Inefficient African Trade Networks

Tilman Graff – University of Oxford

9th February 2018

1 Introduction

2 A model of Optimal Transport Networks

Are African roads where they should be? To gain a notion of transport network efficiency, the first step is to identify the optimal allocation of roads for a given country. If a social planner were to observe a country's geography, where would she place highways and footpaths, and which regions would she leave unconnected? Intuitively, this optimal network should succeed in connecting regions with a heavy interest in beneficial trade, place roads only where they are cheap to build, and avoid overspending on roads that nobody ever uses.

Optimising over the space of possible network has proven to be challenging to researchers and policymakers alike. The many contingencies and non-linearities of the network structure make it hard to obtain a unique closed-form solution to the problem. The large number of possible connections between locations have, furthermore, led to computational challenges. Researchers and practitioners have hence often been left only with slow, iterative algorithms to obtain a notion of local network optimality. A recent contribution by Fajgelbaum and Schaal (2017), however, manages to circumvent these issues. Under a number of reasonable assumptions, it paves the way for obtaining closed-form solutions to the problem of optimal transport networks without straining computational capabilities all too much. It thereby allows for nesting many of the standard models of the trade literature and can thus flexibly be tailored to the question at hand. This thesis harnesses a version of the Fajgelbaum and Schaal framework to compute the optimal transport network for every African country. The model is introduced in the following paragraphs, chapter 3 then describes the procedure used to bring the model to the data.

2.1 Geography

Following the set-up and notation of Fajgelbaum and Schaal (2017), I consider a set of locations $\mathcal{I} = \{1, ..., I\}$. Each location $i \in \mathcal{I}$ inhabits a number of homogeneous consumers L_i. This number is treated as given and fixed for every location, such that consumers are not allowed to move between locations. Each consumer has an identical set of preferences characterised by

$$u = c^{\alpha}$$

where c denotes per capita consumption. Due to the symmetry of consumers in every location, per capita consumption (and therefore utility) is identical for every consumer within each location, such that every consumer in location i consumes c_i. It is sensible to further denote

$$C_i = L_i c_i$$

as total consumption in location i.

There are N goods denoted by $n = \{1, ..., N\}$. Total consumption in each location is defined as the CES aggregation of these goods

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

where σ denotes the standard elasticity of substitution and C_i^n denotes the supply of good n in location i. Locations specialise in the production of goods such that each location only supplies one variety $n \in N$. Let

$$Y_i = Y_i^n$$

denote the total production output of location i. In contrast to Fajgelbaum and Schaal, my version of the model takes output levels as given and hence does not take a stand on how the output is produced. It is conceivable to think of a production function in which consumers produce goods themselves or an environment in which capital does the work for them. Importantly, however, by remaining agnostic about the production process, output levels are treated as fixed throughout the analysis. The only assumption about the supply side of the economy is perfect specialisation on one of the goods, which does not allow locations to choose between different production opportunities.

2.2 Network Topography

Locations $\mathcal{I}$ can be conceived as representing the nodes of an undirected network graph. Each location i is directly connected to a set of neighbours $N(i) \in \mathcal{I}$. I consider locations to lie on a two-dimensional rectangular grid where each node is connected to its eight surrounding nodes to the North, North-East, East, and so on. Nodes at the border of the network graph might have fewer than eight neighbours. Let $\mathcal{E}$ denote the set of edges connecting neighbouring nodes and note that $(\mathcal{I}, \mathcal{E})$ fully describes the underlying network topography.

All goods can be traded within the network. Let $Q_{i,k}^n$ denote the total flow of good n travelling between nodes i and k. The network shapes trade flows in the sense that goods can only be shipped between neighbouring nodes. However, nothing prevents goods to travel far distances through the network by passing multiple locations after each other. To send a good on to a far away destination, the optimal trade route will potentially make multiple stop-overs in intermediate locations. The entire flow-topography of the trade network can hence be modelled simply by considering flows between neighbouring nodes.[1]

Shipping goods from location i to location $k \in N(i)$ incurs trade costs, which are modelled in the canonical iceberg form. In order for 1 unit of good n to arrive at location k, origin location i has to send $(1+\tau_{i,k}^n)$ units on its way. $\tau_{i,k}^n$ can be understood as a tax on transport, causing a fraction of goods to not arrive in the

[1]While I follow the notation of Fajgelbaum and Schaal (2017), the set up so far is very similar to many of the recent contributions in the theory of trade in networks, like Allen and Arkolakis (2016) or Galichon (2016).

destination location. This iceberg formulation has analytical advantages, as it makes modelling an entire transport sector superfluous. Importantly, barriers to trade are the dimension through which geography is introduced to the model. It is now costly for goods to travel elaborate and long routes through the network and notions of distance and connectedness start to play a role. The main assumption in Fajgelbaum and Schaal's derivation of optimal networks is then about the functional form of $\tau^n_{i,k}$. In particular, the authors model iceberg trade costs for shipping good n between neighbouring locations i and k as

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

where $I_{i,k}$ is defined as the level of *infrastructure* on the edge between nodes i and k. Since model parameters are restricted at $\delta^\tau_{j,k}, \beta, \gamma > 0$, more infrastructure on a given link decreases the cost of trading between them. $I_{i,k}$ will later be calibrated more succinctly, but for now it suffices to picture anything that might make trade costs between two locations smaller, like broader and better roads, less detours, faster train tracks, an the like. In a first-best scenario, infrastructure between all nodes would be set infinitely high to wash away all trade costs and enable goods to travel seamlessly through the entire network. As $I_{i,k}$ is independent from n, higher infrastructure between two nodes will benefit the trade of all goods travelling on this edge. $\delta^\tau_{j,k}$ is a scaling parameter, which allows trade costs to be flexibly adjusted for any given origin-destination pair. Note that $\delta^\tau_{j,k}$ is also independent of good n. In later calibrations, the main component of $\delta^\tau_{j,k}$ will be the geographical distance between locations i and k. Distance is thought to increase trade costs independent of infrastructure provision on the travelled link.

The strongest assumption of the Fajgelbaum and Schaal framework is the dependence of trade costs on $Q^n_{i,k}$, the total flow of goods on the link. Higher existing trade volumes on any given edge make sending an additional good more costly, a dynamic the authors refer to as *congestion*. A few things are worth noting about this assumption. First, congestion plays the role of an externality in the trade network. Sending one additional unit of goods from i to k makes all other existing shipments more expensive. The social planner realises this and takes congestion into account when determining optimal trade flows.[2] Second, congestion forces *are* allowed to vary with different goods. While analytically appealing, this assumption might strike the reader as odd at first. Congestion for transporting some good n is only caused by existing trade flows of the same good. No matter how many other goods are being pushed through an edge, each one only causes congestion nuisances for other shipments of the same variety. While Fajgelbaum and Schaal demonstrate a way to circumvent this peculiarity of the model, as will be discussed soon the assumption has clear analytical advantages and is hence kept for the purpose of this thesis.

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

$$C^n_i + \sum_{k \in N(i)} Q^n_{i,k}(1 + \tau^n_{i,k}(Q^n_{i,k}, I_{i,k})) \leq Y^n_i + \sum_{j \in N(i)} Q^n_{j,i} \tag{2}$$

must hold for every n and i. Fajgelbaum and Schaal call this the *Balanced Flows Constraint.*

[2] Perhaps not surprisingly, the existence of a negative externality will lead to over-provision of trade in the decentralised market solution to the problem. In their technical appendix, however, Fajgelbaum and Schaal demonstrate how a moderately straightforward transport tax can internalise the externality and establish the applicability of the two welfare theorems. As this thesis stays in the social-planner conception of the problem, such decentralisation conditions are of minor concern going forward.

2.3 Optimal Infrastructure Provision

So far, we have merely cast a plain model of trade in a geographical network, albeit with a slightly unconventional trade cost functional. Solving this model is more or less straightforward and would result in trade flows caused by heterogeneity in production endowments and population, but constrained by iceberg trade costs. The latter depend partially on the level of infrastructure present on any given node, which until now was taken as exogeneous. It is the main contribution of the Fajgelbaum and Schaal (2017) framework to endogeneise infrastructure provision $I_{i,k}$ in order to facilitate optimal trade flows. This allows for identifying areas in need of further infrastructure investment and computing their optimal level of expansion. Intuitively, a social planner would want to increase infrastructure on a link between locations which would heavily benefit from more mutual trade.

Analytically, this problem nests the static trade flow exercise mentioned above. The social planner chooses an infrastructure network, and given the network proceeds to compute the optimal trade flows subject to the *Balanced Flows Constraint* (2). In joint optimisation, it is the social planner's goal to construct a network in order to then induce optimal trade flows in the nested problem. A priori, the model leaves the possibility of differential infrastructure provision depending on the direction in which goods travel, that is $I_{i,k} \neq I_{k,i}$. However, in order to circumvent solutions in which some export-heavy nodes only have one-way streets leaving the location while consumption-heavy nodes cannot ever be left, I impose symmetry and restrict $I_{i,k} = I_{k,i} \; \forall \; i, k \in N(i)$

Without a constraint on infrastructure provision, the problem is trivially non-sensible: the social planner would just provide each edge with an infinite amount of infrastructure in order to induce free shipment of goods and perfect consumption smoothing over locations. To make the problem more interesting, I follow Fajgelbaum and Schaal in introducing a constraint on infrastructure. This is specified in fairly straightforward manner as the *Network Building Constraint*

$$\sum_{i} \sum_{k \in N(i)} \delta^{i}_{i,k} I_{i,k} \leq K \tag{3}$$

where $\delta^{i}_{i,k}$ denotes the cost of building infrastructure on the edge between nodes i and k. This cost is allowed to vary by link, such that one can model heterogeneity in infrastructure building cost caused by ruggedness, elevation, or the like. Total spending on infrastructure is, furthermore, constrained by K. One can think of K as the total budget allotted to infrastructure expenditure, or the total amount of concrete available in the economy. It is taken as exogeneous and does not compete with other endowments in the network.[3]

In the application of this thesis, K is interpreted as the total cost of originally spent for building the *existing* road network of a country. I observe the current road network of the economy, infer how much it must have cost to build it, and set K equal to this amount. The social planner's task of choosing an optimal network while treating the total amount of concrete in the economy as fixed, amounts to a reallocation exercise. The social planner gathers all the concrete available and gets to redistribute it in a more sensible way. In order to improve infrastructure between two nodes to foster local trade comes at the cost of having to take away infrastructure someplace else. I argue that how much she has to rearrange existing interconnections serves as a sensible measure of spatial efficiency in the existing network. Before we can fully derive interpret this

[3] One could think of an extension in which K has to be produced before being put on the road. The authors briefly discuss this possibility. To keep the model operational, I abstain from this extension and treat K as exogeneous and fixed.

measure, I first present the full planner's problem and ensuing equilibrium.

2.4 Planner's Problem and Equilibrium

In the nested problem, the social planner observes localities, endowments, population, and preferences and solves for trade flows between nodes that optimise overall welfare. She also solves for the optimal transport network which induces welfare-maximising trade flows in the nested problem while respecting the network building constraint (3). The full planner's problem can hence be stated as

$$\max_{\substack{\left\{C_i^n, \{Q_{i,k}^n\}_{k\in N(i)}\right\}_n, \\ c_i, \left\{I_{i,k}\right\}_{k\in N(i)}}} \quad \sum_i L_i u(c_i)$$

$$\begin{aligned}
\text{subject to} \quad & L_i c_i \leq \left(\sum_{n=1}^{N} (C_i^n)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} && \textbf{CES Consumption} \\
& C_i^n + \sum_{k\in N(i)} Q_{i,k}^n (1+\tau_{i,k}^n(Q_{i,k}^n, I_{i,k})) \leq Y_i^n + \sum_{j\in N(i)} Q_{j,i}^n && \textbf{Balanced Flows Constraint} \\
& \sum_i \sum_{k\in N(i)} \delta_{i,k}^i I_{i,k} \leq K && \textbf{Network Building Constraint} \\
& I_{i,k} = I_{k,i} \text{ for all } i \in \mathcal{I}, k \in N(i) && \textbf{Infrastructure Symmetry} \\
& C_i^n, c_i, I_{i,k}, Q_{i,k}^n \geq 0 \text{ for all } i \in \mathcal{I}, n \in N, k \in N(i). && \textbf{Non-Negativity Conditions}
\end{aligned}$$

My version of the planner's problem loosely follows the baseline Fajgelbaum and Schaal (2017) model. However, four important differences need to be emphasised. First, in my model all goods are tradeable and no local amenities exist. Second, I do not allow workers to migrate between places and hence differences in marginal utility might still exist between nodes. Third, my model remains agnostic about the production function of each location and no analysis of the optimal use of input factors is undertaken. Fourth, I impose infrastructure symmetry. All these changes are undertaken with the later calibration and reshuffling exercise in mind.

Solving the planner's problem appears potentially daunting for two reasons. First, it is ex-ante not clear whether a unique optimum exists. Second, since one of the control variables is infrastructure $I_{i,k}$, even if a unique optimum were to exist, it might appear as if one had to optimise over the very large space of possible networks, making the problem numerically intractable. Luckily, Fajgelbaum and Schaal (2017) provide conditions under which both these concerns can be allayed.

To show that how unique optimum can exist, first note that the introduction of congestion to iceberg trade cost causes the balanced flows constraint to be convex if $\beta > \gamma$. Every part of the lengthy constraint is linear, but for the interaction term $Q_{i,k}^n \tau_{i,k}^n(Q_{i,k}^n, I_{i,k})$ representing total trade costs. Since $\tau_{i,k}^n$ was parameterised as in (1), this expands to

$$Q_{i,k}^n \tau_{i,k}^n(Q_{i,k}^n, I_{i,k}) = \delta_{j,k}^{\tau} \frac{(Q_{i,k}^n)^{1+\beta}}{I_{i,k}^{\gamma}} \tag{5}$$

which is convex if $\beta > \gamma$.[4] This condition describes a notion of congestion dominance: increased infrastructure expenditure might alleviate the powers of congestion, but it can never overpower it. Intuitively, 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.

Second, instead of optimising over the large space of networks, Fajgelbaum and Schaal harness the convexity result to instead solve the dual of the problem specified above. Instead of solving for every single infrastructure link, trade flows for all goods, and consumption patterns in each location, we can recast the problem as a set of first-order conditions from the subproblems, which only depend on Lagrange multipliers of each constraint. There are considerably fewer multipliers than primal control variables, namely one for every good in every node (interpretable as prices). We are hence left to only find a price field from which under the convexity assumptions, all other properties follow. As Fajgelbaum and Schaal note, solving the dual is common practice in optimal transport literature as it makes cumbersome large-scale optimisation problems much more tractable.[5] I thus obtain the optimal network by constructing the Lagrange corresponding to the planner's problem, deriving its first-order conditions and recasting them as functions of the Lagrange parameters. These functions are largely equivalent to the ones derived in Fajgelbaum and Schaal's technical appendix, with the only differences attesting to the four changes discussed above. I can then reinsert these formulations into the original Lagrangian, which is now simply a function of the Lagrange parameters. Optimising numerically yields the solution to the dual problem. By inserting the parameters back into the derived first-order conditions, one can immediately derive the optimal infrastructure network $I_{i,k}$, optimal trade flows $Q^n_{i,k}$ over this network, and ensuing consumption patterns C^n_i in each location.

Having laid out the network topography and properties of the planner's problem and having ensured that a unique solution exists and is realistically obtainable, I now proceed to calibrate the model to African data in order to derive a novel dataset of network efficiency for the entire African continent.

3 Bringing the model to the data – Towards a spatial measure of African network inefficiency

Investigating the patterns behind the spatial distribution of network inefficiency involves a series of steps. First, I construct a representation of the topography of economic activity and infrastructure networks for every African country. Using the model by Fajgelbaum and Schaal (2017) outlined above, I then conduct an optimisation exercise in which existing infrastructure is reallocated within each country to maximise overall welfare. This scenario will produce winners and losers, such that I can derive a measure of network inefficiency over regions by comparing current welfare with welfare under the optimised scenario.

Below, I discuss these steps in more detail.

[4] This is Fajgelbaum and Schaal Proposition 1.

[5] Note, however, that is is still a quite demanding task to solve the ensuing dual problem, even numerically. Invoking duality reduces the scale of the problem, but we are still left with optimising over $I \times N$ variables.

3.1 Data Sources

To construct a network database of all African countries, I 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. Using GIS, I locate the geometric centroid of each cell and overlap these points with current political borders to assign countries to each centroid.[6] I then use spatial data on economic and geographic characteristics from a variety of sources and aggregate them onto the grid cell level.

Raster data on 2015 population totals comes from the Socioeconomic Data and Applications Center (2016) and is available on a much finer resolution than the one of my study. I hence aggregate them onto the grid cell level to obtain the total number of people living in each grid cell. On average, a cell is home to 110,000 people, with the median much to the left of that (25,000). The most populous cell contains Cairo and inhabits almost 18 million people. 212 cells are uninhabited.

Geographic characteristics include altitude, soil quality, weather data, malaria prevalence, and terrain ruggedness.[7] These data come from Henderson et al. (2018) and are also aggregated to the slightly coarser resolution of my study. Henderson et al. show that these indicators account for a substantial part of the global variation of economic activity and hence serve as an important set of controls for later empirical examinations.

Lastly, to proxy for heterogeneities in economic activity over space, I rely on the established practise of using satellite imagery of light intensity at night. Following the seminal work by Henderson et al. (2012), more lit places have been associated with more economic activity or growth in numerous settings, including ethnic homelands (?), cities (Storeygard, 2016), or the Korean peninsula (Lee, 2016).[8] I use data on night luminosity from Henderson et al. (2018) which was captured by satellites in 2010. This dataset bears an improvement over other existing night lights data, as it is able to better discern differences at the very right tail of the light distribution and hence prevents many problems related to top-coding of the highest-lit places. Taking means, this data is also aggregated to the 0.5 x 0.5 degree resolution. Night lights are approximately log-normal distributed and hence have to be converted into logarithmic form when serving as the dependent variable in standard regression settings. However, in the study at hand, lights serve as a calibration parameter for differences in economic output and thus do not have to be transformed.

3.2 Road Network

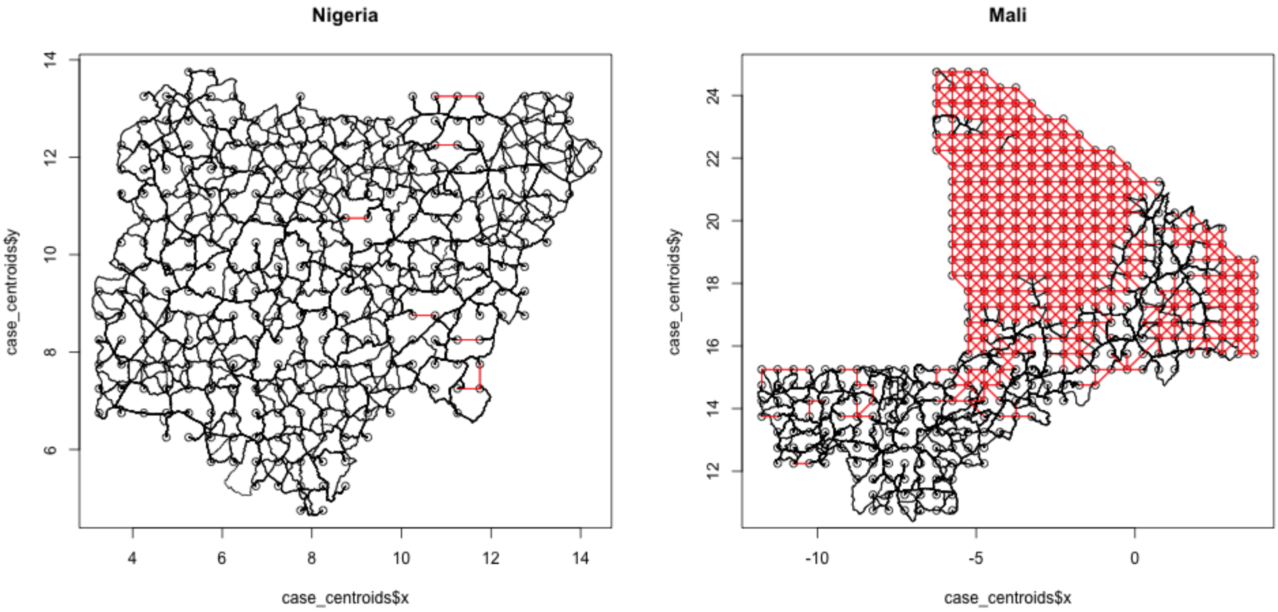
So far, the dataset merely consists of a list of grid cells with their respective characteristics. To gain a conception of their relative position in a trade network, a measure of connectedness between locations is needed. In particular, one needs to know whether two locations are connected at all, how strong the link between the two already is, how costly it is to improve the link, and how costly it is to trade between the

[6] Data on African borders come from Sandvik (2008). Because of its relative age this dataset does not include borders on South Sudan. I hence add the world's youngest country manually using data from OCHA (2017). Taking the centroid position as the decisive statistic to assign countries might lead to situations where cells near oddly shaped borders are assigned country A, even if they mostly lie in country B. However, I don't believe this to be a major concern, especially since colonial legacies have led to African borders being drawn in a particularly straight-line fashion (see Alesina et al., 2011).

[7] Note that this version of the indicator slightly improves upon the original quantification of terrain ruggedness in the work by Nunn and Puga (2012).

[8] See Donaldson and Storeygard (2016) for an excellent review.

Figure 1: Road Networks for different countries as scraped off OSM



(a) Nigeria **(b)** Mali

two. Or, in the notation of Fajgelbaum and Schaal (2017), an infrastructure matrix $I_{i,k}$, an infrastructure investment cost parameter $\delta^{I}_{i,k}$, and a trade cost parameter $\delta^{\tau}_{i,k}$.

3.2.1 Road Data

Obtaining objective data on transportation networks is difficult, especially in developing countries where many transport routes are not available in digitised form. In their own empirical exercise Fajgelbaum and Schaal observe European countries and make use of a large coherent dataset on position and objective characteristics of important European roads. A comparable dataset for Africa does not yet publicly exist, even though a recent project by Jedwab and Storeygard (2017) has undertaken the effort to manually compile and digitise Michelin Maps in order to create a comparable dataset (their data is not yet available for replication).

Instead, I use a different approach and make use of the open source internet routing service Open Street Maps (OSM), which is comparable to Google Maps but allows for unlimited use of its API. For every centroid location, I scrape OSM for the optimal route to their respective eight surrounding neighbours. This is greatly facilitated by the R package `osrmRoute`.[9] Since I am interested solely in within-country transport networks, I perform the exercise for each country separately and do not elicit connections between locations of different countries. Hence, centroids located near a coast or country border often have less than eight immediate neighbours. For all of the resulting almost 90,000 routes, I gather distance travelled, average speed, and step-by-step coordinates of the travel path.

The OSM routing algorithm is specified for cars and takes into account differential speeds attainable on different types of roads. However, if either start or destination location do not directly fall onto a street,

[9]Calculations were performed in November 2017.

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. For some particularly remote areas, the nearest street is very far such that the car routing provided by OSM is not sensible. To counter these cases, I also calculate for all 90,000 connections the outside option of walking the entire link in a straight line at 4 km/h. I then identify cases in which walking directly is actually faster than using OSM's proposed route (plus the travel to and from roads). In these cases, I replace OSM's route with the walking distance and (constant 4 km/h) speed. Figure (1) presents the resulting road networks for two African countries. Figure (1a) displays every optimal route for Nigeria, which appears overall fairly well connected. Commuters mostly seem to be able to drive relatively direct routes between locations, even though cases with substantial detours are also evident at second glance. Connections in which walking were the preferred alternative are displayed in red and fairly rare in Nigeria. Figure (1b) presents the case of Mali, which paints a different picture: for many connections through the Sahara desert in the North-East of the country, walking straight lines in the sand is actually the fastest way to get from A to B.

Relying on the open source community of OSM does come with some drawbacks. Most importantly, data on the position and quality of roads is user-generated and hence subject to reporting bias. Intuitively, 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. As soon as inference is conducted on the relationship between streets and any covariate of development, the resulting estimates will be biased. 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, especially compared to the alternative of digitised Michelin maps. Reporting bias will definitely be an issue when trying to find the optimal route *within* a particular small neighbourhood in Accra, but the OSM database should do a fairly good job in finding the optimal route between Accra and Kumasi. It is nevertheless important to keep this potential flaw of the data in mind when conducting inference later on.[10]

After having collected data on distance and average speed on the optimal route between all neighbouring centroid locations, the next step is to discretise these data in order to have a tractable network representation capable of performing the trade simulations necessary in the remainder of this study.

3.2.2 Infrastructure matrix $I_{i,k}$

To gain a conception of how much two given nodes are connected in the transport network, one needs to derive a numeric measure of how much infrastructure there is on the optimal route between locations. In

[10]There is a second, less troubling problem with using OSM data. As this study merely pertains to within-country transport networks, I only look at connections between neighbouring locations of the same country. However, in some cases the optimal route between these locations still might go through a neighbouring country. For instance, Senegal is effectively split into two parts by the intersecting country The Gambia. Still, when connecting Senegalese cell centroids just to the north and to the south of The Gambia (which are still less than 60km apart), the route will go over foreign soil. This presents a problem only in later policy recommendations, as political leaders of one country cannot necessarily legislate road improvements abroad. The rest of the analysis is not affected by this issue.

their own empirical analysis, Fajgelbaum and Schaal (2017) have data on the average number on lanes of the streets used on a given route, and whether these streets are national or secondary roads. The OSM algorithm does not supply this detailed level of information for Africa. However, I argue that the authors are only proxying for a much more immediate statistic – the average speed with which one can travel on a given road. Obviously there are many factors influencing driving speed other than number of lanes or road classifications. Congestion, altitude differences, or potholes come to mind. However, what the authors are trying to capture is an infrastructure investment vehicle to reduce trade costs. Certainly building more lanes on a given road will reduce trade costs. But it will do so by increasing the speed with which cars can travel on that road. I argue that observing the average speed with which transport can occur between two nodes is a much more immediate measure of how well these nodes are connected. I hence propose

$$I_{i,k} = \text{Average Speed}_{i,k} \tag{6}$$

This measure is naturally bound from below at 4 km/h, as walking the air-line distance is always available as a backup. Empirically, average speeds range between 6 km/h (Mauritania, where most of the distances through the desert have to be covered by walking) and 33 km/h (Swaziland). It is now the objective of the policymaker or social planner to reduce trade costs between suitable trade partners by increasing the average speed $I_{i,k}$ with which transport can occur between them.

3.2.3 Infrastructure building cost matrix $\delta^I_{i,k}$

The constraint of the policymaker in this exercise is that increasing $I_{i,k}$ comes at a cost. In particular, building an additional $I_{i,k}$ between locations i and k costs $\delta^I_{i,k}$. The total infrastructure budget in the economy is fixed at K, such that

$$\sum_i \sum_{k \in N(i)} \delta^I_{i,k} I_{i,k} = K \tag{7}$$

where $N(i)$ denotes the set of surrounding neighbours of node i. $\delta^I_{i,k}$ depends on a variety of inputs, like the distance of a road or the underlying terrain. I follow Fajgelbaum and Schaal who in turn make use of a recent study by Collier et al. (2015) who estimate infrastructure building costs in the Developing World. Following Fajgelbaum and Schaal, I calculate

$$\ln\Big(\frac{\delta^I_{i,k}}{\text{dist}_{i,k}}\Big) = \ln(\delta^I_0) - 0.11 * (\text{dist}_{i,k} > 50km) + 0.12 * \ln(\text{ruggedness}_{i,k}) \tag{8}$$

Note that every route in my sample is longer than 50 kilometres and the corresponding dummy term is hence always 1. δ^I_0 is a scaling parameter. Following Fajgelbaum and Schaal, I normalise $K = 1$ for every country and thus flexibly alter δ^I_0 such that equation (7) is satisfied.

3.2.4 Trade cost parameter $\delta^\tau_{i,k}$

Trade costs between locations are modelled in the standard iceberg formulation, whereby in order for one unit to arrive in destination k, precisely $(1 + \tau_{i,k})$ units have to leave origin i. Perhaps surprisingly, even though trade costs are such an influential concept in the trade literature very few reliable estimates of their magnitude exist. Fajgelbaum and Schaal propose a tractable functional form in which $\tau_{i,k}(Q_{i,k}, I_{i,k})$ depends negatively on the invested infrastructure $I_{i,k}$ (derived above) and positively on goods shipped $Q_{i,k}$ on a given link (a notion they refer to as *congestion*):

$$\tau_{i,k}(Q_{i,k}, I_{i,k}) = \delta^{\tau}_{j,k} \frac{Q^{\beta}_{i,k}}{I^{\gamma}_{i,k}} \tag{9}$$

with $\beta = 1.245$ and $\gamma = 0.5\beta = 0.6225$ ensuring convexity. $\delta^{\tau}_{j,k}$ is a scaling parameter. Fajgelbaum and Schaal calibrate is as a linear function of the distance between i and k and in order to match some particular patterns from Spanish trade data. This calibration is not applicable to the given context. Firstly, it deals with European developed countries which recent evidence suggests face much less stifling trade costs than the African continent (see e.g. Anderson and Van Wincoop, 2004). Secondly, their calibration revolves around different measures for economic output Y_i and infrastructure investment $I_{i,k}$ and will hence produce arbitrary estimates for $\delta^{\tau}_{j,k}$.

Instead, I make use of the recent contribution by Atkin and Donaldson (2015). They use barcode-level data on sale prices of identical goods to back out trade costs between regions within two African countries, Nigeria and Ethiopia. They show that trade costs are significantly increasing in distance between origin and destination and estimate the trade cost elasticity to (log) distance as 0.0374 for Ethiopia and 0.0558 for Nigeria. Directly using the average of these two point estimates, I calculate

$$\delta^{\tau}_{i,k} = 0.466 * log(\text{dist}_{i,k}) \tag{10}$$

This specification treats $\text{dist}_{i,k}$ as fixed. It hence precludes the policy maker from building a completely new road between locations that were until now only connected via an extensive detour. The only possibility on her hand is to make travel along a given route faster, not to design a completely new route. This will prevent the model to conceive new roads that go right through a swamp or big mountain, but instead propose an improvement of the existing road that goes around the mountain.

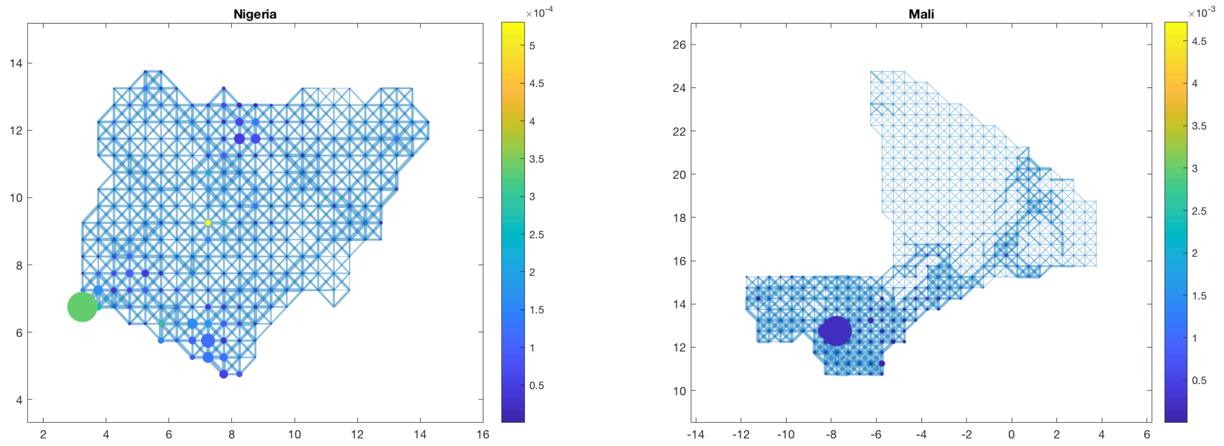
3.3 Heterogeneous goods

To build incentives for trade, I introduce two different goods: an agricultural and an urban good, which are aggregated in canonical CES fashion. Since economic output is proxied by night luminosity, I cannot observe the distribution of different goods in a given grid cell, but only their total production. I hence assume that urban grid cells are solely producing the urban good, while all other grid cells are producing nothing but the agricultural good. This largely follows Fajgelbaum and Schaal, except that they also allow for differentiation even amongst the largest producing cities. For ease of interpretation (and computation), I stay with a two-good economy.

To classify grid cells as urban or rural, I use an iterative procedure which seeks to match each country's 2016 urbanisation rate as reported by The World Bank (2017).[11] With this procedure, 7 per cent of grid

[11] I start by assuming every location is a city and then gradually proceed to re-classify the least densely populated locations, until the ratio of people living in urban areas to total population equals that of the WDI. For three countries, the WDI do not report urbanisation rates. In these cases, I match the overall urbanisation rate for the entire African continent of 42 per cent as reported by Lall et al. (2017).

Figure 2: Discretised Networks for different countries



(a) Nigeria

(b) Mali

cells are classified as urban. These cells inhabit 40 per cent of the continent's population, matching recent figures from Lall et al. (2017) fairly well.

Since labor is fixed, I do not have to impose a strict production function. In fact, the model stays agnostic about how grid cells produce their luminosity output. This is in contrast to Fajgelbaum and Schaal who do allow for endogenous labor allocation across cities and goods and hence need to take a stance on functional forms of total factor productivity and output. The only assumption my model makes is that cells can ever only produce one of the two varieties and hence their productivity in the respective other variety is zero.[12]

3.4 Simulation

After these steps, a discretised network representation now exists for every African country. Nodes in the network are the spaced centroid locations of each grid cell. They combine the characteristics of the entire grid cell (population, output, etc.) in one point. Edges in the network are road connections between centroids. Each edge carries a number of characteristics (average speed, trade costs, and infrastructure building costs). Figure (2) presents this discretised network specification for the two countries from above. Nodes are printed larger proportional to their population. A node's colour scale reflects night light intensity per capita (a statistic which is only calculated for illustrative purposes). Edges are drawn thicker proportional to the initial infrastructure investment (i.e. average attainable speed).

For each country, I proceed to conduct two simulation exercises. In the first, infrastructure $I_{i,k}$ is treated as fixed. This is to obtain a baseline estimate of the spatial variation of welfare in each country. I conduct a social planner exercise whose goal it is to choose trade flows in order to maximise total welfare. This fairly standard static exercise generates the need for trade mostly out of product differentiation. Cities do not produce the agricultural good and hence need to import from surrounding rural backlands, and vice-versa. To facilitate computation and make the problem tractable, Fajgelbaum and Schaal show that instead of

[12]Thanks to this parsimonious specification, I do not rule out the impact of capital or any inputs other than labour to the production function. I also avoid having to divide total lights by total population in order to obtain TFP, a procedure recent literature has explicitly warned against (see e.g. Michalopoulos and Papaioannou, 2018).

optimising over the space of trade flows, the social planner can indeed optimise over the much sparser set of prices. Since strong convexity holds, the solution is unique and a certain price field will induce trade flows and consumption patterns that maximise total welfare. The authors also show how the two welfare theorems hold in this setting and the social planner outcome can hence be supported by a decentralised market solution.

As labor is fixed, the resulting solution will have two properties. Firstly, total output over the entire country will remain untouched. Labour inputs do not shift to more productive regions. Indeed, any welfare gains will be attained solely by shipping the right quantity and mix of goods to the right regions. Second, labor immobility will leave welfare differences between regions as agents cannot simply move to privileged cells. The social planner would like to overcome these differences, but is confronted by trade costs which might leave certain remote areas much worse off than well-connected ones.

Following this static exercise, I proceed to the main task of endogeneising the infrastructure matrix $I_{i,k}$. Intuitively, the social planner is now free to improve connections to a remote area in order to supply them better. Fajgelbaum and Schaal's main contribution is showing that under particular assumptions, this task can be accomplished simply by means of an additional constraint. Improving on infrastructure yields the benefit of more utility smoothing over space, yet comes at cost $\delta^I_{i,k}$. Returns to improving $I_{i,k}$ are not unbounded, however, as the congestion assumption ensures convexity and hence an interior solution to trade flows exists.

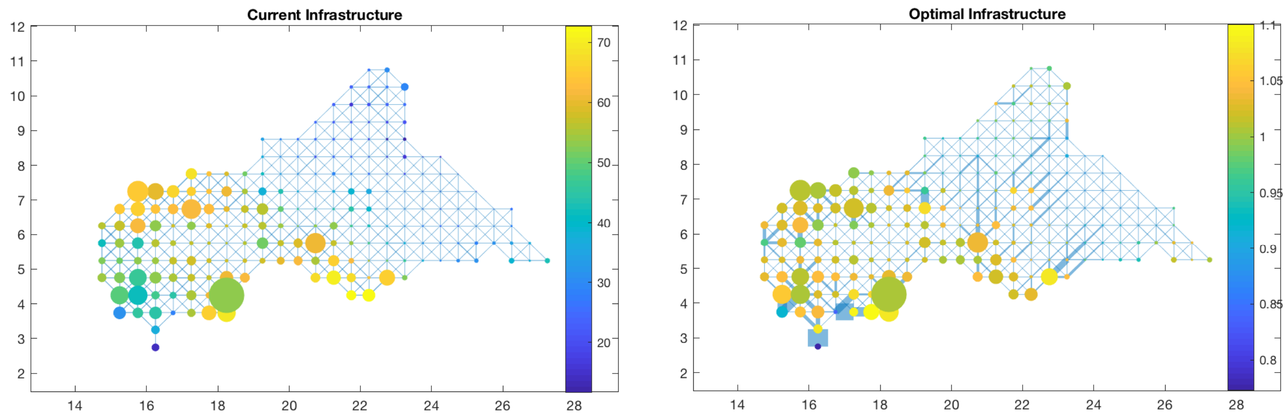
To gain a notion of network inefficiency, I run the following scenario: while the social planner is free to manipulate any given link between neighbouring locations, take away or add to existing infrastructure, the total amount of infrastructure in the economy remains fixed. Formally, this corresponds to $K = 1$ in equation (7) and is what Fajgelbaum and Schaal call the *Optimal Reallocation Scenario.* If the social planner wants to improve the connection between two given locations, she will have to take away infrastructure from somewhere else in the country. The entire exercise does not seek to identify where to place the optimal new investment, but rather represents an utterly fictitious scenario in which every road can be lifted from the ground, reshuffled, and eventually located someplace else.[13] This procedure does not measure how many roads a country has, but rather how well they are placed. It does not look at whether the entire country is full of speedy roads, but rather whether those roads connect the right locations.

Figure (3) visualises this reallocation exercise for the Central African Republic. Subfigure (3a) displays the discretised network representation of the country, comparable to figures (2a) and (2b). The edges to this network are printed almost evenly thick, implying that infrastructure is fairly evenly distributed across the country. Subfigure (3b) then displays the country after the network reshuffling exercise. Two patterns stand out. First, the social planner sees a heavy need to connect the populous areas in the south west of the country to with more speedy roads. For that, she is willing to salvage some of the unnecessary infrastructure in the middle or north of the country. Second, there still seems to be a benefit from having a few trails connecting the south west with the north east of the country. Some clear north-south and east-west highways spanning multiple regions emerge.

I conduct the reallocation scenario for every African country. Five small countries (Cape Verde, Comoros, The Gambia, Mauritius, and Reunion) are too small to form a sensible network as they only show up as

[13] Note that equation (7) 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.

Figure 3: Optimal Reallocation Scenario in the Central African Republic



(a) Before Reallocation

(b) After Reallocation

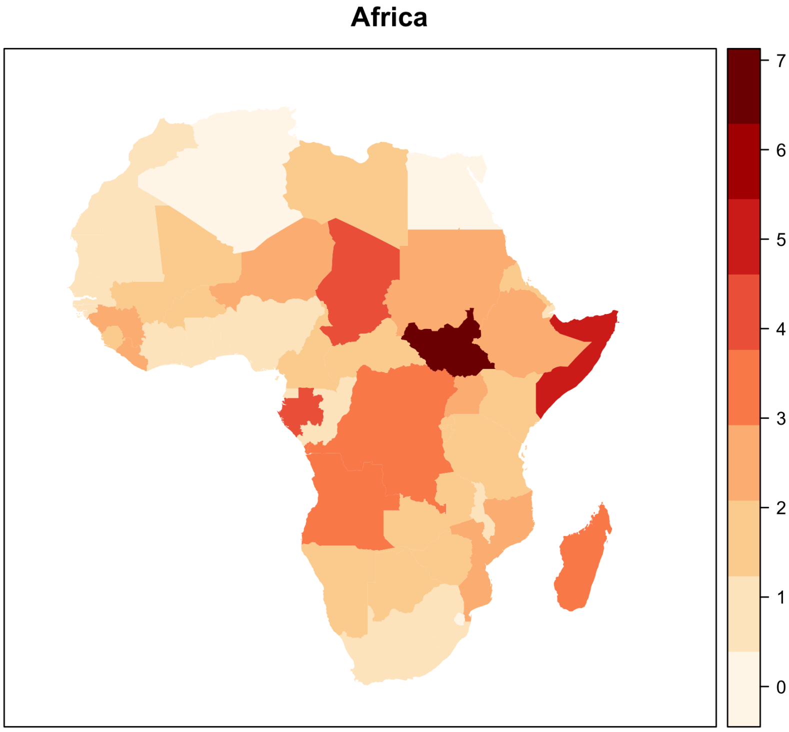
a single location in the dataset and are henceforth no longer considered. Computation times are greatly diminished when exploiting the strong convexity of the optimisation setting and solving the dual problem as outlined in the appendix of Fajgelbaum and Schaal (2017). Optimisations are performed via Matlab's `fmincon` command. When conducting the simulations, I bound the social planner's set of permissible roads from below, at 4 km/h (such that $I_{i,k} \geq 4 \; \forall i, k$). This is motivated by the assumption at the beginning that walking straight lines at this speed is conceived as an outside option and always available to any commuter. The social planner should not be able to force commuters to travel slower than walking, in order to build a faster road elsewhere.[14]

After successfully reshuffling a country's transport network, overall welfare will necessarily (weakly) increase. It is the social planner's objective to maximise overall welfare, and since the original network composition is always still available, the entire country cannot on aggregate be worse off than before. Note as before that due to labor immobility, overall production (light output) will be the same as before. Welfare gains are solely due to enabling mutual benefits from trade by connecting the right locations. The Central African Republic of figure (3), for instance, stands to gain 1.84% of overall welfare just by reshuffling roads. Since this welfare gain is rather hard to materialise, I prefer thinking of it as a measure of network inefficiency. The higher the hypothetical gains from merely reshuffling existing roads, the more haphazard the existing allocation of roads, and hence the more inefficient the current transport network.

Figure (4) displays all African countries and their measure of network inefficiency. The closer the forgone welfare gain to zero (the lighter the country's colour), the more efficient the current allocation of roads. The Central African Republic's measure of 1.84% makes the country look rather good in comparison. Some (mostly more developed) countries like South Africa (0.47%) or Tunisia (0.24%) perform even better. Many countries are leaving much more on the table like Somalia (4.76%) or Chad (4.28%). No African country, however, has a more ill advised road network than South Sudan (6.66%). This might not come as a surprise,

[14]Contrarily, I do not restrict possible investments from above (or at least not in addition to the restriction imposed by equation (7)), as this could violate the strong convexity of the problem. Not bounding the problem in principle allows the social planner to build supersonic speed highways. However, the model is calibrated in a way that makes this very unattractive to the planner anyway. After simulating reallocation in every African country, less than 0.8% of all built roads were suggested to be over 260 km/h. Still, one outlier of 2007 km/h (in Egypt) and one of 1755 km/h (in South Africa) remain.

Figure 4: African countries by network inefficiency



as the world's youngest country has largely inherited a road network that was not conceived to sustain an independent nation, but connect it to its former capital up north. On a simple cross-country level, more inefficient networks are significantly correlated with more corruption ($p < 0.01$), less property rights ($p < 0.01$) and less 2010 log GDP ($p = 0.07$). Note that these are merely descriptive correlations which are far from implying any form of causation.[15]

While each country only stands to gain overall welfare from this reallocation procedure, individual 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 can help identify regions which are currently over or under-provided for. More formally, I define

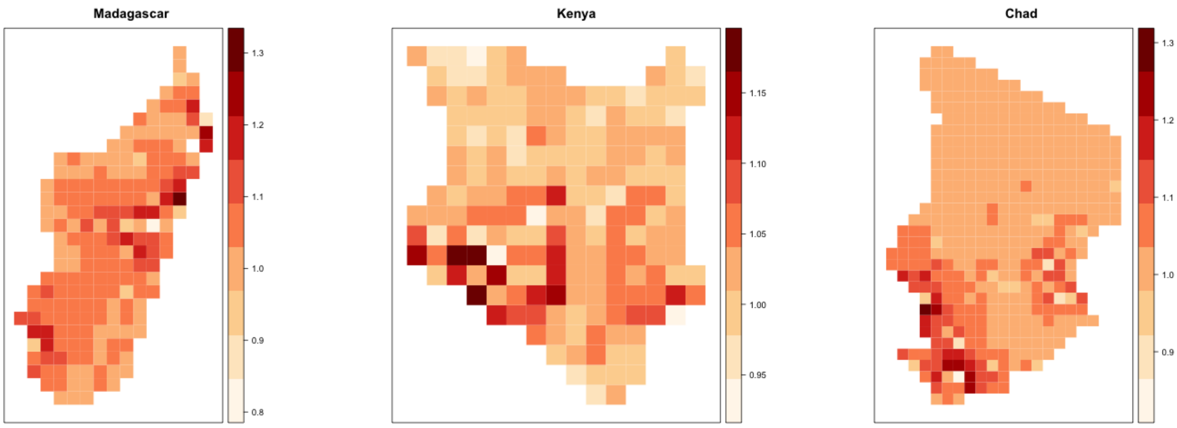
$$\zeta_i = \frac{\text{Welfare under the optimal Infrastructure}_i}{\text{Welfare under the current Infrastructure}_i} \tag{11}$$

as the *Infrastructure Gap* for each grid cell i. Figure (5) displays the spatial distribution of the infrastructure gap ζ_i for Madagascar, Kenya, and Chad. The darker a grid cell's shade, the more it is disadvantaged by the inefficiencies of the current network. The lighter a cell, the more overprovided a region is with infrastructure.

Figure (6), lastly, displays the spatial variation of ζ_i over all 10,000+ grid cells of the entire African continent. When interpreting this map, note that grid cells are undergoing the reshuffling scenario solely within their

[15] Data from The World Bank (2017). For corruption and property rights, data is only available for 35 countries and correlations are hence performed on this truncated sample.

Figure 5: Spatial Distribution of ζ_i for sample countries



(a) Madagascar **(b)** Kenya **(c)** Chad

Figure 6: Spatial Distribution of ζ_i for entire sample

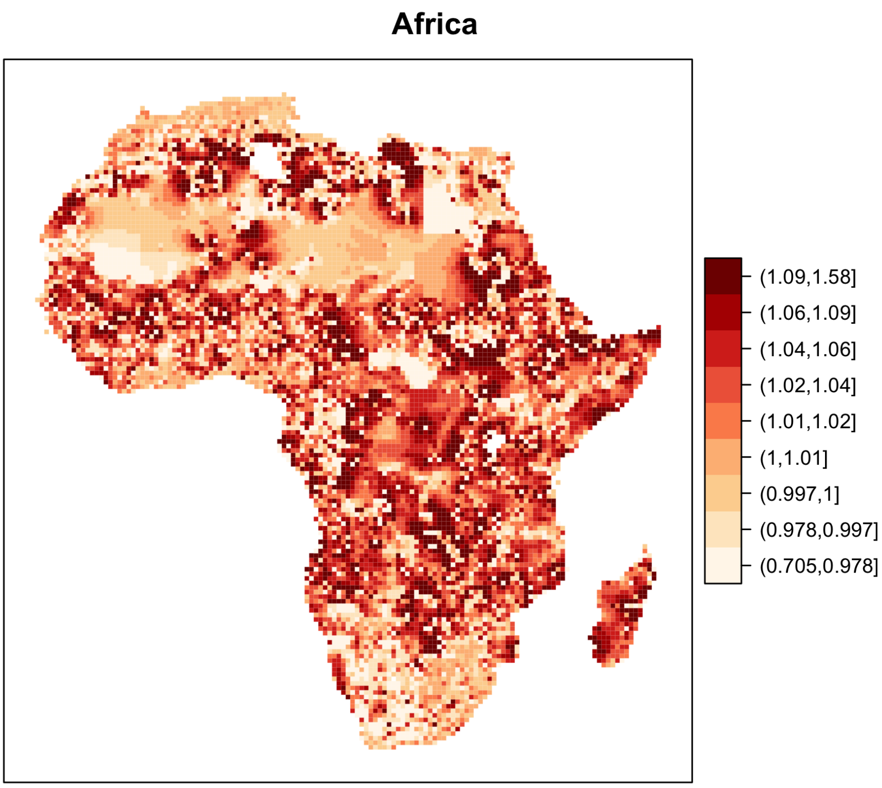


Table 1: Colonial Railroads and Infrastructure Gap

	Dependent variable:							
	Infrastructure Gap ζ_i							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KM of Colonial Railroads	-0.0002^{***}	-0.0001^{***}	-0.0002^{***}	-0.0002^{***}				
	(0.0001)	(0.0001)	(0.0001)	(0.0001)				
KM of Colonial Placebo Railroads					0.00004	-0.0002	-0.0003	-0.0003
					(0.0003)	(0.0003)	(0.0003)	(0.0003)
Country FE		Yes	Yes	Yes		Yes	Yes	Yes
Geographic controls			Yes	Yes			Yes	Yes
Simulation controls				Yes				Yes
Observations	10,158	10,158	10,158	10,158	10,158	10,158	10,158	10,158
R^2	0.001	0.099	0.112	0.114	0.00000	0.098	0.111	0.113
Note:								*p<0.1; **p<0.05; ***p<0.01

This table displays results of estimation of equation (??) on the sample of 0.5x0.5 degree grid cells for the entire African continent (excluding five small countries, see text). Dependent variable is the infrastructure gap for each grid cell. Columns (1)-(4) estimate the effect of colonial infrastructure investments on today's infrastructure gap. Starting with a simple univariate cross-section in (1), column (2) adds 49 country-fixed effects. Column (3) adds geographic controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, 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. Simulation controls are added in column (4) and are comprised of population, night lights, ruggedness, and a dummy for whether a cell is classified as urban. These are indicators that went into the original infrastructure re-allocation simulation and are hence not orthogonal to ζ. Columns (5)-(8) repeat these calculations with railroads that were planned, but never built ("placebo railroads"). Results are robust to using only the subsample of 33 countries with any colonial infrastructure investment as reported by Jedwab and Moradi (2016), plus South Africa. Heteroskedasticity-robust standard errors are clustered on the 3x3 degree level and are shown in parantheses.

respective country. 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 infrastructure gap across the African continent. The luckiest region (in Namibia) stands to loose almost 30% of total welfare if the fictitious social planner intervened and reshuffled roads away from. On the other hand of the spectrum, the residents of one grid cell in Gabon are missing out on a welfare hike of more than 50%. Moreover, abandoned regions are clearly displaying spatial correlation with large neighbouring swaths of land collectively missing out on infrastructure improvements in certain countries. This begs the conclusion that countries do not just overlook single grid cells but rather live with vast stretches of disadvantaged regions.

In the following section, I proceed to investigate the patterns behind this heterogeneity of network inefficiency over space.

4 Results

Table 2: General Equilibrium Effects of Colonial Railroads

	Dependent variable:							
	Infrastructure Gap ζ_i							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
< 10 KM to Colonial Railroad	−0.013***	−0.015***	−0.017***				−0.014***	
	(0.003)	(0.004)	(0.004)				(0.004)	
10 − 20 KM to Colonial Railroad	−0.013***	−0.015***	−0.017***				−0.015***	
	(0.005)	(0.005)	(0.005)				(0.005)	
20 − 30 KM to Colonial Railroad	−0.002	−0.004	−0.005				−0.005	
	(0.004)	(0.004)	(0.004)				(0.004)	
30 − 40 KM to Colonial Railroad	0.010**	0.008*	0.007				0.008	
	(0.005)	(0.005)	(0.005)				(0.005)	
< 10 KM to Colonial Placebo Railroad				−0.005	−0.005	−0.006		−0.007*
				(0.004)	(0.004)	(0.004)		(0.004)
10 − 20 KM to Colonial Placebo Railroad				−0.003	−0.004	−0.004		−0.005
				(0.005)	(0.005)	(0.005)		(0.005)
20 − 30 KM to Colonial Placebo Railroad				−0.001	−0.001	−0.001		−0.004
				(0.004)	(0.004)	(0.004)		(0.004)
30 − 40 KM to Colonial Placebo Railroad				0.007	0.006	0.005		0.003
				(0.004)	(0.004)	(0.004)		(0.004)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls		Yes	Yes		Yes	Yes	Yes	Yes
Simulation controls			Yes			Yes	Yes	Yes
Observations	10,158	10,158	10,158	10,158	10,158	10,158	6,362	6,362
R^2	0.101	0.115	0.118	0.099	0.111	0.114	0.116	0.110

Note: *p<0.1; **p<0.05; ***p<0.01

This table displays effects of various distance-intervals on the infrastructure gap ζ. Explanatory covariates are dummy-variables indicating whether a cell's centroid is within X kilometres to its closest colonial railroad. Geographic controls consist of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, 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. Simulation controls are comprised of population, night lights, ruggedness, and a dummy for whether a cell is classified as urban. Columns (1)-(3) examine the effect of actually built colonial railroads. Columns (4)-(6) repeat these calculations with railroads that were planned, but never built ("placebo railroads"). Columns (7)-(8) restrict the sample to the 32 countries on which data for colonial railways is available. Heteroskedasticity-robust standard errors are clustered on the 3x3 degree level and are shown in parantheses.

Table 3: Heterogeneous Effects of Colonial Railroads

	Dependent variable:					
	Infrastructure Gap ζ_i					
	(1)	(2)	(3)	(4)	(5)	(6)
KM of Colonial Rails for Military Purposes	-0.0002^{***}	-0.0002^{***}			-0.0002^{***}	-0.0002^{***}
	(0.0001)	(0.0001)			(0.0001)	(0.0001)
KM of Colonial Rails for Mining Purposes			-0.0001	-0.0001	-0.0001	-0.0001
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Simulation controls		Yes		Yes		Yes
Observations	10,158	10,158	10,158	10,158	10,158	10,158
R^2	0.112	0.114	0.111	0.113	0.112	0.114
Note:						*p<0.1; **p<0.05; ***p<0.01

This table replicated the estimations of Table (1) *in estimating effects of colonial railroads on the infrastructure gap* ζ. *Colonial rails are classified as built for military or mining purposes (or neither or both) by Jedwab and Moradi (2016). Geographic controls consist of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, 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. Simulation controls are comprised of population, night lights, ruggedness, and a dummy for whether a cell is classified as urban. Results are mostly robust for using only the sub-sample of 32 countries for which this data is available, however the p-value of negative impact of military rails increases to* $p = 0.013$ *(Country FE and Geographic Controls) and* $p = 0.096$ *(Country FE, Geographic, and Simulation Controls). Heteroskedasticity-robust standard errors are clustered on the 3x3 degree level and are shown in parantheses.*

References

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

Allen, Treb and Costas Arkolakis (**2016**). The Welfare Effects of Transportation Infrastructure Improvements. Working Paper, Dartmouth and Yale

Anderson, James E. and Eric Van Wincoop (**2004**). Trade Costs. *Journal of Economic Literature* 42(3), pp. 691–751

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

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

Donaldson, Dave and Adam Storeygard (**2016**). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives* 30(4), pp. 171–198

Fajgelbaum, Pablo D. and Edouard Schaal (**2017**). Optimal Transport Networks in Spatial Equilibrium. Working Paper, National Bureau of Economic Research

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

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)

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

Jedwab, Remi and Alexander Moradi (**2016**). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *Review of Economics and Statistics* 98(2), pp. 268–284

Jedwab, Remi and Adam Storeygard (**2017**). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010. Working Paper

Lall, Somik Vinay, J. Vernon Henderson, and Anthony J. Venables (**2017**). *Africa's Cities: Opening Doors to the World*. World Bank Publications

Lee, Yong Suk (**2016**). International Isolation and Regional Inequality: Evidence from Sanctions on North Korea. Working Paper, Stanford University

Michalopoulos, Stelios and Elias Papaioannou (**2018**). Spatial Patterns of Development: A Meso Approach. *Annual Review of Economics* forthcoming

Nunn, Nathan and Diego Puga (**2012**). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics* 94(1), pp. 20–36

OCHA, United Nations Office for the Coordination of Humanitarian Affairs (**2017**). South Sudan - Administrative Boundaries - Humanitarian Data Exchange. `https://data.humdata.org/dataset/south-sudan-administrative-boundaries` (Accessed: 2017-12-17)

Sandvik, Bjorn (**2008**). World Borders Dataset. `http://thematicmapping.org/downloads/world_borders.php` (Accessed: 2018-01-04)

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

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

The World Bank (**2017**). World Development Indicators. `https://data.worldbank.org/data-catalog/world-development-indicators` (Accessed: 2017-12-12)