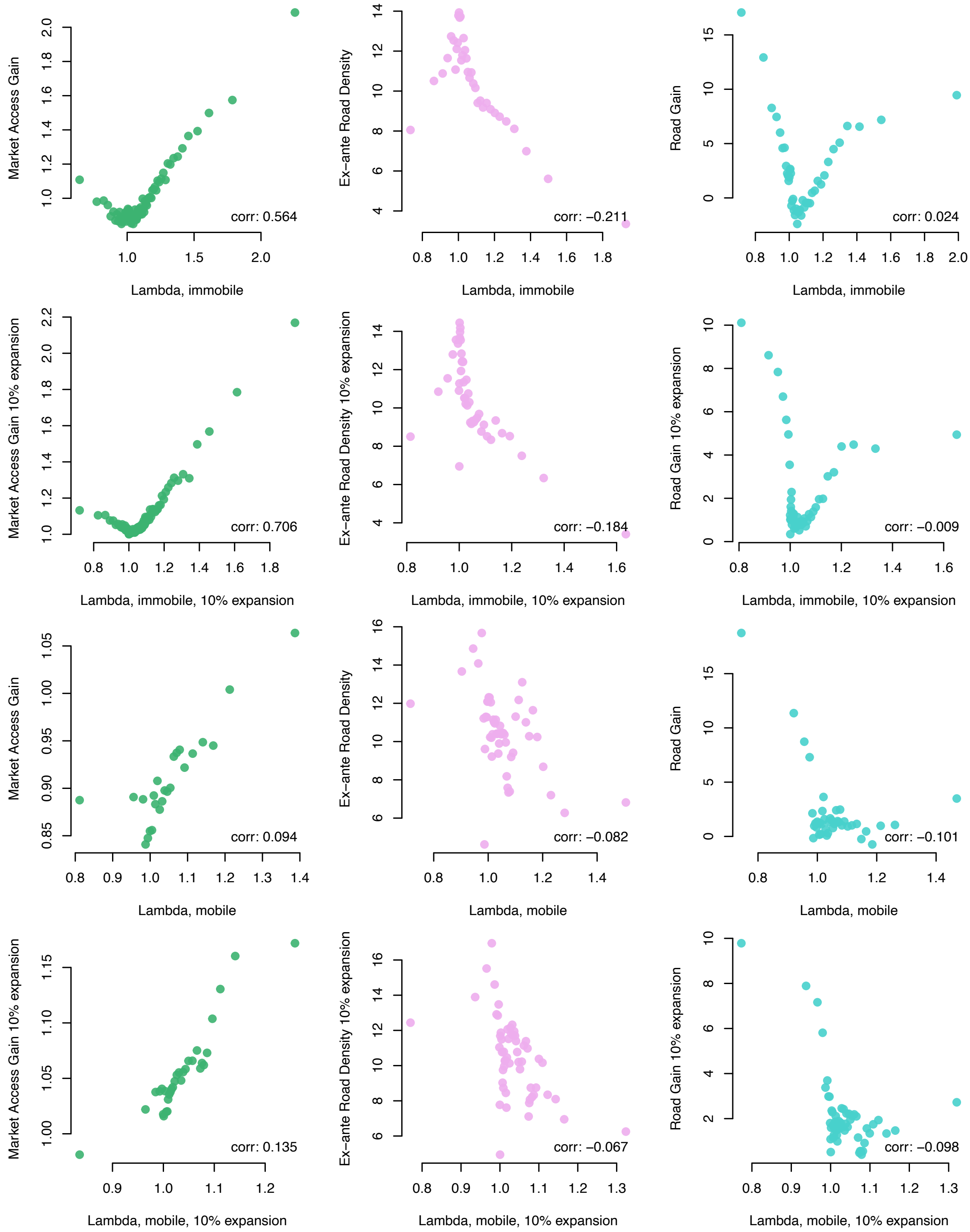

*Example of how optimal expansion can lead to losses in some regions, even when they do not lose any infrastructure. In this example, locations A and C produce the same variety and are small. Location B produces a second variety and is much closer to C but ex-ante only connected to A. Optimal expansion connect B and C, which leads to welfare increases on aggregate, but hurts location A, who can now sell less to B and hence get less of B's variety.*

Figure A.3: Reallocation and expansion scenarios



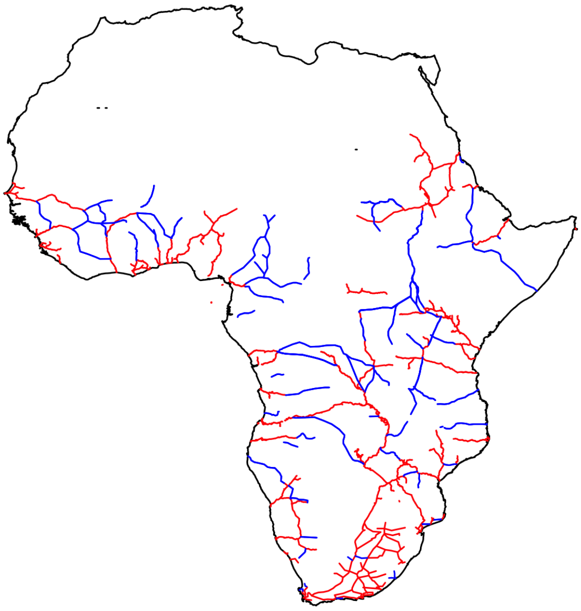

*Raw correlations of $\Lambda_{imm}$ and $\Lambda_{imm}^{10\%}$ and $\Lambda_{mob}$ and $\Lambda_{mob}^{10\%}$, respectively.*

Figure A.4: Correlations of $\Lambda$ with other road network measures



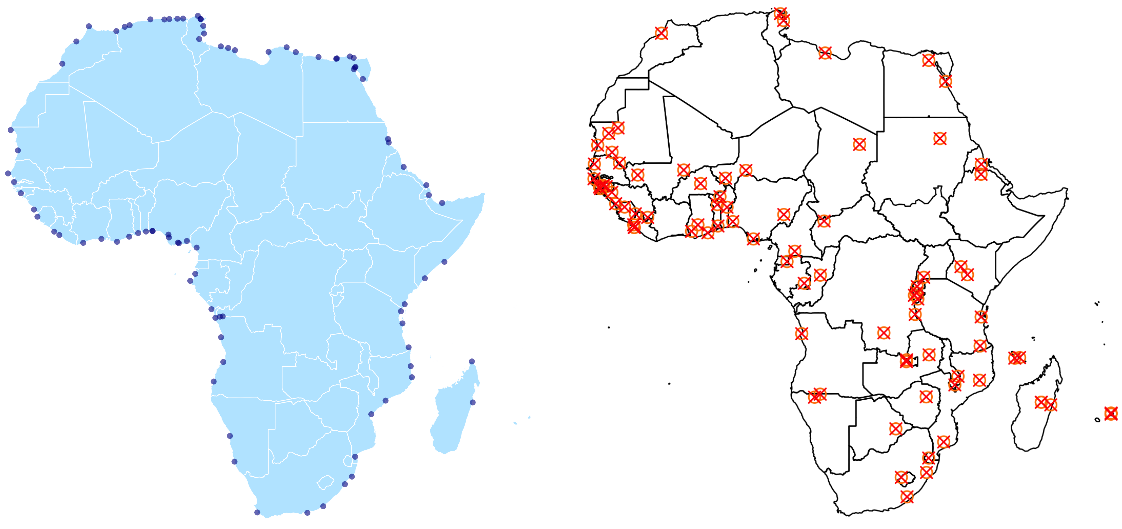

*Raw correlations of various $\Lambda$ measures of infrastructure discrimination with plausible other road network measures. The first row reports scatter plots of $\Lambda_{imm}$ without labor mobility against a measure of market access change, ex-ante road density, and road gain. Market access is defined as $MA_i = \sum_j (1+\tau_{ij})^{-\sigma} z_j L_j$ where $\tau_{ij}$ corresponds to the cost of shipping a quantity of $Q = 1$ over the network. I compute the change in MA from the static network before re-allocation to the optimal one post re-allocation. Ex-ante road density is defined as $I_{i,initial} / L_i$ where $I_{i,initial}$ is the average infrastructure on all links originating at i before re-allocation. Lastly, infrastructure gain is defined as $I_{i,optimal} - I_{i,initial}$, the ratio of average infrastructure orginating in a link after vs before the reallocation exercise. The remaining rows repeat this exercise for $\Lambda^{10\%}_{imm}$, $\Lambda_{mob}$, and $\Lambda^{10\%}_{mob}$*

Figure A.5: Colonial railway network



*Maps displaying the network of railway lines (red) and placebo railroads (blue). Data from Jedwab and Moradi (2016) and Herranz-Loncán and Fourie (2017). Railroads built by the colonial powers between 1890 and 1960 are printed in red. Lines that were initially planned but never actually built are printed in blue.*

Figure A.6: Further spatial data used



(a) Locations of ports

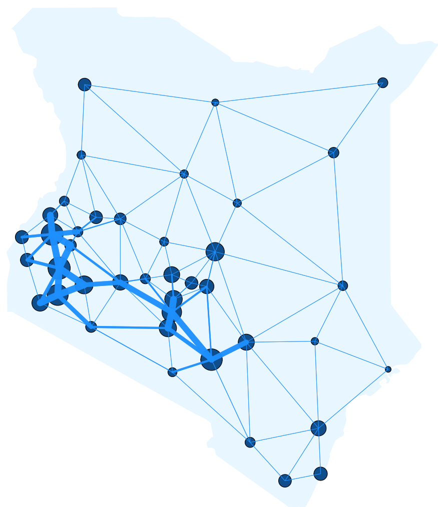
(b) Leaders' birthplaces

*Spatial distribution of ports and leaders' birthplaces across the continent. Ports data is hand-coded from Lloyd's list at https://directories.lloydslist.com/port and corresponds to the 90 biggest ports in Africa. Birthplace data from Dreher et al. (2019).*

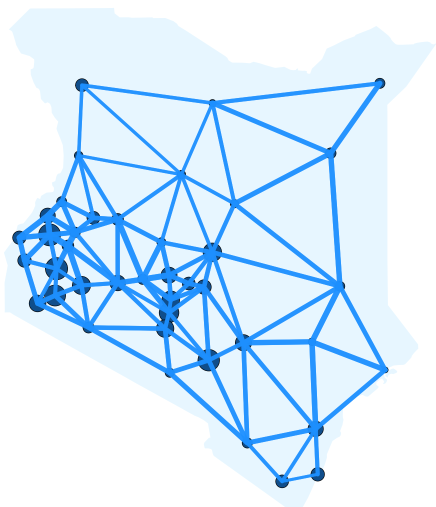
Figure A.7: Reallocation of historical road networks in Kenya

(a) Kenya 1963: before reallocation

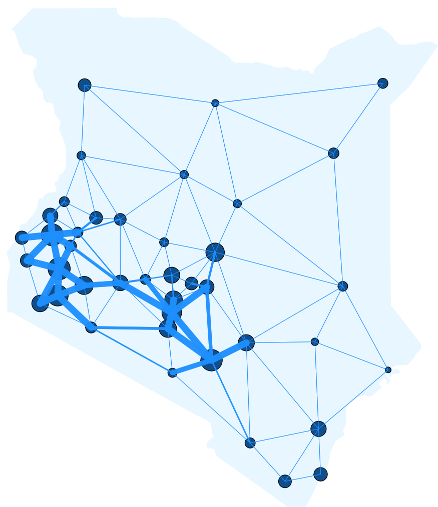
(b) Kenya 1963: after reallocation



(c) Kenya 2007: before reallocation

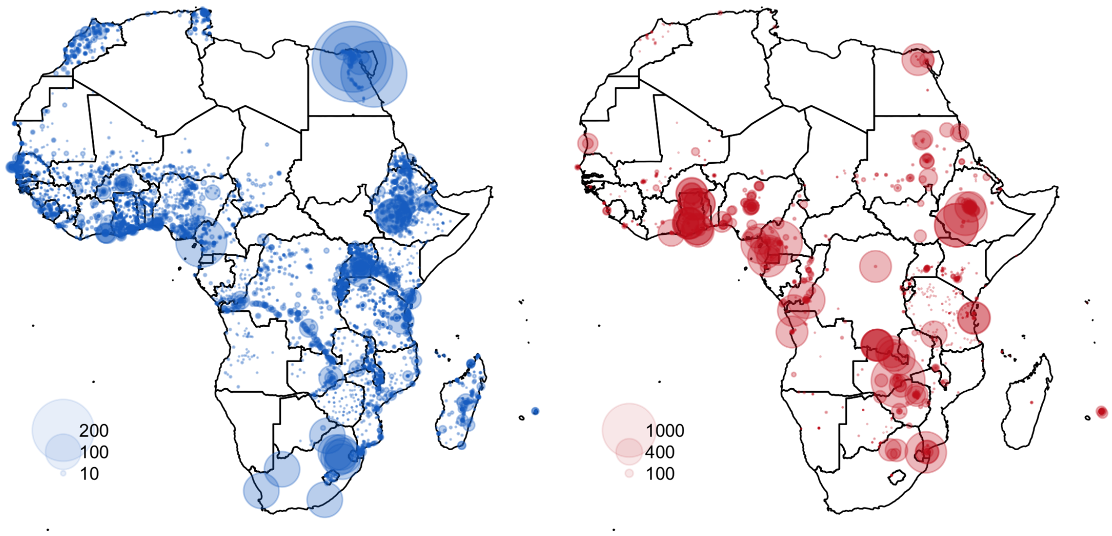


(d) Kenya 2007: after reallocation



*Ex-ante and optimally reallocated road networks for Kenya in 1963 and 2007. Ex-ante roads data from Jedwab and Storeygard (2022). Each dot represents a district with size proportional to its 1962 population (from Burgess et al., 2015). Maps print the reallocation with immobile labor.*

Figure A.8: Spatial distribution of development aid projects to African nations




(a) World Bank aid

(b) Chinese aid

*Foreign aid projects funded by the World Bank* (A.8a) *and China* (A.8b). *Each dot represents one project site with radius proportional to the logarithm of total disbursements flowing to each site. World Bank data comprise all projects approved between 1996–2014. Chinese data include tracked projects between 2000–2011. Map only depicts projects coded with sufficient precision to not be excluded (see Appendix text). If a project has multiple sites, total disbursements are assumed evenly distributed between locations. Data from AidData (2017) and Strange et al. (2017). Legend denotes disbursement values in million 2011 US dollars. Note that the legends have different scales.*

Table A.1: Geographic correlates of infrastructure discrimination measures

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | $\Lambda_{\text{imm}}$ | $\Lambda^{10\%}_{\text{imm}}$ | $\Lambda_{\text{mob}}$ | $\Lambda^{10\%}_{\text{mob}}$ | Housing | Housing (pc) |
| Population (in 100,000) | 0.00227 | 0.00226 | -0.0104** | -0.0144** | 0.00468 | 0.00996*** |
| | (0.00263) | (0.00263) | (0.00458) | (0.00619) | (0.00289) | (0.00349) |
| Ruggedness | -0.00000267 | -0.00000370 | 0.00000470 | 0.00000499 | -0.00000355 | -0.00000551 |
| | (0.00000333) | (0.00000326) | (0.00000308) | (0.00000314) | (0.00000418) | (0.00000340) |
| Night lights | -0.0148*** | -0.0138*** | -0.0230*** | -0.0222*** | -0.00766** | -0.0122** |
| | (0.00430) | (0.00424) | (0.00496) | (0.00535) | (0.00347) | (0.00592) |
| Altitude | -0.0000726 | -0.0000276 | 0.000110* | 0.000112* | -0.0000418 | -0.000238*** |
| | (0.0000532) | (0.0000537) | (0.0000585) | (0.0000626) | (0.0000473) | (0.0000775) |
| Agr. suitability index | 0.0658 | 0.0808 | -0.0225 | -0.0404 | -0.0741 | -0.0844* |
| | (0.0534) | (0.0511) | (0.0477) | (0.0585) | (0.0514) | (0.0484) |
| Temperature | -0.0141 | -0.00688 | -0.00672 | 0.00204 | 0.0109 | -0.0427** |
| | (0.0119) | (0.0120) | (0.0117) | (0.0122) | (0.0127) | (0.0170) |
| Precipitation | 0.0000230 | -0.00000789 | 0.00109* | 0.00162** | 0.00825 | 0.00404 |
| | (0.000437) | (0.000399) | (0.000594) | (0.000680) | (0.00707) | (0.00373) |
| Yearly growing days | 0.0000557 | 0.000187 | -0.000367 | -0.000390 | -0.00117 | -0.000551 |
| | (0.000293) | (0.000270) | (0.000337) | (0.000391) | (0.00115) | (0.000607) |
| Malaria prevalence | 0.00264 | 0.00405 | -0.00162 | -0.000610 | -0.000882 | 0.00108 |
| | (0.00257) | (0.00247) | (0.00198) | (0.00212) | (0.00229) | (0.00262) |
| $<$ 25 KM from suitable harbor | -0.130 | -0.0509 | -0.0524 | -0.216 | 1.685 | 2.477 |
| | (0.0840) | (0.0757) | (0.147) | (0.179) | (1.578) | (2.420) |
| $<$ 25 KM from navigable river | -0.319** | -0.333** | -0.107 | -0.0706 | 0.0991 | 0.138* |
| | (0.158) | (0.155) | (0.125) | (0.125) | (0.0831) | (0.0733) |
| $<$ 25 KM from navigable lake | 0.0397 | 0.0370 | -0.138** | -0.157* | 0.0127 | -0.00541 |
| | (0.0587) | (0.0643) | (0.0601) | (0.0830) | (0.0519) | (0.0337) |
| National capital | -0.00962 | -0.0259 | 0.0544 | -0.0546 | -0.137 | -0.231* |
| | (0.0825) | (0.0850) | (0.0837) | (0.111) | (0.0934) | (0.118) |
| At national border | -0.152*** | -0.152*** | -0.0134 | 0.00541 | 0.0386 | 0.0530** |
| | (0.0189) | (0.0187) | (0.0176) | (0.0179) | (0.0335) | (0.0259) |
| Road density (z-scored) | -0.187*** | -0.197*** | -0.155*** | -0.186*** | -0.0106 | -0.115** |
| | (0.0256) | (0.0254) | (0.0227) | (0.0256) | (0.0514) | (0.0461) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Remaining Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 8580 | 8384 |
| R2 | 0.769 | 0.782 | 0.887 | 0.863 | 0.0533 | 0.466 |

*Geographic correlates of the various model-implied measures, obtained from a joint regression of $\Lambda$ on the vector $\mathbf{X}$. Fourth-order polynomials of latitue and longitude, as well as country fixed effects are not printed.*

Table A.2: Geographic correlates of realised and placebo railway lines

| | (1) Rail | (2) Placebo | (3) p-value | (4) p-value (joint) | (5) Bias ($\Lambda_{\text{imm}}$) |
|---|---|---|---|---|---|
| Population (in 100,000) | 3.27 | 1.96 | **0.00** | **0.00** | *(unclear)* |
| | (7.03) | (4.24) | | | |
| Ruggedness | 2463.68 | 1801.06 | **0.00** | 0.87 | *(unclear)* |
| | (3193.47) | (2458.56) | | | |
| Night lights | 1.36 | 0.24 | **0.00** | 0.33 | *(unclear)* |
| | (5.18) | (0.98) | | | |
| Altitude | 793.71 | 786.20 | 0.43 | 0.46 | *(unclear)* |
| | (521.08) | (492.70) | | | |
| Agr. suitability index | 0.37 | 0.42 | **0.00** | 0.97 | *(unclear)* |
| | (0.26) | (0.26) | | | |
| Temperature | 22.74 | 23.93 | **0.00** | 0.60 | *(unclear)* |
| | (4.14) | (3.24) | | | |
| Precipitation | 73.40 | 91.06 | **0.00** | 0.18 | *(unclear)* |
| | (45.39) | (37.65) | | | |
| Yearly growing days | 155.05 | 199.77 | **0.00** | **0.01** | *(unclear)* |
| | (85.70) | (86.08) | | | |
| Malaria prevalence | 9.87 | 12.79 | **0.00** | 0.33 | *(unclear)* |
| | (9.84) | (9.65) | | | |
| $<$ 25 KM from suitable harbor | 0.01 | 0.00 | **0.05** | 0.43 | *(unclear)* |
| | (0.08) | (0.04) | | | |
| $<$ 25 KM from navigable river | 0.01 | 0.01 | 0.17 | 0.33 | *(unclear)* |
| | (0.12) | (0.08) | | | |
| $<$ 25 KM from navigable lake | 0.01 | 0.02 | 0.15 | 0.83 | *(unclear)* |
| | (0.10) | (0.15) | | | |
| National capital | 0.03 | 0.02 | **0.02** | 0.46 | *(unclear)* |
| | (0.16) | (0.13) | | | |
| At national border | 0.31 | 0.38 | **0.00** | **0.00** | **Against** |
| | (0.46) | (0.48) | | | |
| Road density (z-scored) | 0.35 | 0.24 | **0.00** | **0.00** | **In favour** |
| | (0.74) | (0.89) | | | |

*Raw means of different geographic variables for grid cells touched by a colonial railroad (1) or a placebo railroad (2). Column (3) prints p-values of a t-test of no means difference between (1) and (2). Column (4) prints p-values of a regression of all geographic variables at once, plus the other variables used in the main regression* (6)*, ie country fixed-effects, and higher-order lat-lon polynomials:* $IsRail_{i,c} = \beta_0 + \mathbf{X}_{i,c}\gamma + \delta_c + \epsilon_i$ *Column (5) prints the direction any significant difference from column (4) might bias the main result. To do so, it multiplies direction of the difference from (4) with the direction of the correlation of the covariate with* $\Lambda_{imm}$*, taken from column (1) of Table A.1. For example, rail cells have significantly higher road density than placebo cells, and high road density is associatied with lower* $\Lambda_{imm}$ *values (see Table A.1), so this might bias in favour of the main hypothesis that rail cells have lower* $\Lambda_{imm}$ *values than placebo cells. Bias is coded as* (unclear) *if either the p-value in column (4) is lower than 0.05, or the p-value against the* $H_0$ *of zero of the respective covariate in column (1) of Table A.1 is less than 0.05.*

Table A.3: Colonial railroads: mobile labor (upper panel) and ex-ante road density (lower panel)

| | Infrastructure discrimination $\Lambda_{\text{mob}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.106$^{***}$ | | | | | |
| | (0.0195) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.101 | | | | |
| | | (0.0672) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.119$^{***}$ | | | |
| | | | (0.0311) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.0982$^{***}$ | | |
| | | | | (0.0314) | | |
| <10KM to railroad | | | | | -0.147$^{***}$ | 0.00884 |
| | | | | | (0.0276) | (0.0244) |
| 10-20KM to railroad | | | | | -0.135$^{***}$ | 0.0152 |
| | | | | | (0.0268) | (0.0264) |
| 20-30KM to railroad | | | | | -0.0792$^{***}$ | 0.00881 |
| | | | | | (0.0268) | (0.0273) |
| 30-40KM to railroad | | | | | -0.0496$^{**}$ | -0.000634 |
| | | | | | (0.0246) | (0.0179) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.884 | 0.884 | 0.884 | 0.884 | 0.885 | 0.884 |

| | Ex-ante road density (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | 0.115$^{***}$ | | | | | |
| | (0.0204) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.0533 | | | | |
| | | (0.105) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | 0.133$^{***}$ | | | |
| | | | (0.0384) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | 0.116$^{***}$ | | |
| | | | | (0.0308) | | |
| <10KM to railroad | | | | | 0.180$^{***}$ | 0.0636$^{*}$ |
| | | | | | (0.0307) | (0.0346) |
| 10-20KM to railroad | | | | | 0.107$^{***}$ | 0.0433 |
| | | | | | (0.0264) | (0.0392) |
| 20-30KM to railroad | | | | | 0.144$^{***}$ | -0.00385 |
| | | | | | (0.0268) | (0.0311) |
| 30-40KM to railroad | | | | | 0.0775$^{***}$ | 0.0458 |
| | | | | | (0.0285) | (0.0325) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.861 | 0.860 | 0.860 | 0.860 | 0.861 | 0.860 |

*Replication of Table 1, but with $\Lambda_{mob}$ (upper panel) and ex-ante infrastructure density $I^c$ (lower panel) as dependent variables, both z-scored.*

Table A.4: Colonial railroads: New investments

| | Infrastructure discrimination $\Lambda^{10\%}_{\text{imm}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.0689*** | | | | | |
| | (0.0227) | | | | | |
| 50 KM of Colonial Placebo Railroads | | -0.0331 | | | | |
| | | (0.0562) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.0850** | | | |
| | | | (0.0357) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.0554* | | |
| | | | | (0.0299) | | |
| <10KM to railroad | | | | | -0.0945*** | -0.0340 |
| | | | | | (0.0321) | (0.0267) |
| 10-20KM to railroad | | | | | -0.111*** | -0.0159 |
| | | | | | (0.0299) | (0.0275) |
| 20-30KM to railroad | | | | | -0.0222 | -0.0163 |
| | | | | | (0.0326) | (0.0385) |
| 30-40KM to railroad | | | | | 0.0497 | 0.0365 |
| | | | | | (0.0307) | (0.0294) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.777 | 0.777 | 0.777 | 0.777 | 0.777 | 0.777 |
| | Infrastructure discrimination $\Lambda^{10\%}_{\text{mob}}$ (z-scored) | | | | Real | Placebo |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.145*** | | | | | |
| | (0.0239) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.181** | | | | |
| | | (0.0764) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.161*** | | | |
| | | | (0.0382) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.147*** | | |
| | | | | (0.0384) | | |
| <10KM to railroad | | | | | -0.202*** | 0.0226 |
| | | | | | (0.0345) | (0.0271) |
| 10-20KM to railroad | | | | | -0.171*** | 0.0304 |
| | | | | | (0.0324) | (0.0302) |
| 20-30KM to railroad | | | | | -0.118*** | 0.0257 |
| | | | | | (0.0315) | (0.0331) |
| 30-40KM to railroad | | | | | -0.0735** | 0.0118 |
| | | | | | (0.0305) | (0.0194) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.859 | 0.858 | 0.859 | 0.859 | 0.860 | 0.858 |

*Replication of Table 1, but with $\Lambda^{10\%}_{imm}$ (upper panel) and $\Lambda^{10\%}_{mob}$ (lower panel) as dependent variables, both z-scored.*

Table A.5: Colonial railroads: no market power (optimal reallocation)

| | Infrastructure discrimination $\Lambda_{\text{imm}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.126*** | | | | | |
| | (0.0357) | | | | | |
| 50 KM of Colonial Placebo Railroads | | -0.0610 | | | | |
| | | (0.0955) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.149*** | | | |
| | | | (0.0508) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.124** | | |
| | | | | (0.0498) | | |
| <10KM to railroad | | | | | -0.174*** | -0.0521 |
| | | | | | (0.0493) | (0.0407) |
| 10-20KM to railroad | | | | | -0.187*** | -0.0739 |
| | | | | | (0.0452) | (0.0453) |
| 20-30KM to railroad | | | | | -0.0472 | -0.0302 |
| | | | | | (0.0530) | (0.0651) |
| 30-40KM to railroad | | | | | 0.0945* | 0.0564 |
| | | | | | (0.0481) | (0.0531) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.236 | 0.235 | 0.236 | 0.235 | 0.237 | 0.235 |
| | Infrastructure discrimination $\Lambda_{\text{mob}}$ (z-scored) | | | | Real | Placebo |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.254*** | | | | | |
| | (0.0504) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.0270 | | | | |
| | | (0.162) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.298*** | | | |
| | | | (0.0697) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.214*** | | |
| | | | | (0.0724) | | |
| <10KM to railroad | | | | | -0.331*** | -0.0494 |
| | | | | | (0.0632) | (0.0588) |
| 10-20KM to railroad | | | | | -0.342*** | -0.0490 |
| | | | | | (0.0580) | (0.0680) |
| 20-30KM to railroad | | | | | -0.233*** | -0.0386 |
| | | | | | (0.0741) | (0.0655) |
| 30-40KM to railroad | | | | | -0.170*** | -0.0428 |
| | | | | | (0.0632) | (0.0545) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.505 | 0.502 | 0.504 | 0.503 | 0.508 | 0.502 |

*Replication of Table 1, but with $\Lambda_{imm}$ (upper panel) and $\Lambda_{mob}$ (lower panel) computed without the assumption of traders having market power (from Atkin and Donaldson, 2015) as dependent variables, both z-scored.*

Table A.6: Colonial railroads: no market power (optimal expansion)

| | Infrastructure discrimination $\Lambda^{10\%}_{\text{imm}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.117*** | | | | | |
| | (0.0336) | | | | | |
| 50 KM of Colonial Placebo Railroads | | -0.120 | | | | |
| | | (0.104) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.127*** | | | |
| | | | (0.0471) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.128** | | |
| | | | | (0.0538) | | |
| <10KM to railroad | | | | | -0.168*** | -0.0777* |
| | | | | | (0.0467) | (0.0443) |
| 10-20KM to railroad | | | | | -0.167*** | -0.0931* |
| | | | | | (0.0417) | (0.0486) |
| 20-30KM to railroad | | | | | -0.0387 | -0.0539 |
| | | | | | (0.0514) | (0.0671) |
| 30-40KM to railroad | | | | | 0.0917* | 0.0634 |
| | | | | | (0.0529) | (0.0569) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.241 | 0.240 | 0.240 | 0.240 | 0.242 | 0.240 |

| | Infrastructure discrimination $\Lambda^{10\%}_{\text{mob}}$ (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.317*** | | | | | |
| | (0.0539) | | | | | |
| 50 KM of Colonial Placebo Railroads | | -0.00492 | | | | |
| | | (0.168) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.352*** | | | |
| | | | (0.0759) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.299*** | | |
| | | | | (0.0781) | | |
| <10KM to railroad | | | | | -0.417*** | -0.0581 |
| | | | | | (0.0703) | (0.0629) |
| 10-20KM to railroad | | | | | -0.401*** | -0.0543 |
| | | | | | (0.0622) | (0.0713) |
| 20-30KM to railroad | | | | | -0.309*** | -0.0399 |
| | | | | | (0.0799) | (0.0707) |
| 30-40KM to railroad | | | | | -0.215*** | -0.0457 |
| | | | | | (0.0702) | (0.0595) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.477 | 0.472 | 0.475 | 0.473 | 0.481 | 0.472 |

*Replication of Table 1, but with $\Lambda^{10\%}_{imm}$ (upper panel) and $\Lambda^{10\%}_{mob}$ (lower panel) computed without the assumption of traders having market power (from Atkin and Donaldson, 2015) as dependent variables, both z-scored.*

Table A.7: Colonial railroads and local amenities

| | Housing (z-scored) | | | | Real | Placebo |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.0161 | | | | | |
| | (0.0134) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.0852 | | | | |
| | | (0.0581) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.00118 | | | |
| | | | (0.0260) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.0291 | | |
| | | | | (0.0337) | | |
| <10KM to railroad | | | | | -0.0133 | 0.0286 |
| | | | | | (0.0195) | (0.0260) |
| 10-20KM to railroad | | | | | -0.0108 | 0.0292 |
| | | | | | (0.0176) | (0.0250) |
| 20-30KM to railroad | | | | | 0.0125 | 0.0159 |
| | | | | | (0.0203) | (0.0202) |
| 30-40KM to railroad | | | | | 0.00718 | 0.0210 |
| | | | | | (0.0170) | (0.0159) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 8580 | 8580 | 8580 | 8580 | 8580 | 8580 |
| R2 | 0.0533 | 0.0533 | 0.0533 | 0.0533 | 0.0533 | 0.0533 |
| | Housing per capita (z-scored) | | | | Real | Placebo |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 50 KM of Colonial Railroads | -0.0583*** | | | | | |
| | (0.0151) | | | | | |
| 50 KM of Colonial Placebo Railroads | | 0.0195 | | | | |
| | | (0.0541) | | | | |
| 50 KM of Colonial Railroads for Military Purposes | | | -0.0441* | | | |
| | | | (0.0224) | | | |
| 50 KM of Colonial Railroads for Mining Purposes | | | | -0.0389 | | |
| | | | | (0.0250) | | |
| <10KM to railroad | | | | | -0.0656*** | 0.00740 |
| | | | | | (0.0214) | (0.0204) |
| 10-20KM to railroad | | | | | -0.0619*** | -0.0114 |
| | | | | | (0.0221) | (0.0219) |
| 20-30KM to railroad | | | | | -0.0569*** | -0.00989 |
| | | | | | (0.0201) | (0.0172) |
| 30-40KM to railroad | | | | | -0.0247 | 0.00361 |
| | | | | | (0.0184) | (0.0150) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 8384 | 8384 | 8384 | 8384 | 8384 | 8384 |
| R2 | 0.464 | 0.464 | 0.464 | 0.464 | 0.464 | 0.464 |

*Replication of Table 1, but with total model-implied housing $H_i$ (upper panel) and per-capita model implied housing $h_i$ (lower panel) as dependent variables, both z-scored.*

Table A.8: Regional favoritism: Interaction effects

| | Relative road expenditure | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A: Kenyan road expenditure over time* | | | | |
| $\Lambda_{\text{imm}}$ (non-democracy) | -2.550** | | -2.548** | |
| | (1.114) | | (1.111) | |
| $\Lambda_{\text{imm}}$ (democracy) | | 0.503 | | 0.344 |
| | | (0.756) | | (0.740) |
| Coethnic district | 1.433*** | 1.206** | 1.577*** | 0.940* |
| | (0.402) | (0.588) | (0.520) | (0.541) |
| $\Lambda_{\text{imm}} \times$ Coethnic district | | | -0.378 | 0.793* |
| | | | (0.725) | (0.446) |
| Year and District FE | Yes | Yes | Yes | Yes |
| (Demographic, political, economic, geographic) $\times$ trend | Yes | Yes | Yes | Yes |
| N | 451 | 410 | 451 | 410 |
| R2 | 0.362 | 0.215 | 0.363 | 0.216 |

*Replication of Table 2, columns 3-4. Yet also adding an interaction effect $\Lambda_{dt} \times coethnic_{dt}$.*

Table A.9: Regional favouritism: Optimal expansion immobile labor (top), mobile labor reallocation (middle), mobile labor expansion (bottom)

| | Discrimination $\Lambda_{\text{imm}}^{10\%}$ | | Relative road expenditure | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A: entire sample* | | | | |
| Ever in power dummy | -0.103* | | | |
| | (0.0532) | | | |
| log(1 + Total years in power) | | -0.0335 | | |
| | | (0.0215) | | |
| *Panel B: Kenyan road expenditure over time* | | | | |
| $\Lambda_{\text{imm}}^{10\%}$ (non-democracy) | | | -0.864 | |
| | | | (0.724) | |
| $\Lambda_{\text{imm}}^{10\%}$ (democracy) | | | | 0.00296 |
| | | | | (0.447) |
| Year and District FE | | | Yes | Yes |
| (Demographic, political, economic, geographic) $\times$ trend | | | Yes | Yes |
| Country FE | Yes | Yes | | |
| Geography Controls | Yes | Yes | | |
| N | 10158 | 10158 | 451 | 410 |
| R2 | 0.777 | 0.777 | 0.356 | 0.215 |

| | Discrimination $\Lambda_{\text{mob}}$ | | Relative road expenditure | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A: entire sample* | | | | |
| Ever in power dummy | -0.0374 | | | |
| | (0.0607) | | | |
| log(1 + Total years in power) | | -0.0111 | | |
| | | (0.0275) | | |
| *Panel B: Kenyan road expenditure over time* | | | | |
| $\Lambda_{\text{mob}}$ (non-democracy) | | | -0.239 | |
| | | | (1.346) | |
| $\Lambda_{\text{mob}}$ (democracy) | | | | 0.582 |
| | | | | (0.931) |
| Year and District FE | | | Yes | Yes |
| (Demographic, political, economic, geographic) $\times$ trend | | | Yes | Yes |
| Country FE | Yes | Yes | | |
| Geography Controls | Yes | Yes | | |
| N | 10158 | 10158 | 451 | 410 |
| R2 | 0.884 | 0.884 | 0.354 | 0.215 |

| | Discrimination $\Lambda_{\text{mob}}^{10\%}$ | | Relative road expenditure | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Panel A: entire sample* | | | | |
| Ever in power dummy | -0.0463 | | | |
| | (0.0775) | | | |
| log(1 + Total years in power) | | -0.0145 | | |
| | | (0.0335) | | |
| *Panel B: Kenyan road expenditure over time* | | | | |
| $\Lambda_{\text{mob}}^{10\%}$ (non-democracy) | | | -0.810 | |
| | | | (1.115) | |
| $\Lambda_{\text{mob}}^{10\%}$ (democracy) | | | | 0.0635 |
| | | | | (0.567) |
| Year and District FE | | | Yes | Yes |
| (Demographic, political, economic, geographic) $\times$ trend | | | Yes | Yes |
| Country FE | Yes | Yes | | |
| Geography Controls | Yes | Yes | | |
| N | 10158 | 10158 | 451 | 410 |
| R2 | 0.858 | 0.858 | 0.355 | 0.215 |

*Replication of Table 2, yet with $\Lambda_{imm}^{10\%}$ (upper panel), $\Lambda_{mob}$ (middle panel), and $\Lambda_{mob}^{10\%}$ (lower) panel as dependent variables, all z-scored.*

## B Numerically solving the planner's problem

The full planner's problem on page 5 consists of a very large number of choice variables and hence requires vast computation efforts when solved directly. Fortunately, Fajgelbaum and Schaal (2020) provide guidance on how to transform this *primal* problem into its much simpler *dual* representation. The following section illustrates how to use their derivation to numerically solve my version of the model.

To show how a unique global optimum exists, first note that every constraint of the social planner's problem is convex but potentially for the *Balanced Flows Constraint*. However, the introduction of congestion causes even the *Balanced Flows Constraint* to be convex if $\beta > \gamma$. To see this, note that every part of the lengthy constraint is linear, but for the interaction term $Q^n_{i,k}\tau^n_{i,k}(Q^n_{i,k}, I_{i,k})$ representing total trade costs. Since $\tau^n_{i,k}$ was parameterised as in (1), this expands to

$$Q^n_{i,k}\tau^n_{i,k}(Q^n_{i,k}, I_{i,k}) = \delta^\tau_{i,k}\frac{(Q^n_{i,k})^{1+\beta}}{I^\gamma_{i,k}} \tag{A.1}$$

which is convex if $\beta > \gamma$. Under this condition, the social planner's problem is to maximise a concave objective over a convex set of constraints, guaranteeing that any local optimum is indeed a global maximum.[A.1] $\beta > \gamma$ describes a notion of congestion dominance: increased infrastructure expenditure might alleviate the powers of congestion, but it can never overpower it. It precludes corner solutions in which all available concrete is spent on one link, all but washing away trade costs and leading to overwhelming transport flows on this one edge. If $\beta > \gamma$, geography always wins.

Consider first the full Lagrangian of the primal planner's problem

$$\mathcal{L} = \sum_i L_i u(c_i) - \sum_i \lambda^C_i\left[L_i c_i - \left(\sum_{n=1}^N (C^n_i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right] \tag{A.2}$$

$$- \sum_i\sum_n \lambda^P_{i,n}\left[C^n_i + \sum_{k\in N(i)} Q^n_{i,k}(1+\tau^n_{i,k}(Q^n_{i,k}, I_{i,k})) - Y^n_i - \sum_{j\in N(i)} Q^n_{j,i}\right] \tag{A.3}$$

$$- \lambda^I\left[\sum_i\sum_{k\in N(i)} \delta^i_{i,k} I_{i,k} - K\right] - \sum_i\sum_{k\in N(i)} \zeta^S_{i,k}\left[I_{i,k} - I_{k,i}\right] \tag{A.4}$$

$$+ \sum_i\sum_{k\in N(i)}\sum_n \zeta^Q_{i,k,n} Q^n_{i,k} + \sum_i\sum_n \zeta^C_{i,n} C^n_i + \sum_i\sum_n \zeta^c_i c_i - \sum_i\sum_{k\in N(i)} \zeta^I_{i,k}\left[4 - I_{i,k}\right] \tag{A.5}$$

This is a function of the choice variables $(C^n_i, Q^n_{i,k}, c_i, I_{i,k})$ in all dimensions $\langle i,k,n\rangle$ and the Lagrange multipliers $(\boldsymbol{\lambda^C}, \boldsymbol{\lambda^P}, \boldsymbol{\lambda^I}, \boldsymbol{\zeta^Q}, \boldsymbol{\zeta^C}, \boldsymbol{\zeta^c}, \boldsymbol{\zeta^I})$ also in $\langle i,k,n\rangle$. Standard optimisation yields first-

[A.1] This is Fajgelbaum and Schaal Proposition 1.

order conditions which can be collapsed to the following set of equations

$$
\begin{aligned}
c_i &= \left(\frac{1}{\alpha}\Big(\sum_{n'}(\lambda^P_{i,n'})^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}\right)^{\frac{1}{\alpha-1}} \\
C^n_i &= \left[\frac{\lambda^P_{i,n}}{(\sum_{n'}(\lambda^P_{i,n'})^{1-\sigma})^{\frac{1}{1-\sigma}}}\right]^{-\sigma} L_i c_i \\
Q^n_{i,k} &= \left[\frac{1}{1+\beta}\frac{I^\gamma_{i,k}}{\delta^\tau_{i,k}}\max\Big\{\frac{\lambda^P_{k,n}}{\lambda^P_{i,n}}-1,0\Big\}\right]^{\frac{1}{\beta}} \\
I_{i,k} &= \max\Bigg\{\Bigg[\frac{\kappa}{\lambda^I(\delta^I_{i,k}+\delta^I_{k,i})}\Bigg(\sum_n \max\Big\{(\delta^\tau_{i,k})^{-\frac{1}{\beta}}\lambda^P_{i,n}\Big(\frac{\lambda^P_{k,n}}{\lambda^P_{i,n}}-1\Big)^{\frac{1+\beta}{\beta}},0\Big\} \\
&\qquad +\sum_n \max\Big\{(\delta^\tau_{k,i})^{-\frac{1}{\beta}}\lambda^P_{k,n}\Big(\frac{\lambda^P_{i,n}}{\lambda^P_{k,n}}-1\Big)^{\frac{1+\beta}{\beta}},0\Big\}\Bigg)\Bigg]^{\frac{\beta}{\beta-\gamma}},4\Bigg\}
\end{aligned}
\tag{A.6}
$$

These directly follow the more general framework outlined in the technical appendix of Fajgelbaum and Schaal applied to my version of the model. In the final equation denoting optimal infrastructure supply, $\kappa = \gamma(1+\beta)^{-\frac{1+\beta}{\beta}}$, and the multiplier $\lambda^I$ is such that adherence to the *Network Building Constraint* is ensured. Through these algebraic manipulations, I have expressed all choice variables as functions of merely the Lagrange parameters $\boldsymbol{\lambda}^{\boldsymbol{P}}$ over dimensions $\langle i,k,n\rangle$. I can hence recast the entire Lagrangian in much simpler form as

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\lambda}, x(\boldsymbol{\lambda})) &= \sum_i L_i u(c_i(\boldsymbol{\lambda})) \\
&- \sum_i\sum_n \lambda^P_{i,n}\Big[C^n_i(\boldsymbol{\lambda}) + \sum_{k\in N(i)} Q^n_{i,k}(\boldsymbol{\lambda})(1+\tau^n_{i,k}(Q_{i,k}(\boldsymbol{\lambda})^n, I_{i,k}(\boldsymbol{\lambda}))) - Y^n_i - \sum_{j\in N(i)} Q^n_{j,i}(\boldsymbol{\lambda})\Big]
\end{aligned}
\tag{A.7}
$$

where $x(\boldsymbol{\lambda})$ denote the choice variables as functions of the Lagrange parameters as derived above. Fajgelbaum and Schaal note that thanks to complementary slackness, all other constraints can be readily dropped from consideration and only the *Balanced Flows Constraint* remains part of the problem.

As Fajgelbaum and Schaal further explain, the dual of this problem can now be conceived as the minimisation of

$$\min_{\boldsymbol{\lambda}\geq 0} \mathcal{L}(\boldsymbol{\lambda}, x(\boldsymbol{\lambda}))$$

which is an optimisation problem over merely $\|\boldsymbol{\lambda}^{\boldsymbol{P}}\| = I \times N$ variables. Fajgelbaum and Schaal interpret $\boldsymbol{\lambda}^{\boldsymbol{P}}$ as a field of prices varying over goods and locations. I am left only to minimise equation (A.7) to obtain the price-field $\boldsymbol{\lambda}^{\boldsymbol{P}}$. I implement constrained optimisations within the `fmincon` environment in MATLAB and achieve fairly fast convergence. Solving for smaller networks (like Rwanda or Djibouti) is a matter of seconds, yet the largest countries (Algeria, Angola, DRC, and Sudan) each take about a day of computation time (on a five-year old device, nonetheless). Plugging the derived $\boldsymbol{\lambda}^{\boldsymbol{P}}$ parameters into the various FOCs in (A.6) yields the optimal transport network $I_{i,k}$, trade flows between locations $Q^n_{i,k}$, and consumption patterns $C^n_i$ and $c_i$.

# C Calibration details

## C.1 Calibrating structural parameters

**Tradable budget share $\alpha$** This parameter captures the Cobb-Douglas budget share households spend on tradable goods. This includes consumption of home-produced goods (ie. food staples) that could in principle be traded. To calibrate this expenditure share, I rely on analysis by Porteous (2022) who calibrate budget shares using data from Nigeria and Angola. They set the share for agricultural tradables to 0.4, non-agricultural tradables to 0.3, and non-agricultural non-tradables to 0.3, respectively. Summing together the two tradables categories, I calibrate $\alpha = 0.7$.

**Elasticity of substitution $\sigma$** There are notoriously many divergent estimates for the trade elasticity of substitution $\sigma$ (for a review see Head and Mayer, 2014). I rely on a recent review paper by Atkin and Donaldson (2022), which summarises the literature estimating trade elasticities in developing country contexts and recommends using a parameter of $\sigma = 5$. This is slightly higher than the parameter used by Fajgelbaum and Schaal (2020) in their calibration of interregional European trade ($\sigma = 4$), which is not too surprising given that homogeneous agricultural products play a more important role in African regional trade vis-a-vis Europe.

## C.2 Calibrating trade costs $\tau_{i,k}$

As described in the main text, I follow Fajgelbaum and Schaal (2020) to assume the following functional form for trade costs between locations $i, k$ for good $n$:

$$\tau^n_{i,k}(Q^n_{i,k}, I_{i,k}) = \delta^\tau_{i,k} \frac{(Q^n_{i,k})^\beta}{I^\gamma_{i,k}}$$

Trade flows $Q^n_{i,k}$ on the link increase trade costs through the congestion elasticity parameter $\beta$, infrastructure $I_{i,k}$ on the link decreases trade costs through the infrastructure elasticity $\gamma$. $\delta^\tau_{i,k}$ captures inherent trade costs between the locations, which could in principle depend on any exogeneous or geographical characteristics of the link.

**Infrastructure elasticity $\gamma$** As described in the main text, infrastructure $I_{i,k}$ is parameterised as the average attainable speed between locations $i$ and $k$. Hence, $\gamma$ captures the elasticity of trade costs to speed improvements. To calibrate it, I use data from a survey of trucking costs across Africa from Teravaninthorn and Raballand (2009). In particular, the authors investigate transport costs across the "Southern Corridor" from Durban to Dar and estimate that a 20% reduction of delays at border posts across this corridor would reduce total aggregate transport costs by 3-4% (Teravaninthorn and Raballand, 2009, Table 1.3). Delays at border posts are one of the major ways international truck traffic in Africa gets slowed down, so this yields helpful insights into the cost benefits of speeding up transport. The authors note that current delays at the border posts along this corridor amount to about four days or 96hrs (Teravaninthorn and Raballand, 2009, page 9). A search of the route on Google Maps indicates that without border delays, the total driving time of the corridor is about 59hrs, so that a total of 62% of the entire trip is currently spent waiting. The authors further report that about 29.8% of all kilometre-weighted trips in the region go along this corridor. Putting it all together; a reduction of 20% of the 62% of time spent waiting at borders for 29.8% of trips amounts to a 3-4% reduction in aggregate transport costs, or

$$(1 - 0.2 \times 0.62 \times 0.298)^{\gamma} = 1 - 0.035$$

or $\gamma = 0.946$.

**Congestion elasticity** $\beta$ To calibrate how much additional cars on the road affect trade costs, I rely on the exercise by Fajgelbaum and Schaal (2017), who compile and aggregate estimates from Wang et al. (2011) of the relationship between car density and speed using data from Georgia, USA. They estimate an average relationship of

$$\text{Speed}_{i,k} \propto Q_{i,k}^{1.2446}$$

Since I posit the relationship between speed and trade costs to be $\gamma = 0.946$, I nest the two and arrive at

$$\beta = 1.2446 \times \gamma = 1.1774$$

**Exogeneous trade costs** $\delta^{\tau}_{i,k}$ To calibrate $\delta^{\tau}_{i,k}$, I make use of the work of Atkin and Donaldson (2015), who investigate price gaps across space in Ethiopia and Nigeria. The authors have barcode-level data of the same product at different locations across the two countries and find that, in general, prices become higher in areas further away from the supposed origin location of the product (ie. the main port of entry for imported products, or the factory location for home-produced products). They posit that the absolute price gap of a product sold at a location $k$ to the price at the origin location $i$ is a function of transport costs $t(\cdot)$ and a markup charged by intermediaries with market power $\mu(\cdot)$:[A.2]

$$P_k - P_i = t\left(\mathbf{X}_k\right) + \mu\left(\mathbf{X}_k\right),$$

Both $t$ and $\mu$ are allowed to vary according to observable characteristics $\mathbf{X}_k$ of the selling location $k$.

I follow the evidence brought forward in Atkin and Donaldson, who analyse price gaps mainly as a function of distance between origin and destination location. In particular, the authors estimate a log-linear relationship between the two:

$$P_k - P_i = \xi \log\left(\text{Distance (miles)}_{i,k}\right)$$

They estimate that (among trading pairs), this elasticity $\widehat{\xi}$ is 0.0248 in Ethiopia and 0.0254 in Nigeria (Atkin and Donaldson (2015), Table 2, columns (2) and (5)). Interestingly, this elasticity is mediated by market power: more remote areas get charged lower markups by intermediaries who realise that inhabitants of these areas tend to be poorer and thus on a more elastic part of their demand curve. The "pure" distance elasticity of transport costs $t$ is larger: 0.0374 in Ethiopia and 0.0558 in Nigeria.

In the Fajgelbaum and Schaal (2020) framework, price gaps for a good $n$ between locations are

[A.2] This corresponds to equation (2) of Atkin and Donaldson (2015), where I have relabeled their notation for transport costs $\tau$ as $t$, to avoid confusion with the endogeneous trade costs $\tau$ in my model.

given by the ad-valorem trade cost parameter $\tau^n_{i,k}$:

$$\begin{aligned} P^n_k - P^n_i &= P^n_i \tau^n_{i,k} \\ \implies \frac{P^n_k - P^n_i}{P^n_i} &= \tau^n_{i,k} \\ \implies \frac{\widehat{\xi} \log \left(\text{Distance (miles)}_{i,k}\right)}{P^n_i} &= \tau^n_{i,k} \end{aligned}$$

Hence, to "translate" the Atkin and Donaldson evidence on absolute (dollar-value) price gaps to my object of inherent ad-valorem "iceberg" trade costs, I need to divide their predicted values of price gaps by the origin price of goods in the barcode-level dataset. Atkin and Donaldson provide these moments in their main text (page 24): the average product in their Ethiopia dataset costs $\overline{P_i} = 43$ cents, while the average product in their Nigeria dataset costs 1.03 dollars. Putting this together, the inherent distance elasticity of ad-valorem price gaps $\widehat{\xi}/\overline{P_i}$ is $0.0248/0.43 = 0.0577$ in Ethiopia and $0.0254/1.03 = 0.02466$ in Nigeria.

However, in the Fajgelbaum and Schaal framework, trade costs are, furthermore, subject to congestion and infrastructure effects, which are non-linear and endogeneous. In other words, the above calibration would be correct for shipping 1 good at speed of 1km/hr across the network, or $Q$ goods at speed $Q^{\beta/\gamma}$. If the actually shipped quantity $Q$ were higher (lower) than that in the current observed equilibrium, my calibration would overstate (understate) this elasticity. Since Atkin and Donaldson (2015) do not have access to data on quantities shipped (which is generally hard to come by), we have no way to directly test for this.

I hence run a fixed-point algorithm to account for this: I treat the price gaps reported in Atkin and Donaldson as equilibrium values, and adjust $\delta^\tau_{i,k}$ until I match them.

In particular, I start by setting $\delta^\tau_{i,k,0} = \widehat{\xi}/\overline{P_i}$ from above, compute the current equilibrium, use it to determine the model-implied ratio of $R \equiv (Q^n_{i,k})^\beta / I^\gamma_{i,k}$, update $\delta^\tau_{i,k,\ell+1} = \delta^\tau_{i,k,\ell}/R$, and iterate until convergence (ie. until equilibrium price gaps coincide with the ones reported in Atkin and Donaldson). This is achieved at

$$\begin{aligned} \delta_0^{\tau,\text{ETH}} &= \frac{\widehat{\xi}^{\text{ETH}}}{\overline{P}_i^{\text{ETH}}} \cdot 3.915 = \frac{0.0248 \cdot 3.915}{0.43} = 0.2258 \qquad \text{for Ethiopia} \\ \delta_0^{\tau,\text{NGA}} &= \frac{\widehat{\xi}^{\text{NGA}}}{\overline{P}_i^{\text{NGA}}} \cdot 0.2414 = \frac{0.0254 \cdot 0.2414}{1.03} = 0.006 \qquad \text{for Nigeria} \end{aligned}$$

Averaging the two estimates yields an elasticity of 0.1159, and hence an inherent trade cost term of

$$\delta^\tau_{i,k} = 0.1159 \times \log \left(\text{Distance (miles)}_{i,k}\right)$$

As a robustness exercise, I also use estimates the authors use to purge spatial price gaps of the impact of market power. These can readily be read of columns (3) and (6) of Table 2 of their paper. Using the same procedure as above, I obtain and average elasticity of

$$\delta^{\tau,\text{no-market-power}}_{i,k} = 0.3525 \times \log \left(\text{Distance (miles)}_{i,k}\right)$$

as an inherent trade cost term in a world where intermediaries have no market power.

## C.3 Additional details on historical analysis

To calibrate my model to historical roads data from Kenya, I use data from Burgess et al. (2015) and Jedwab and Storeygard (2022). The Burgess et al. (2015) paper provides population, road expenditure shares, as well as centroids of Kenyan districts and their ethnic affiliation.

Using digitised Michelin maps by Jedwab and Storeygard (2022), I merge each road segment to their nearest district-centroid. Segments come with a roads classification (highway, paved, unpaved, and so on), which I translate into average speeds using the same methodology as Jedwab and Storeygard. This leaves me with a *district-level* measure of infrastructure density. To translate this into *edge-level* values, I calibrate the amount of infrastructure $I_{j,k}$ between districts $j$ and $k$ as the average infrastructure density of $j$ and $k$. I treat two districts as adjacent if the voronoi-diagrams around their centroids touch.

I don't have information on population and productivity of each of Kenya's districts across time. I do, however, have population information for 1962. I project this outward using country-wide population totals (ie. assuming that the relative population share of each district stays constant over time). I also calibrate productivity such that the country's total production matches data on overall GDP of Kenya in each year (using data on population totals and GDP from the World Bank). I again treat the 4 most populous districts as producing their own variety, with the remaining districts producing an "agricultural" fifth variety. Note that I do not include a foreign-country buffer around Kenya, as I don't have historical information on Tanzania's, Uganda's, (South) Sudan's, Ethiopia's, or Somalia's population and productivity.

Assuming all structural parameters from the rest of the paper (which are calibrated on much more recent data) stay the same, this leaves me with enough to compute the optimally reallocated and expanded network with mobile and immobile labor, at each point at which a digitised Michelin map exists.

# D Foreign aid and infrastructure discrimination

To investigate whether international development aid is quantitatively associated to my measure of trade network inefficiency, I make use of two datasets of geo-referenced aid flows to Africa. Firstly, AidData (2017) in cooperation with the World Bank, tracks over 5,600 lending lines from the World Bank to African nations and reports precise coordinates of over 60,000 projects financed through these funds, totalling more than 300 billion US dollars. The sample comprises all projects approved between 1996–2014. As Strandow et al. (2011) describe, attributing projects to locations relies on a double-blind coding procedure of various World Bank documents. Secondly, I explore patterns from a similar database on Chinese aid projects by Strange et al. (2017). They resort to reports from numerous local and international media outlets to track official and unofficial financing lines to over 1,500 projects worth 73 billion US dollars in the period 2000–2011.[A.3]

For the purpose of this study, I exclude aid projects with no clear-cut geographical target like unconditional lending lines to the central government or assistance for political parties. I also exclude flows with unknown or only vague information on eventual project location.[A.4] I also ignore

[A.3]As Strange et al. point out, media reports are often based on initial press releases and do not necessarily follow up on the eventual disbursement of every promised dollar. In that, the dataset is likely to capture Chinese funding *commitments* rather than actual *disbursements*. Insofar as donors usually commit to more than they eventually deliver, these figures present an upper bound of realised development assistance. Furthermore, while AidData (2017) claim their dataset on World Bank projects to be exhaustive, the dataset on Chinese aid will naturally miss some unofficial flows, as significant parts of Chinese involvement remain untracked.

[A.4]Specifically, I exclude all projects with a precision code of more than 3 – this corresponds to projects only identified at projects which were still under construction or otherwise not fully completed by the end of 2017. Together, these steps truncate the World Bank sample by 35% and the China sample by 52%. In Figure A.8, I map the spatial distribution of aid projects from both remaining samples. I aggregate the total value of aid disbursements from the remaining 10,786 World Bank projects and 1,420 Chinese projects onto the grid cell level. Of the 10,158 grid cells of my sample, more than 21% have received some form of assistance from either source.[A.5]

Do donor institutions identify places most in need of additional infrastructure? I employ various indicators of aid provision in the standard grid cell level framework based on equation (6). I rely on two measures to quantify the prevalence of foreign aid: the total value of aid disbursements to a grid cell in 2011 US dollars and the number of distinct project sites within a given cell. I also put additional emphasis on infrastructure by separately analysing variation in funds going only to infrastructure projects in the transportation sector.

Table A.10 reports results. Columns (1–4) investigate the spatial distribution of World Bank assistance. The estimates reveal seemingly opposing objectives between the Bank and the social planner. Negative estimates in columns (1) through (4) imply that grid cells receiving more World Bank assistance score lower on the discrimination index $\Lambda_i$. Every additional million US dollar flowing into an area is associated with the grid cell being about 0.004 standard deviations too well off. Focusing on transport sector projects only, results are qualitatively similar, yet much stronger. The average transport infrastructure project size of around 3 million US dollars goes to grid cells which stand to lose 0.005 standard deviations of welfare under the reallocation exercise. Similar effects hold on the extensive margin reported in columns (3) and (4). Columns (5–8) present very similar results for Chinese aid. Chinese assistance also systematically flows into privileged cells, with intensive margin point estimates of the association ranging between a quarter and a tenth of the World Bank results. On the extensive margin, more Chinese projects are similarly associated with higher trade network imbalances. For each new development site financed by China in a certain cell, the social planner intervenes and allocates about 0.03-0.04 standard deviations of welfare away from the cell (columns 7–8).

These relationships should by not interpreted as causal effects. Since the placement of aid projects is not random, numerous other channels could account for the patterns depicted in Table A.10. The donor's investment strategies might for example be motivated by increasing returns to scale. If the World Bank believes in an environment with multiple equilibria, where small initial investments set in motion a dynamic of spillover externalities, labour migration, and follow-up investments, it is often the right decision to fund projects in places that will not immediately harness their full capabilities (Krugman, 1991; Duranton and Venables, 2017). These investments will necessarily appear inefficient in promoting optimal trade *today*, yet spur transformative development *tomorrow* (see Michaels et al., 2021). Embedding the reallocation exercise in a New Economic Geography framework of increasing returns and labour mobility might be a valuable extension to better evaluate specific place-based policies.

province-level or above. The remaining entries are geo-coded either exactly (61%), within a 25 kilometre radius (4%), or with municipality-level precision (35%) (Strandow et al., 2011)

[A.5] All disbursements are adjusted to 2011 US dollars. For projects with multiple sites, I assume total disbursement value to be split evenly between sites. On average, these cells receive aid volumes of more than 30 million US dollars. The area receiving the most total World Bank funding is the grid cell containing Uganda's capital Kampala. The biggest beneficiary of Chinese development assistance is a grid cell in the south of Congo-Kinshasa, where Chinese funds of almost 5 billion US dollars helped construct a vast copper mining infrastructure.

Table A.10: International aid and local infrastructure discrimination

| | Λ: World Bank | | | | Λ: China | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Total disbursements (mil $) | -0.00187*** | | | | -0.0000909 | | | |
| | (0.000635) | | | | (0.0000684) | | | |
| Total transport sector disbursements (mil $) | | -0.00159 | | | | -0.000256* | | |
| | | (0.00180) | | | | (0.000133) | | |
| Number of projects | | | -0.0116*** | | | | -0.0127*** | |
| | | | (0.00253) | | | | (0.00450) | |
| Number of transport sector disbursements | | | | -0.0147*** | | | | -0.0278** |
| | | | | (0.00367) | | | | (0.0133) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 | 10158 |
| R2 | 0.764 | 0.764 | 0.765 | 0.764 | 0.764 | 0.764 | 0.764 | 0.764 |

*Grid cell level estimations of equation* (6) *with z-scored local infrastructure discrimination* $\tilde{\Lambda}_i$ *as dependent variable and different measures of foreign aid flows into grid cells as explanatory covariates. Columns (1–4) investigate World Bank assistance. Column (1) analyses total disbursement value from World Bank projects approved from 1996–2014 in 2011 US dollars, which were completed by 2017. (2) only uses a subset of projects in the transport sector. (3)–(4) use the same data but focus on the number of distinct project sites within each grid cell. Columns (5–8) repeat the same estimations, but with data on Chinese aid projects between 2000–2011. Geography controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, indicators for the 12 predominant agricultural biomes, indicators for whether a cell is a capital, within 25 KM of a natural harbour, navigable river, or lake, the fourth-order polynomial of latitude and longitude, and an indicator of whether the grid cell lies on the border of a country's network. They also include population, night lights, and ruggedness. Chinese aid data are more likely to reflect commitments rather than actual disbursements. Standard errors are clustered on the* $3 \times 3$ *degree level and are shown in parentheses.*