Within-City Roads and Urban Growth*

P. Brandily F. Rauch

September, 2020

Abstract

In this paper we study the role of within-city roads layout in fostering city growth. Within-city roads networks have not been studied extensively in economics although they are essential to facilitate interactions, at the core of agglomeration economies. We build and compute several simple measures of roads network and construct a sample of over 1800 cities and towns from Sub-Saharan Africa. Using a simple econometric model and two instrumental variable strategies based on the history of African cities, we then estimate the causal impact of within-city roads layout on urban growth. We find that over the recent decades, cities with greater road density and road evenness in the centre grew faster.

JEL-classification: O18, R4 Key words: Road layout, Urban planning, Urbanization, Sub-Saharan Africa

^{*}Brandily: Paris School of Economics, Rauch: Oxford University. Contact email address: ferdinand.rauch@economics.ox.ac.uk. We thank Julia Bird, Paul Collier, Victor Couture, Vernon Henderson, Richard Hornbeck, Guy Michaels, Matt Turner, Tony Venables and various seminar participants for helpful comments. We gratefully acknowledge the generous support of a Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. The usual disclaimer applies.

1 Introduction

It is an established fact in economics that cities enhance productivity and consumption benefits through various mechanisms. Within this tradition, all mechanisms proposed for agglomeration effects relate to one essential feature of the urban environment: it facilitates short-distance interactions between economic agents. This feature is at the core of the main theories, regardless of whether they focus on production or consumption. In his seminal typology Marshall (1920) suggested knowledge spillovers, labour market externalities and production linkages as drivers of urban productivity; more recently, Duranton and Puga (2004) highlight the importance of sharing, matching and learning. In this paper we investigate the role of roads and streets network within cities. We thus focus on a central component of the urban setting that could directly promote or discourage interactions between agents. In particular, we study the impact of the density and shape of the road grid within African cities on population growth in recent decades.

To study these specific urban features we measure various statistics on the road networks within urban centres for a large number of African cities and show the association between some of the layout's features and urban growth. Combining satellite imagery and Open Street Map, an open-source geographic data set mapping roads and streets all over the world, we build simple statistics of the road layouts at a precise geographic scale. We believe our focus on within-city layouts to be a conceptual innovation to the economic literature where the topic has remained largely unstudied so far. We show that measures of the central road grid, in particular its density and evenness, are associated with recent population growth.

We focus on the link between these within-cities roads layout and urbanization specifically in the context of sub-Saharan Africa. We chose this focus for four main reasons. First, understanding better the role of this urban infrastructure in the capacity of cities to generate growth could help designing policies that benefit the billions of people that live or are expected to live in African cities in the next decades. All the more so in the context of cities and towns that rely heavily on road transportation, and currently do not have many urban train or underground networks. Second, it is costly to change a road layout once it is in place. Planning ahead is therefore critical in Africa's urbanization, which is characterized by low income levels and where "two-thirds of the cities are to be built" (Bryan, Glaeser and Tsivanidis 2019). Third, various studies report high level of traffic congestion in African cities and the need for better road investments. Hidden external costs of congestion were estimated at up to 5% of cities' GDPs in Dakar or Abidjan, which is more than twice the estimates for European cities (Cervero, 2013) while "in a sample of 30 cities around the world, the 8 African cities rank in bottom 12 spots for road density" (Lall, Henderson and Venables 2017). Finally, some argue in favor of a "grid structure, which (...) enhanced travel efficiency" (as stated in the last major report from the World Bank on Africa's Cities - idem), although to the best of our knowledge there is little empirical research in economics so far to assess this hypothesis.

To estimate the impact of road layout we rely on two observations: the road layout is both difficult to change and dependent of the context and available transportation technologies at the time of its construction. The persistence of road layout over time is exemplified in cities of the ancient world (like Jerusalem or Paris, among many others, where the Roman's cardo are still important streets today) or the new (as New York where the "Commissioners' Plan" of 1811 still structures Manhattan today). More recently, this was also one conclusion from a large scale analysis of cities around the world by Barrington-Leigh and Millard-Ball (2019). In the case of Africa in particular, cities are often much newer and Baruah et al (2017b) provide anecdotal evidence of this layout persistence. These observations imply that due to historic reasons some cities may be accidentally stuck with better or worse road grids, which would benefit or hurt their growth potential. If a city by historic accident develops

a good road layout it will be rewarded by long lasting population growth, as people move in to benefit from the agglomeration benefits generated in the city. If on the other hand a city is stuck with a sub-optimal grid, it may reach a peak population level beyond which it finds it difficult to grow any further.

African cities also prove a useful environment to empirically study this question. First, these cities rely heavily on road transportation, which allows us to approximate connectivity with the road network in cities with less noise than elsewhere. Second, the timing of African urbanization being recent, we are able to build and gather data with higher level of precision and so observe more variation to study this question.

Our main measure of the quality of the road network is a simple measure of road density in the centre of towns and cities. We also compute a second and secondary simple measure of the distance to the nearest road for a random point in the city centre, as well as a number of measures of regularity and orientation of the road grid. All measures we use in this paper have the advantage to be intuitive, straightforward to compute and simple to interpret. Our conclusion focuses on the linear effect of our measures of road density and spatial distribution, which are the most robust and quantitatively most important of all the network features we consider. To address measurement error and potential endogeneity, we rely on two instrumental variable strategies based on the age of a city. We accept that the data available to us are limited, especially that we cannot observe historic road grids for a large panel of cities in a way that is comparable, and have to use proxy data. Despite the unavoidable limitations in this setting, we hope to show that our proximate measures are still good enough for this analysis to uncover a real effect. Our IV approaches also sometimes suffer from weak first stages in some specifications. Yet we think that the consistency of results across OLS and various IV specifications, despite their occasional weaknesses (which we show for completeness), overall points to suggestive evidence concerning the importance of roads in the city centre.

While there is a substantial literature on the effect of roads or other forms or transportation on urban shape (recent examples include Baum-Snow 2007, Duranton and Turner 2012, Faber 2014, Jedwab and Moradi 2016, Jedwab and Storeygard 2016, Michaels 2008 and Storeygard 2016), this literature typically has concentrated on roads connecting cities with one another. Less is known on roads and road layout within cities (exceptions include Baum-Snow et al 2017, who consider the effect of urban highways in China and Couture et al 2018 who suggest that restricting traffic would bring welfare gains). Our approach is more focused on within-city variation, and includes smaller roads than these existing studies. As we mentioned, the importance of road grid for urban development has long been recognized. Fuller and Romer (2014) write: "getting the grid right will likely be more important than enforcing building codes on structures or imposing limits on density". Yet, to our knowledge, there is a gap in research quantitatively assessing the importance of cities' road layout in economics. The topic has of course been studied in urban planning or environmental studies. Layout shapes have been shown to be associated with transport decisions, travel time and local CO2 emissions. Notably, in a recent series of articles, Barrington-Leigh and Millard-Ball (2019 and 2020) also use Open Street Map to provide global-level measures of within-city roads networks and confirm results about transport decisions. Their focus is not on agglomeration economies nor do they specifically study the sub-Saharan African context. While this region has known a rapid urbanization in recent decades and will most likely continue to do so, research is still needed on this context. In a literature review on developing-world cities, Bryan et al. (2019) insist on the need for more research on urban mobility in African cities.

Overall, our results suggest that denser road network are associated with more population growth. Our interpretation of this empirical relationship is that better connected city-centre foster interactions, agglomeration economies and eventually the growth of cities. We also find evidence that, given a specific density of road, the even distribution of roads across space in the centre is also a significant determinant of growth, and more so than, for instance the grid-orientation or the number of nodes in the network. As Barrington-Leigh and Millard-Ball (2019) show, the type of features we analyse here are associated with increased number of destinations that can be reach within a given time frame. And we take away from our results that such increase in connectivity is important for the growth of cities. Although we acknowledge limits to our identification strategies, we hope that the consistency of our results is indeed indicative of a causal relationship between road network and city performance (an identification mostly absent from the literature outside of economics). By these findings, we also provide evidence for a hypothesis advanced by Collier and Venables (2016) that the road network is an important predictor of recent population growth, and that many cities and towns in Sub Saharan Africa are constraint in their relative growth due to lack of road density.

The paper is organised as follows: Section 2 develops our hypothesis in greater detail, discusses its underlying assumptions and present our main measures of roads network. Section 3 describes the construction of the datasets used and presents descriptive statistics. Section 4 develops the empirical strategy and discusses our instrumental variable strategies we use. Section 5 presents the main results. We present a series of robustness checks in Section 6. Section 7 concludes.

2 Hypothesis and main measures

Our thinking about this problem, our hypothesis and our choice of instrumental variable are shaped by on six findings that are commonly reported in the urban economics literature. Given that they are central to our argument, we discuss each of these assumptions one by one in this section.

First, we assume that the returns from agglomeration are largely generated in the city centre. This is one of the standard assumptions used in monocentric city models. This suggests that people live around a city centre, to which they travel to work and enjoy leisure time. The agglomeration returns are overwhelmingly generated in that centre. We take from this assumption that we pay special attention to the road grid in the centre of cities rather than the whole road grid structure over the entire urban agglomeration. Although we also study the city as a whole for comparison for some of our measures, our preferred measure captures only the centre. Empirical evidence for the monocentric model has been found for a variety of cities, by demonstrating gradients of wages and house prices, as well as from commuter flows (for recent evidence for Africa see Antos et al 2016 or Larcom et al 2017). If anything, monocentricity has been found to be more pronounced in Africa than in the developed world (Cervero 2013, Grover Groswami and Lall 2016). Following this assumption, we define a city centre by the brightest spot of night light of a city, and draw a circle around it. In our main specification, this circle has a radius of 1km. This approach has the additional advantage that we do not need to define the edge of a city and how it changes over time, and that having similar sized areas makes the measure more comparable across cities. We do, however adjust the central area of a city in the case where we find rivers, coasts, or a boarder (see section 3.1 for more details).

Second, we make the assumption that the city centre is unique for each city we study, and we define only one centre for each town or city. This again is a standard assumption made in monocentric city models. In the context of our study this assumption is strengthened given that we include not only big cities, but also many small and medium sized towns, for which this assumption is more plausible. We also exclude capital cities in our preferred specification, which excludes some of the largest cities, which are more likely to have more than one centre, from our analysis. Antos et al (2006) provide evidence that is consistent with monocentric centres for a few large African cities, it is even more justified for small and medium sized cities, which are not large enough to support two centres. We can verify in our dataset that this assumption does not seem to be contradicted by the data.

Third, we assume that cities grow in a circular way around a given centre. Or to rephrase, we assume that the location of the city centre does not change much over time. This is true if the foundation of a city followed locational fundamentals that do not change much over time. Evidence can be found from all the many cities in the world that are still structured around Medieval or Ancient sites. A recent empirical paper making this assumption is Harari (2016). By this assumption, a city with a better working city centre will develop a larger commuter zone around it over time. This assumption is important to inform our identification strategy, since it allows us to assume that the historic centre and the modern centre coincide.

Fourth, we assume that road variables that we compute are important to facilitate the agglomeration returns, whatever these returns may be. This could be violated if most of the agglomeration returns happen overwhelmingly within buildings, and are less dependent on connection in the centre. Even in this case people would have to get to these buildings, but quantitatively the true agglomeration effects may be poorly approximated by our measures. This would also be violated if true connection in a city depends on factors other than those that we measure. It could be violated by some strict version of the 'fundamental law of road congestion' (Duranton and Turner 2011) applies, which could imply that road construction does not relieve congestion.

Fifth, we assume that there is path dependence of the road layout. Once a layout has been established, it is costly to change it, and the grid from one year overlaps strongly with the grid from another year. While there are examples of radical changes in the layout of city maps (Haussmann's renovation of Paris in the 19th century is a famous example), the great majority of cases that come to mind suggest indeed a tendency of lock-in.¹ Using a large data extraction from Open Street Map and covering a large share of world's cities, Barrington-Leigh and Millard-Ball (2019) study within cities roads layout and notably conclude that "street-network sprawl is a path-dependent process. Because streets are one of the most permanently defining features of cities". In economics, Baruah et al (2017b) compare colonial and modern street maps for some African cities. They find that "the spatial structures of cities in Sub-Saharan Africa are strongly influenced by the type of colonial rule experienced" and they point to "high costs of acquiring new rights of way in an already built-up city". Their examples that contrast colonial and modern maps show strong persistence of road layout. Shertzer et al. (2016) argue that Chicago zoning ordinances from 1923 have a larger effect on the spatial distribution of economic activity in the city today than geography or transport networks. Without this assumption cities would be able to self correct, our first stages would not be statistically significant, and our measures of the road grid would be poor predictors of future population growth.

Sixth, we assume that the optimal grid depends on the available transport technology. The ideal grid in the age of walking and the age of the horse differs from the ideal grid in the age of the car. For example, the latter might require broader roads and might avoid crossings of roads harder. If so, the age of a city influences its initial grid, which in turn influences its grid later on. This assumption helps us to develop our instrumental variable, which uses a correlate of the age of a city as an instrument for the initial road grid specification.

Taken together, these six results imply that initial differences of road layout, even if from a very distant past, can lead to different long run population developments. Due to small differences in geography and local history, some towns may accidentally stumble upon

¹Examples of large and successful cities whose central road layouts still reflect their Ancient or Medieval past abound in Europe. For evidence of path dependence for cities see Michaels and Rauch (2017). For evidence of path dependence of infrastructure in Africa see Michaels et al (2018).

a more successful layout than others. These initial differences depend on the age of a city, which is potentially observable.

These assumptions do not tell us what a successful layout should look like. We use the Open Street Map data set to measure several features of within-city roads like total length, number of nodes and intersections, orientation of each segment of road (bearing), etc. We then use these measurement either directly or in order to compute relevant characteristics of the layout. In particular, we build a simple measure of density of roads in the city, measured by road kilometres over area considered (often, the city centre only).² This strikes us as a simple and straightforward measure of road layout in cities. It also has the advantage of not taking any strong stand on the correct relationship between the observable features and the actual travel speed of residents. The latter not being obvious as having roads with multiple lanes may be helpful in bigger cities, but wasteful in smaller ones; crossings may slow traffic down, but enable useful travel combinations, etc. Since part of this question is also empirical, we will study the impact of directly-measured features like the number of interactions on city growth (either controlling for total road length or not).

In addition, an ideal measure to characterise a useful grid should capture how well connected people and firms in the city centre are. We try to approximate this ideal by measuring two main indices of the grid characteristics. We first try to capture the "evenness" of the layout and second its distance to a perfect grid in terms of "orientation". We build the evenness approach as a measure of how easy it is to access road network from a random point in the city centre. Facing a similar problem, Donaldson and Hornbeck (2016) create 200 random points within US counties, calculate the distance from each point to the nearest railroad, and take the average of these nearest distances. Following this algorithm, we create 12 points in each circle, measure the distance to the nearest road for each of these points, and compute the average nearest distance to the road from these 12 points.³ Because this measure decreases as the roads are less evenly distributed across space, we actually call it "unevenness index". To illustrate what we have in mind, consider Figure 1. This figure presents two cities with stylised road layouts. Both cities have the same road density in the centre by construction. The city on the right has a shorter distance from a random point to the nearest road. An example of a city with the 12 points is shown in Figure 5.

To capture the grid-aspect of a road network, we build a second index based on orientation of roads. Using the distribution of the (segments of) streets' bearing we compute an Herfindahl index of concentration. This simple and transparent measure takes its highest values when all streets run in either one of two directions, and two only (like a perfect grid does). We call this measure the "orientation Herfindahl index". Figure 6 illustrates two cities with high (top panel) and low (bottom panel) orientation Herfindahl. We describe further the construction of all measures in 3.3.

3 Data

We build a complete and consistent sample of cities, combining satellite imagery and geospatial population data in addition to other data sources. For each of these cities, we observe centre's road layout and measure a set of descriptive statistics. This section briefly describes datasets and measurement, all sources and processes are detailed further in the data

 $^{^2 \}rm When$ computing this density variable we count a roads recorded in the open street map data with n lanes n times.

³Using 12 points is a low number in comparison to 200. We do so for computational convenience. As a justification, we also record the mean distance from the centre of the circle to the nearest road, which is effectively the same exercise with that parameter set to one. The correlation between the exercise of 1 and 12 points is around 0.9. Given this large correlation, we believe that all such measures would give fairly similar numeric estimates.

appendices A and B.

3.1 Defining Cities: Boundaries and Centres

Our main source for the measurement of cities location and boundaries is satellite imagery of luminosity at night. This source has the double advantage of being complete (i.e. it covers the entire continent) and consistent (i.e. all the cities are defined in the same way). First, we identify lit areas from NOAA's DMSP-OLS satellite record (cf. appendix A) by keeping pixels that emit light at least twice over the five yearly observations from 2008 to 2012. Contiguous lit areas are then aggregated using a GIS software to create unions. Each union represent a city, and we consider its footprint as the city boundaries. Figure 2 illustrates our sample. As the Figure shows, we consider cities in almost all countries of Sub-Saharan Africa, with healthy spatial variation.

Our focus is on city centres, where we assume agglomeration effects to take place. Because, to the best of our knowledge, there is no dataset providing precise coordinates of cities' centre we rather follow a systematic data driven definition. For each city, we take as "centre" the centroid of the brightest area. This area could be a single lights pixel, but it could be a larger area of contiguous pixels, which also allow us to compute a centroid. In case of multiple centres of equal brightness in a city we choose the larger one. In theory there could be a problem of multiple, separate centres of equal brightness and size for a single city, but in practise we notice that this does not appear to occur in our dataset.

Two reasons at least support this definition of a city centre: first, because nightlights glow, it is often the case that brightest area is close to the geographic centroid of the city. Second, at the same time, nightlights are also known to be linked with economic activity, even at the local level. Figure 3 illustrates the derived city centres for two small towns. The figure also shows the city extent, computed from lights data, and the road network of these towns. Importantly, we chose to use early images of nightlights (1994-1996) to define city centre. This is for both empirical and theoretical reasons. Empirically, African cities emitted less light in the early period which facilitates the identification of the main historical centre. Theoretically we are interested in the lock-in of historical centre and therefore want to identify where this centre was as early as possible. This also reduces the sample to cities that were important enough in the mid-1990 to emit some light. This selection mainly exclude very small towns and is coherent with our instrumental variable strategies (cf. section 4.2).

3.2 City growth

Our main outcome of interest is the population growth of cities. We measure this outcome relying on the GHS dataset provided by the European Commission. The GHS is a spatial grid that provides a population estimate for each cell (of 250 by 250m resolution) in four target years: 1975, 1990, 2000 and 2015. The population estimates are obtained from the spatial distribution of census data, guided by GHS modelling of built-up presence (derived from Landsat day-time satellite images). For each city in our sample, we sum all the GHS cells falling within its boundaries and extract the population counts for the four target years. In appendix B.3 we discuss this measurement further.

Although or focus is on city growth from 2000 to 2015, where we think the data quality reaches its best, we also present robustness analysis using alternative outcomes and data sources.

3.3 Centre and Road layout

We construct city-centres' roads layout using the publicly available data from Open Street Map (OSM), a volunteered geographic information project consisting of information from

over 2 millions users. As with any user generated data, there is some concern that the dataset may be badly measured or biased. In their quantitative evaluation of OSM quality, Barrington-Leigh and Millard-Ball (2017) however conclude that the completeness of OSM was already high as of 2016, including for sub-Saharan Africa. In appendix B.2.1 we discuss this point in more details and present the robustness of our results using only higest OSM-quality African countries. Our main conclusions hold even in this smaller sample. Seidel (2019) also studies OSM bias in Africa by comparing OSM maps with information from other sources, primarily for hospitals. His findings imply that our standard specification with country fixed effects and additionally controls for the initial size of a city (which we always include) would address this bias to some degree. A specification with (within-country) province fixed effects would absorb most of the bias. We provide evidence in this paper that suggests our main coefficients are robust to both these versions.

For each city we gather data on the road network at two different levels. At the city level, we take into account any road that falls within a city boundaries; at the centre level, we draw a 1km-radius circle around the centre point and keep only the (segments of) roads that fall within. Figure 4 gives an example of a small city and highlights the derived circle of the defined central part. At both city and centre level, we measure a series of descriptive statistics about the road network. In particular we compute total length of roads, number of nodes and intersections, as well as the density of each of these items (i.e. dividing them by the considered area).⁴

As previously mentioned, we also compute two indices of the network spatial organisation: we call the first index "unevenness" and the second "orientation Herfindahl". For the unevenness index, we first create 12 points within each city centre circle and compute the average nearest distance to the road from these 12 points. Figure 5 illustrates the construction of this evenness measure, plotting the 12 points over a city centre (also showing the city and centre road networks). The index is thus expressed in a distance unit. The lower the index (i.e. the distance) the more likely the road network is evenly distributed over space. Second, we follow G. Boeing (2018) in constructing an orientation measure we then summarize in an Herfindahl index. To do so, we first compute the bearing of each edge (i.e. segment of road between two nodes) within a city centre and then observe the distribution of all edges across equal-sized bins of compass orientation (each bin covers 10 degrees). We then summarise the concentration of roads among potential orientations by taking the Herfindahl index of this edge bearing distribution. Figure 6 illustrates the construction of this orientation Herfindahl. The left two panels show the street grid in two city centres (Umm Ruwaba from Sudan at the top, and Mohale's Hoek from Lesotho at the bottom), defined as one kilometer radius circles around the centre point. The middle panels take the road segments from the first, and sort them by their orientation, weighting segments by their length. The final two panels take this information and display them as standard histograms. It is over these shares that we compute the Herfindahl indices of street orientation. In this context, the Herfindahl index is large if many roads run in parallel, and small if the grid is more chaotic. In practice, the highest values of the orientation Herfindahl correspond to an almost perfect grid (as the one displayed in the top panel of figure 6). Both the evenness and the orientation Herfindahl are related, and their correlation in the sample is pretty strong (0.42). If one think of a perfectly dense and grid-like city-centre, both indices would most likely reach their minimum and maximum (for the unevenness and orientation Herfindahl, respectively). But the two indices still capture different things: one

⁴Obviously, when considering only 1km radius city-centres, road density is just a multiple of total road length. There are however two cases where density is, to our view, a better measure: first, some 1km centre radius overlap water bodies or foreign countries that we exclude from the denominator; second, we also study city-wide measures and want to take city size into account in that case.

⁵Intersections are nodes with strictly more than one street emanating from; thus excluding cul-de-sacs, Boeing (2017).

can imagine a city where most roads are concentrated in one single neighbourhood (i.e. not evenly distributed) yet run in parallel (i.e. close to a perfect grid), and conversely. The role of any of these network features on a city growth is an empirical question.

Ideally we would want to measure the road layout at the beginning of the period over which we measure population growth, and not at the end. For a sample as large as ours this is however data we were unable to obtain. Instead we use road maps that are closer to the end of the period as a proxy. We make the assumption that the road grid does not change fast over time, particularly in the city centre. As discussed in our 5th assumption (cf. section 2), this assumption is supported by many evidences, including recent largescale research. We also note that our instrument corrects potential bias arising from this measurement error.

3.4 Sample description

By following the steps described in subsections 3.1, 3.2 and 3.3 we obtain a total sample of 1,850 cities from 40 sub-Saharan countries (cf. appendix C.1). From this sample our analysis notably excludes capital cities that we expect to follow unique development paths within their countries. We also exclude two countries from the main analysis. First, we exclude Nigeria because nightlights data suffer from potential issues in the Gulf of Guinea. Second, we exclude Madagascar as its own history is incompatible with one of our historical IV. These three exclusions (capital, Nigeria and Madagscar) reduce our sample to 1,412 cities from 38 countries (cf. appendix C.2). Given the importance of these last three exclusions, we do present robustness results including all 1,850 cities and show that our results remain similar and that our conclusions hold true. More generally, appendix C details all the data constraints and decisions that define our sample.

Table 1 provides a break-down of our samples by country. In 2000, the median city size is around 33 thousand inhabitants, varying greatly across countries (from 2,500 in Djibouti to above 100,000 in Chad and South Sudan, where we identify only 12 and 3 cities, respectively). Average population growth rate over the 2000-2015 period is estimated to be 2.04%. And there again, countries experienced different paths with city population growth spanning between 5.66% a year in Equatorial Guinea to negative growth in Swaziland (-1.04%).

As for the road network of these cities, Table 2 provides simple descriptive statistics. Here again, we find a high variation in the total amount of road length within cities (from virtually 0 to 28,621km) and even within centres (0 to 87km). Although the maximum values seem high, they reflect that cities defined by nightlights can reach very large size, magnified by lights glowing. In our sample, the largest area encompasses Lagos as well as peripheral cities like Abeokuta in the north or Ijebu Ode in the east. Table 2 also reports a basic descriptive of the distances from cities' centre to the sea, and to the four cities (Dakar, Dijbouti, Cairo and Cape Town) we use four our IV specification.

4 Empirical strategy

To uncover the role of street networks in fostering city growth we first present a simple regression model in section 4.1. Because this model may suffer from classic limitations, we also present and discuss two IV strategies in the following section 4.2.

4.1 Baseline model

The estimation equation we use for both sets of results is of the form:

$$\Delta_1^T Pop = f(Street \ network) + \mu_c + \beta_1 ln(Pop_{t=1}) + \beta_2 ln(dist. \ sea) + \gamma X + \epsilon.$$
(1)

We compute a measure of annualised population growth defined as $\Delta_1^T Pop = (ln(Pop_{t=T}) - ln(Pop_{t=1}))/(T-1)$. In the main regression the two years for which we observe population levels are 2015 and 2000, and so T-1 is equal to 14 here. f is a function of our main measures of the road network in a city centre. We sometimes use country fixed effects or regional (i.e. within-country areas), denoted as μ_c above. In our main specifications we control for the initial population density of a city $(ln(pop_{t=1}) \text{ and } (log)$ distance to the sea (ln(dist. sea)), which is important to condition on for one of our instruments. Finally, we also show results including a richer set of control variables such as city-specific ruggedness, elevation, minimum temperature, maximum temperature, average temperature, precipitation, distance to the big lakes and climatic conditions for malaria. We use robust standard errors throughout.

The coefficient associated to f(Street network) identifies the causal impact of a feature of the city-centre road network on city's population growth only if that feature is independent of any factor influencing city growth (that are not controlled for). Put differently, this amounts to consider that two cities within a same country, with similar initial population and distance to the sea would grow similarly in expectation if they had the same city-centre road network. In particular, we must ensure that cities growing faster (or slower) do not change their road network. This is another reason why we focus on the city centre, which is less prone to new road construction than city fringes would be. We sometimes include city-specific control variables that could capture geographic endowments influencing both the city-centre road network and population growth. In robustness checks we also include market access measures.

Table 3 displays the change in the coefficient for different versions our main specification, and highlights that the magnitude does not seem to be greatly affected by the inclusion of controls, country fixed effects or different ways of selecting the sample. The independent variable of interest here is city-centre road density, but we are more interested by the comparison across specifications; we interpret the results thoroughly in section 5. Our conclusion from Table 3 is that the main coefficient of interest remains close to unchanged across alternative specifications. Put differently, the estimate of our favourite specification (column (1)) is not statistically different from any of the alternatives displayed. More specifically, in column (2) we exclude the country fixed effect; in column (3) we instead include a (withincountry) province fixed effect; in column (4) we also add control variables (cf. above); in column (5) (respectively (6)) we also add the cities of Nigeria (and Madagascar) to our sample. Both sign and magnitude of the coefficient of interest remain statistically close.

4.2 Instrumental Variable strategies

One concern with the empirical strategy based on equation 1 is the potential role of reverse causality. If cities that grow faster also build more road or improve the OSM data set faster then our first model would not identify the causality of the road network on population growth. To circumvent this threat, we implement two related instrumental variable (IV) strategies using a city's age as an instrument for its city-centre road network. As discussed in Section 2, the current layout of a city-centre is likely to be a direct product of the initial, historic layout. In the IV regressions we control for initial population density in these regressions. We expect to see some road network feature changing across cities of different ages. For example, a city founded before the invention and adoption of cars would have a higher density suitable for older technologies. We expect more modern cities, built in the era of increasing adoption of cars, having bigger roads and also smaller road densities.

To measure the foundation age of a city, we rely on the Africapolis (2019) dataset, that records historic population for a large number of cities in Sub-Saharan Africa. The oldest year in that dataset is 1950, and our instrument consists of a simple dummy variable, indicating those cities that existed in that year according to this dataset (i.e. that had 1 or more inhabitants in 1950). One downside of this "age dataset" is that it is only available for a subset of our cities, and reduces our sample.

Given remaining concerns that the age of a city, even if crudely measured, might influence recent population growth through channels other than the road layout in the city centre. To address this concern, we also construct a second, related instrument, measuring the distances from the cities of Cape Town, Cairo, Dakar and Djibouti. These four cities represent connection nodes of the two main colonial powers of the continent. If a significant part of colonial expansion was conducted via these ports, distance to these four cities should influence the year of creation of modern cities on the continent. In these specifications we control for the distance to the sea and again for initial population density. (And, as argued for the first instrument: if the age of a city correlates with the initial road layout, then the distance to these four cities will also correlate with the initial road layout.) Using the distance from cities to each of these four points is a simplistic vision of history that captures enough information that it might function as an instrument.

Urbanisation in Africa was facilitated to a large degree by the colonial powers, especially the British and the French, the two main parties in the scramble for Africa. As late as 1880, about 80 percent of Africa were ruled by Africa's own kings and queens. Colonial powers were found near the coast, with a British stronghold around the port of Cape Town, and French holdings around Dakar (Boahen 1985). The British invasion of Cairo of 1882 established a second stronghold on the continent for the British that lasted well into the 20th century (Hourani et al 2004). Soon it became the vision of the British to connect these two centres, and build the "Cape to Cairo" Railway line. Historically, these two cities played an important part as connection points of British colonial Africa, and many expeditions, travels and trade took off in one or the other. France's influence on the eastern coast of Africa was initiated by treaties with the rulers of what is today Djibouti from 1883, and the creation of outposts in modern day Dakar. Soon the French ambition became to establish an East-West link between its posts in Dakar and Djibouti. French and British expansions clashed near the town of Fashoda, in the Fashoda incident of 1898 (Bates 1984). In schoolbooks, this episode in history is sometimes illustrated with maps that show the cities of Cairo, Cape Town, Dakar and Djibouti, with four arrows meeting near Fashoda. This simplified schoolbook view of colonial expansion in Africa is the model we follow in the construction of our instruments, taking the distance to these colonial origin ports as instrument for the age of a town, and consequently its initial road grid.

4.2.1 First-stage regressions

Table 5 shows the correlation between our main right hand side variables, which are the measures of road density and unevenness, and the distance to the four instrument cities. This is effectively a first stage regression. As expected, we observe a significant negative or no relationship between the distance to the ports and road density in the city centre in Columns (1) - (4), which could reflect that newer cities that were designed with modern transportation in mind rely on a less dense road grid. It is negative in the case of three of our four cities and insignificantly different from zero at five percent level of statistical significance in the case of Cape Town.⁶ When we jointly include all four cities in one regression and add various control variables as in column (5) we continue to observe negative coefficients, typically at a high level of statistical significance. Coefficients become weaker and even flip sign when we include country fixed effects, as in column (6). This seems plausible to us, since we expect the distance to these colonial ports to matters less within countries than for

 $^{^{6}}$ On average, the coefficient in columns (1)-(4) is -0.87, perhaps close to the distance coefficient of -1 sometimes estimated for distance coefficients (Rauch 2016).

the longer distances across the whole continent. Because of this weak instrument problem, we show the main results later with and without country fixed effects. In our IV strategy we prefer to use the four distances as separate variables to simply using the minimum distance given that British and French colonial legacies could have influenced cities in different ways (Baruah et al. 2017b). In columns (7) and (8) we repeat the estimation from columns (5) and (6) but using unevenness on the left hand side, rather than density in just the city centre. We find a strong relationship without country fixed effects and a much weaker one when we include country fixed effects.

In columns (9) and (10) we provide more direct evidence that distance to these historic ports is indeed correlated with the age of a city. For this analysis we use the data from the "Africapolis" project. Keeping only cities in our dataset that are also in the Africapolis reduces the sample to 1,070 observations. We then compute an indicator for cities that, according to Africapolis, did not have any population in 1950. As columns (9) and (10) show, there is a significant relationship between these distances and the age of a city. Once we include our control variables and country fixed effects, the relationship is positive, which confirms that cities that are further away from these ports are more likely to be new. Columns (11) and (12) are then first stage regressions when using the age indicator for cities as an instrument. Finally, in Columns (11) and (12) we show that this simple indicator variable for cities that are older than 1950 correlates strongly significantly with the density of roads in the city centre and unevenness in the expected way.

4.2.2 Exclusion restriction

These IV strategies could uncover the causal relationship between the road network and population growth only under the exclusion restriction that the age of a city influences its recent population growth only through its impact on the road layout. It is important to immediately add that in our 2SLS estimation we always include: initial population size and distance to the sea coast as well as country (or province) fixed effects. In other words, the effect of a city's age through these two channels or that is time-invariant but country-specific are taken into account and do not threaten the exclusion restriction.

If, as we argued, age and location of cities are related then geographic factors could potentially influence city growth through other channels than just the city-centre road layout. The inclusion of the distance to the coast does not only address the concern that the four ports (Cairo, Cape Town, Dakar and Djibouti) still play a large role as transportation hubs today but it also picks up other influences the coast may have on the local economy. We also note that there are large ports available along the African coast, apart from these four cities.⁷ In addition to seaports, airports now play a substantial road in connecting Africa, which further reduces the importance of these four historic ports. Also, to address concern relating to the location of cities within the continent we also provide robustness specifications in which we control for market access measures of each city in our data set. Including such measures barely change the magnitude and significance of our coefficients of interest.

Another concern for our instrumental variable strategies would include all colonial legacy that depends on age. For example, there may be physical infrastructure other than roads whose age depends on these distances in a similar way, and that have a similar effect on future development. We can't dismiss this concern entirely, but take note of some limits to this concern. First, as we argued previously, we think that the road layout is particularly prone to lock-in, compared to other types of infrastructure. Additionally, road construction and city planning were of central importance of all potential colonial investments. Second,

⁷The busiest ports of Africa are (in order descending in size according to total cargo volume): Richard's Bay, Saldanha Bay, Alexandria, Damietta, East Port, Apapa, Casablanca, Skikda, Pointe Noire, Mombasa, Abidjan, Tanger, Bejaia, Jorf Lasfar and Tin Can Island (AAPA 2015).

building reconstruction rates in Africa are very high. Baruah et al. (2017a) estimate an annual replacement rate of five percent for housing in Tanzania. Henderson et al. (2017) find a replacement rate of 3.6 percent in Kenya. Given these high replacement rates, most houses and most other physical infrastructure would have been rebuilt several times since the early 20th century, and have much time to converge from an initial steady state to a new one. Path dependent behaviour of infrastructure quality is far less likely in such a rapidly changing environment. Of course also the physical characteristics (the "quality" and size) of the roads were changed and updated multiple times in the great majority of our towns and cities. It is however the layout of roads, the plan, that we think is harder to change. Third, a large share of public infrastructures in Sub-Saharan Africa happen to be concentrated in capital cities (Bekker and Therborn 2012), which we exclude from the analysis. If public policy programs are less common outside then there are fewer channels through which administrative legacies could manifest. Note also that, in our baseline specifications we always include a country (or province) fixed effect. This also addresses other concerns related to institutional legacies. Any institutional setting that is set (and invariant) at the country (province) level would be absorbed by the fixed effects.

Finally, we are reassured by the fact that the key results from the IV specifications are robust to many alternative specifications or sample changes. Furthermore, their main results tend to either confirm or strengthen the estimates from the simple OLS estimation.

5 Results

5.1 Results from OLS estimation

To estimate the impact of various features of the city-centre road network on cities' growth and prosperity, we first estimate of equation 1. In this specification, we always include a country fixed effect as well as (logs) initial population density and distance to the sea coast (cf. section 5.1 for more details). We investigate the relationship between our simple variables of road network and population growth.

Table 4 reports the results. Column (1) shows that a greater road density in the city centre correlates positively and strongly statistically significantly with population growth in the period 2000-2015. The measure of road density in the city centre "d centre" is measured in km/km^2 , and has a mean of around 6. Increasing the road density in the city centre by one km per km^2 , is associated with more than one tenth of a percent of higher population growth annually. In column (2) we show the estimated impact of the number of nodes (i.e. any intersection with one road or more) within the city-centre also correlates positively with city growth but column (3) shows that this effect disappears when the density of roads is taken into account. In other words, the number of nodes correlates with city growth in-so-far as it is a direct product of the total amount of road.

In columns (4) and (6) we turn to our two indices of spatial organisation, the unevenness and orientation Herfindahl indices, respectively. We describe this indices in section 3.3. The main result in column (4) indicates that as the evenness of the road grid (respectively, the index) increases (decreases) so does the population growth over 2000-2015. In column (5) we also add city-center road density as an independent variable. The concern here is that two cities with identical road density as measured by our *d centre* measure may nevertheless have quite different flows in the centre. Similar to the result in column (4), the negative sign associated to the unevenness index indicates that larger average distances to the nearest road harm population growth, holding road density constant, as expected.⁸ In column (5) we

⁸The unevenness index is expressed in meters so the coefficient in column (4) can be read as: within a given country, cities where the evenness is 1m smaller also had a population growth a 3.10^{-5} of a percent higher (given initial city population and its distance to the sea coast).

estimate the correlation between the orientation Herfindahl index and population growth: a small negative correlation appears (only significant at the 10%level) but the coefficient shrinks and become insignificant in column (7), when road density is taken into account. In other words, two cities (within a same country, with the same initial population and distance to the sea coast) with the same amount of city-centre roads do not benefit more from having all its roads sharing a few main bearings. As explained in section 3.3, both the unevenness and the orientation Herfindahl indices eventually measure different things, although they correlate quite strongly (0.42). One could for instance think of a city-centre where all roads run north to south (highest Herfindahl index) yet are all concentrated in one specific neighborhood (highest unevenness); of these two forces, the unevenness seems a dominant factor for city growth.

In column (8) we estimate a model including all the road network features at once (i.e. street density, number of nodes, the unevenness index and the orientation Herfindahl index). We draw to main conclusions from this exercise: first, the only feature which impacts remain strongly significant is the unevenness index. Such results indicate that the spatial distribution of the network matters even after the total amount of roads and nodes, as well as their overall bearing concentration are taken into account. Second, although the significance of the road density collapses, the point estimate remains quite similar in most specifications (0.00103 here, vs. 0.00110 in column (1)). We take this result as an indication of the robustness of our result and of the importance of road density. Note also that city-centre road density directly correlates with all other independent variables, and it is therefore not surprising that the standard errors almost triples here compared to column (1). In column (9) and (10) we substitute either (within-country) province fixed effects or no fixed effect, respectively, to the country fixed effect. Column (9) is a strategy to reduce the potential measurement bias (Seidel, 2019; cf. section 3.3) while column (10) mirrors subsequent specification in the remainder of this paper. In both case, the effect of the unevenness index is reduced and remains significant only at the 10% level; while the coefficient of road density increases (and even becomes significant in the last specification) but still remains statistically close to other estimates.

5.2 Results from the 2SLS estimation

We next turn to the IV estimates corresponding to the OLS results derived so far. First, we use an indicator variable for new cities as an instrument for the road layout in the city centre in Table 6. This variable indicates cities that were founded after 1950 according to the Africapolis dataset (cf. section 4.2 for more details). Given that not all our towns and cities feature in the Africapolis dataset, this reduces the size of our sample to 1,070. This instrument follows the same logic as the port distances, which is that older cities are more likely to be stuck with sub-optimal grids; but it measures the age of a city directly. The advantage of this strategy is that we get a strong first stage, with the exception of the specification in column (5). All coefficients on road density are positive and statistically significant, while magnitudes on unevenness are negative and statistically significant. Magnitudes are larger than in the OLS estimate, which might be explained by a reverse causality problem in the OLS regression, and some scope for measurement error. Qualitatively, this table supports the OLS findings on density and unevenness, and shows that both are strongly associated with population growth in recent years.

Table 7 presents results corresponding to the OLS results in Table 4, but this time estimated as 2SLS using distances to colonial ports as instruments. Coefficients retain the expected signs, positive for road density and negative for unevenness. This table suffers from weak first stages in columns with country fixed effects. We still report the coefficient for comparison. We note that coefficients are similar in magnitude to the ones estimated using the other instrument.

Overall, both IV strategies seem to confirm the qualitative results from our baseline OLS estimation. The impact of city-centre road length density remains positive and significant with the 2SLS estimations. Similarly, the coefficient associated to the unevenness index indicates that better spread roads have a positive impact on city growth over the 2000 2015 period and 2SLS only confirms this conclusion. For both measures and both instrument, the point estimate of the coefficient is greater in 2SLS than in OLS, suggesting that the potential bias in OLS attenuate the true effect of roads. We acknowledge than these two IVs have limitations, but we find it worth reporting that they yield comparable conclusions regarding both the direction of the effect of road network and the direction of the potential bias in OLS.

In terms of magnitude, a coefficient of 0.005 on road density implies that if a city increases its central road density by 1km per square kilometre, its total population would grow by an additional 0.005 per year. We can compare this with the mean of 6.4km in our sample, and a standard deviation of 5.2. On unevenness, a coefficient of -0.0002 suggests that increasing the unevenness measure by 100 reduces population growth by 0.02 percent per year. Given that the mean unevenness is only around 200, this is a less quantitatively strong effect than the one we report concerning road density.

6 Robustness Checks

In this subsection, we show that our results are robust to some of the choices guiding our main specification.

In Table 8 we present the results associated to alternative models or samples. The 6columns reports our 6 main estimations: the main independent variable is road density in columns (1) to (3) and unevenness in columns (4) to (6); for both independent variables we present the results associated to our basic OLS (columns (1) and (4), respectively), our age IV (columns (2) and (5)) and our distance IV (columns (3) and (6)). Panel A displays the baseline results (i.e. including log initial population and distance to the sea as well as a country fixed effect) and therefore reproduces coefficients found in previous tables. Each of the following 5 panels report the estimates of the coefficient of interest when relying on alternative models or samples. Panel B adds measures of market access to the baseline model. We measure market access by distance weighted populations of all other cities in our data set. The weights we use are -1, -2 and -4, as in Maurer and Rauch (2019); and we include all three market access measures in logs to the main model. Controlling for market access should capture the geography-induced trade factors that favours (or impede) city growth. This is a potentially important channel which also threaten the exclusion restriction of the distance instrument. Panel C replaces the baseline country fixed effect with a within-country province fixed effect. Such change may absorb more of the measurement bias in OSM data (as suggested by Seidel, 2019). Panel D and E change the sample, introducing back: capital cities and Nigeria, respectively. We argued against the inclusion of capital cities (because of their particular status) and Nigeria (because of potential measurement errors) but also provide results when including one or the other so as to be comprehensive. Finally, in Panel F we replace city-centre measure of road density by city-wide measure of road density. Since there is potential errors in identification of the historical centre of cities, taking all roads within the nightlights boundaries is a sensible test. The sign of the coefficient in each robustness check remains similar to the baseline; significance is often similar as well, except with the distance IV where precision varies substantially. Finally, the magnitudes also remain pretty stable across specifications, with larger estimates for IV specifications.

In Table 9 we again reproduce our 6 main estimations where panels A to C report the coefficients associated to road density while panels D to F report the coefficients associated

to unevenness; OLS is in panels A and D, age IV in panels B and E, and distance IV in panel C and F. The column (1) reports the baselines results while columns (2) to (5)present alternative outcome variables. In column (2) we replace the main dependent variable (annualised population growth rate over 2000-2015) by the annualised population growth rate over the longer 1975 to 2015 period. In column (3), we use the $(\log of)$ population level in 2015; this is equivalent to a long run population growth rate starting at population levels close to 0. Finally, in columns (4) and (5) we use alternative data sources to compute population growth. Namely, we first use the Gridded Population of the World (GPW) data set in its 4th version; this source is actually one of the components entering the GHS data we mainly rely on throughout the paper. Second, we directly use the Africapolis data set from the OECD that we only used so far in order to build the age instrument. We think GHS data has both a greater coverage and a higher quality than the two alternatives presented here, and we discuss this aspect in Appendix B.3. In all cases, for each robustness test we use a respective measure of the initial population. As in the first robustness table, most results are qualitatively similar and our main conclusion remains the same. We note that the magnitude of the effects is lower when using alternative data sources and that we lack precision when using the Africapolis data set.

7 Conclusion

Having a road layout in the centre of a city that facilitates interaction between people is an essential urban public good. Providing this good is an important policy concern, particularly in the fast growing and urbanising Africa. In this paper we highlight that the layout of roads in the centre of a city influence population growth of African cities in recent years. Cities that have either a road density that is too low in the city centre, or an uneven road network in the centre grow less than other cities. Cities with too few roads or a too uneven network may hit constraints on growth and fall behind other cities. We hope to have provided evidence of the causal relationship between road networks and city growth.

Our own reading of the results consider connectivity and agglomeration economies (or density) to be the key factors at play. We also consider that, given the magnitudes of the issue, local policy makers are most likely already aware of the problem, yet unable to retrofit better road layouts. Correcting infrastructure ex-post seems be prohibitively costly. Rather, our take away is to stress the importance of carefully planning network ahead and to consider the importance of connectivity. The later is the product of other aspects of within cities network we do not specifically study here, such as the quality of the infrastructure and other means of transportation. But given its primacy in transport infrastructure investment, we think the very shape of the road system is a key aspect, especially in fast urbanising contexts.

References

- AAPA American Association of Port Authorities, "World Port Rankings 2015", downloaded from http://www.aapa-ports.org on March 22nd, 2017.
- Africapolis, Aggomeration data, downloaded from http://africapolis.org/data in March, 2019.
- Antos, Sarah Elizabeth, Somik V. Lall, and Nancy Lozano-Gracia. "The morphology of African cities." (2016).
- Ahlfeldt, Gabriel M., et al. "The economics of density: Evidence from the Berlin Wall." Econometrica 83.6 (2015): 2127-2189.
- Barrington-Leigh, Christopher, and Millard-Ball, Adam, (2017). "The world's usergenerated road map is more than 80% complete", PLOS ONE, 12(8), 1-20.
- Barrington-Leigh, Christopher, and Millard-Ball, Adam, (2019). "A global assessment of street-network sprawl". PloS one, 14(11), e0223078.
- Barrington-Leigh, Christopher, and Millard-Ball, Adam, (2020). "Global trends toward urban street-network sprawl", Proceedings of the National Academy of Sciences, Jan, 117 (4) 1941-1950.
- Baruah, Neeraj, Guy Michaels, Dzhamilya Nigmatulina, Ferdinand Rauch and Tanner Regan, 'Upgrade slums or invest before they form? Evidence from Tanzania', Mimeo, 2017a.
- Baruah, Neeraj G., J. Vernon Henderson, and Cong Peng. "Colonial legacies: shaping African cities." (2017b).
- Bates, Darrell. The Fashoda incident of 1898: encounter on the Nile. Oxford Univ Pr, 1984.
- Baum-Snow, Nathaniel. "Did highways cause suburbanization?." The Quarterly Journal of Economics 122.2 (2007): 775-805.
- Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics, 99(3), 435-448.2
- Bekker, Simon B., and Göran Therborn. Capital cities in Africa: power and powerlessness. Hsrc Press, 2012.
- Bluhm, Richard and Melanie Krause, 'Top Lights Bright cities and their contribution to economic development', Draft paper, 2018.
- Boahen, Adu, "Africa and the colonial challenge", in Boahen (edt.) 'General History of Africa VII', United Nations Educational, Heinemann Books, London, 1985.
- Boeing, Geoff. "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks." Computers Environment and Urban Systems, vol. 65, pp.126-139, 2017.
- Boeing, Geoff, "Urban Spatial Order: Street Network Orientation, Configuration, and Entropy". 2018.

- Bryan, Gharad, Edward Glaeser and Nick Tsivanidis, 'Cities in the Developing World', NBER Working paper No. 26390, 2019.
- Cervero, Robert, (2013) 'Linking Urban Transport and Land Use in Developing countries', Journal of Transport and Land Use, 6(1), 7-24.
- Collier, Paul, and Anthony J. Venables. "Urban infrastructure for development." Oxford Review of Economic Policy 32.3 (2016): 391-409.
- Couture, Victor, Gilles Duranton, and Matthew A. Turner. "Speed." Review of Economics and Statistics 100.4 (2018): 725-739.
- Donaldson, Dave, and Richard Hornbeck. "Railroads and American economic growth: A "market access" approach." The Quarterly Journal of Economics (2016).
- Duranton G, Puga D. 2004. Micro-Foundations of Urban Agglomeration Economies. In Handbook of Regional and Urban Economics, ed. J. V. Henderson and J.-F. Thisse, Amsterdam: Elsevier. Vol. 4: 2063-2117.
- Duranton, Gilles, and Matthew A. Turner. "The fundamental law of road congestion: Evidence from US cities." American Economic Review 101.6 (2011): 2616-52.
- Duranton, Gilles, and Matthew A. Turner. "Urban growth and transportation." Review of Economic Studies 79.4 (2012): 1407-1440.
- Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. Review of Economic Studies, 81(3), 1046-1070.
- Fetzer T., Henderson V., Nigmatulina D. and Shanghavi A. "What Happens to Cities When Countries Become Democratic?" (2016).
- Fuller, Brandon, and Paul Romer. "Urbanization as opportunity". The World Bank, 2014.
- Grover Goswami, Arti G., and Lall, Somik V., (2016), 'Jobs in the City: Explaining Urban Spatial Structure in Kampala', World Bank Policy Research Working Paper 7655, World Bank, Washington D.C.
- Harari, Mariaflavia. "Cities in bad shape: Urban geometry in india." Job market paper, Massachusetts Institute of Technology (2015).
- J. V. Henderson, T. Regan, A. Venables, "Building the city: Sunk capital and sequencing" (Discussion paper 11211, Centre for Economic Policy Research, London, 2016).
- Hourani, Albert Habib; Khoury, Philip Shukry; Wilson, Mary Christina (2004). The Modern Middle East: A Reader (2nd ed.). London: I.B. Tauris. ISBN 1-86064-963-7.
- Jedwab, R., & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: evidence from Africa. Review of economics and statistics, 98(2), 268-284.
- Jedwab, R., and A. Storeygard. "The Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010." Unpublished manuscript, Tufts Univ., Medford, (2016).

- Kocornik-Mina, Adriana, Thomas McDermott, Guy Michaels and Ferdinand Rauch (2019), 'Flooded Cities', American Economic Journal: Applied Economics, forthcoming.
- Lall, Somik V. and Henderson, J. Vernon and Venables, Anthony J, (2017), Africa's Cities: Opening Doors to the World, World Bank, Washington DC.
- Larcom, Shaun, Ferdinand Rauch, and Tim Willems. "The benefits of forced experimentation: striking evidence from the London underground network." The Quarterly Journal of Economics 132.4 (2017): 2019-2055.
- Marshall A. 1920. Principles of Economics. London: Macmillan.
- Maurer, Stephan, and Ferdinand Rauch. "Economic Geography Aspects of the Panama Canal." (2019), Oxford Economics Working Paper.
- Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. The Review of Economics and Statistics, 90(4), 683-701.
- Michaels, G., Nigmatulina, D., Rauch, F., Regan, T., Baruah, N., and Dahlstrand-Rudin, A. (2017). Planning Ahead for Better Neighborhoods: Long Run Evidence from Tanzania.
- Michaels, Guy, and Ferdinand Rauch. "Resetting the urban network: 117–2012." The Economic Journal 128.608 (2018): 378-412.
- Rauch, Ferdinand. "The geometry of the distance coefficient in gravity equations in international trade." Review of International Economics 24.5 (2016): 1167-1177.
- Seidel, Andre, (2019) "Regional Public Goods Provision, Decentralization and Social Heterogeneity", Draft paper.
- Shertzer, A., T. Twinam, and R. P. Walsh (2016, July). "Race, Ethnicity, and Discriminatory" Zoning. American Economic Journal: Applied Economics 8 (3), 217-246.
- Storeygard, Adam. "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa." The Review of economic studies 83.3 (2016): 1263-1295.

8 Tables and Figures

	Cities	Sample Share $(\%)$	Median Pop.	Av. Pop.	Av. Pop. Growth
Angola	33	1.78	54130	163514	4.78
Benin	27	1.46	43352	58856	1.47
Botswana	29	1.57	18382	32626	0.67
Burkina Faso	27	1.46	44998	62813	1.94
Cameroon	69	3.73	34104	93522	1.01
Central African Republic	10	0.54	25628	24958	0.74
Chad	12	0.65	101813	109497	4.10
Congo, Dem Republic of	52	2.81	88960	177904	2.30
Congo, Republic of	15	0.81	18752	65410	3.75
Côte d'Ivoire	131	7.08	17728	68461	1.03
Djibouti	5	0.27	2479	2581	3.86
Equatorial Guinea	2	0.11	47880	47880	2.69
Eritrea	5	0.27	4317	21045	5.66
Ethiopia	113	6.11	50527	70516	3.41
Gabon	25	1.35	10507	18436	1.93
Gambia	5	0.27	22254	26840	1.75
Ghana	88	4.76	41778	81420	0.77
Guinea	23	1.24	40946	52615	2.54
Kenya	80	4.32	29856	71457	2.92
Lesotho	10	0.54	19638	38970	0.15
Liberia	2	0.11	37678	37678	0.59
Madagascar	33	1.78	15292	19509	2.88
Malawi	57	3.08	13495	30477	2.79
Mali	30	1.62	30466	46283	2.94
Mauritania	15	0.81	17271	20955	2.50
Mozambique	62	3.35	33707	68073	1.64
Namibia	42	2.27	5246	13361	1.71
Niger	34	1.84	27652	69991	3.21
Nigeria	405	21.89	50889	150988	1.60
Rwanda	9	0.49	63907	57128	2.40
Senegal	54	2.92	29195	74327	2.20
Sierra Leone	6	0.32	92277	80577	1.93
South Sudan	3	0.16	101410	142367	1.64
Sudan	52	2.81	91025	141978	3.63
Swaziland	7	0.38	13838	30454	-1.04
Tanzania, United Rep of	100	5.41	26104	56960	2.59
Togo	22	1.19	51132	62932	0.50
Uganda	38	2.05	33671	54929	2.97
Zambia	56	3.03	45171	72605	2.12
Zimbabwe	62	3.35	22623	50864	0.95
Total	1850	100.00	32972	87565	2.04

Table 1: In total, we consider 1850 cities from 40 countries. In our main specification we nevertheless exclude Nigeria and Madagascar, reducing our sample to 1,412 cities. The first two columns detail each country total number of cities and contribution to the sample, respectively. The following three column gives indication of poulation size in 2000 (median and average) as well as the average population growth rate in cities over the 2000-2015 period.

	Ν	Mean	sd	Min	Median	Max
Whole City						
Total Road Length (km)	1850	243.41	878.49	0	73	28621
Meters of Road per Sq. Km	1850	2490.88	2176.68	0	1850	17569
Area (Sq.Km.)	1850	94.64	278.85	2	40	7408
City Centre (1km. radius)						
Total Road Length (km)	1850	19.55	16.06	0	15	87
Meters of Road per Sq. Km	1850	6402.65	5160.03	2	5205	27703
Area (Sq.Km.)	1850	3.10	0.21	1	3	3
Dist. Centre to closest Road (m)	1850	133.29	166.94	0	62	975
Orientation Herfindahl index	1850	0.19	0.19	0	0	1
Unevenness index (m)	1850	189.15	155.37	10	150	1078
Distances from City Centre						
Distance to Dakar (km)	1850	4065.62	2070.82	78	3487	8272
Distance to Djibouti (km)	1850	3583.33	1450.88	34	3773	6507
Distance to Cairo (km)	1850	4007.16	1046.31	918	3875	6738
Distance to Cape Town (km)	1850	4337.90	1311.26	628	4838	7168
Distance to the closest (km)	1850	2064.75	830.10	34	2135	3678
Distance to the sea (km)	1850	451.14	350.46	0	389	1711

Table 2: Cities: Descriptive statistics

Table 3: Baseline model and alternative OLS. *d centre* measures road density in the city centre. The set of controls include variables measuring ruggedness, elevation, minimum, maximum and average temperature, precipitation, distance to the lakes and the coast and malaria risk. Robust standard error. Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta^{2015}_{2000} Pop$					
d centre	0.0011^{***}	0.0013^{***}	0.0011^{***}	0.0013^{***}	0.0011^{***}	0.0011^{***}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
$\ln pop 2000$	-0.0083***	-0.0072***	-0.0085***	-0.0077***	-0.0059***	-0.0060***
	(0.0018)	(0.0016)	(0.0017)	(0.0017)	(0.0013)	(0.0013)
ln dist sea	0.0000	0.0011**	0.0000	-0.0006	-0.0006	-0.0009
	(0.0007)	(0.00049)	(0.0014)	(0.0017)	(0.0016)	(0.001)
Observations	1.412	1.412	1.412	1.412	1.817	1.850
Country FE	Yes	_,	_,	_,	_,=	_,
Province FE				Yes	Yes	Yes
Controls			Yes	Yes	Yes	Yes
Nigeria includ					Yes	Yes
Madaga. includ.						Yes

Table 4: Results from simple regression model. d centre measures road density in the city centre; n centre is the number of roads' nodes in the city centre; *uneven* is the unevenness index; *orientation* H is the orientation Herfindahl index. Robust standard error. Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

	$\Delta^{(10)}_{2000} Pop$	0.0014^{*}	(0.0008)	(000.0)	-0.00001^{*}	(0.000)	0.0031	(0.0053)	-0.0075***	(0.0016)	0.0011^{**}	(0.0005)	1,412	
	$\Delta^{(0)}_{2000} Pop$	0.0013	(0.0008)	(000.0)	-0.00002^{*}	(0.00)	-0.0003	(0.0051)	-0.0078***	(0.0017)	0.0007	(0.0014)	1,412	Yes
	$\Delta^{(8)}_{2000} Pop$	0.0010	(0.0008)	10000.0)	-0.00002^{***}	(0.000)	0.0017	(0.0053)	-0.0087***	(0.0018)	0.0000	(0.0007)	1,412 Ves	
	$\Delta^{(7)}_{2000} Pop$	0.0011^{***}	(0.0003)				-0.0027	(0.0050)	-0.0084^{***}	(0.0018)	0.0000	(0.0007)	1,412 Ves	
	$\Delta^{(6)}_{2000} Pop$						-0.0103^{*}	(0.0053)	-0.0068***	(0.0016)	0.0001	(0.0007)	1,412 Ves	-
	$\Delta^{(5)}_{2000} Pop$	0.00073^{***}	(0.0003)		-0.00002^{***}	(0.00)			-0.0087***	(0.0018)	0.0000	(0.0007)	1,412 Ves	-
	$\Delta^{2015}_{2000} Pop$				-0.00003^{***}	(0.000)			-0.0078***	(0.0017)	-0.0000	(0.0007)	1,412 Ves	
	$\Delta^{(3)}_{2000} Pop$	0.0017^{**}	(0.0007)	-0.000)	~				-0.0085***	(0.0017)	0.0001	(0.0007)	1,412 Ves	
•	$\Delta^{(2)}_{2000} Pop$		***6000000	(0.000)	~				-0.0077***	(0.0018)	0.0000	(0.0007)	1,412 Ves	2
	$\Delta^{(1)}_{2000} Pop$	0.0011^{***}	(0.0003)						-0.0083***	(0.0018)	0.0000	(0.0007)	1,412 Ves	-
þ		d centre		n centre	uneven		orientation H.		$\ln pop \ 2000$		ln dist sea		Observations Country FE	Province FE

	(12) uneven	55.30^{***} (9.842)	7.900	(5.261) -31.17***	(4.395)									1,070	\mathbf{Yes}	Yes
akar, ables akes,	(11) d centre	-2.288^{***} (0.359)	0.0976	(0.167) 1.370***	(0.148)									1,070	$\mathbf{Y}_{\mathbf{es}}$	\mathbf{Yes}
oe Town, De sists of varie to the big la cent.	(10) new		0.0150	(0.0188) -0.139***	(0.0121)	0.675^{***}	(0.239)	0.151	(0.134)	0.0939	(0.0928)	0.711^{**}	(0.275)	1,070	$\mathbf{Y}_{\mathbf{es}}$	Yes
er Cairo, Car <i>controls</i> con ion, distance 1 1 (***) perc	(9) new		0.0275^{***}	(0.00892) - $0.128***$	(0.0116)	0.110	(0.0998)	0.128^{***}	(0.0447)	-0.103^{***}	(0.0329)	0.256^{**}	(0.111)	1,070		
ance to eith t. The set of e, precipitat), 5 (**) and	(8) uneven		10.54^{**}	(4.924) -42.89***	(3.215)	32.71	(63.99)	38.07	(30.10)	-67.45^{*}	(37.36)	150.2^{**}	(71.27)	1,412	$\mathbf{Y}_{\mathbf{es}}$	Yes
s to log dist enness indes temperatur nce at 10 (*	(7) uneven		5.111	(4.243) -43.26***	(3.244)	142.3^{***}	(30.32)	76.12^{***}	(13.91)	-9.998	(12.83)	150.3^{***}	(31.16)	1,412	$\mathbf{Y}_{\mathbf{es}}$	
In dist refer. i is the unev ure, average tote significa	(6) d centre		-0.141	(0.144) 1.816^{***}	(0.105)	-1.380	(1.836)	-0.513	(1.097)	3.367^{***}	(1.222)	-8.681***	(2.660)	1,412	\mathbf{Yes}	Yes
after 1950. <i>I</i> intre; <i>unever</i> im temperat ble Stars den	(5) d centre		-0.149	(0.129) 1_{850***}	(0.101)	-5.572***	(0.971)	-2.694^{***}	(0.495)	0.0579	(0.422)	-9.633^{***}	(1.141)	1,412	\mathbf{Yes}	
es founded a n the city ce ure, maximu laria. Variah	(4) d centre		-0.0232	(0.0694) 1.780***	(0.0992)	0.385	(0.310)							1,412		
ator for citi ad density i m temperat ions for ma	(3) d centre		-0.145^{**}	(0.0733) 1.702***	(0.0966)							-2.876^{***}	(0.545)	1,412		
⁰ is an indic measures re ion, minimu matic condi	(2) d centre		-0.0617	(0.0701) 1.782***	(0.0980)					-0.595^{***}	(0.200)			1,412		
age. <i>new</i> ₁₉₅ ea. <i>d centre</i> dness, elevat oast and cli	(1) d centre		0.0191	(0.0696) 1.795***	(0.0983)			-0.386^{**}	(0.164)	r				1,412		
Table 5: First si Djibouti or the s measuring rugge distance to the c		new_{1950}	ln dist sea	In non prowth		ln dist Cap		ln dist Dhak		$\ln \operatorname{dist} \operatorname{Dji}$		ln Dist Cai		Observations	Controls	Country fe

Table 6: 2SLS using age as instrument. d centre measures road density in the city centre; uneven is the unevenness index. The First stage F statistic reports the F test of the excluded instrument, which is an indicator for cities that were founded after 1950. The set of controls consists of variables measuring ruggedness, elevation, minimum temperature, maximum temperature, average temperature, precipitation, distance to the big lakes, distance to the coast and climatic conditions for malaria. Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

	$\Delta^{(1)}_{2000} Pop$	$\Delta^{(2)}_{2000} Pop$	$\Delta^{(3)}_{2000} Pop$	$\Delta^{(4)}_{2000} Pop$	$(5) \ \Delta^{2015}_{2000} Pop$	$\Delta^{(6)}_{2000} Pop$	$(7) \ \Delta^{2015}_{2000} Pop$	$\Delta^{(8)}_{2000} Pop$
d centre	0.00564^{**} (0.00225)		0.00635^{***} (0.00172)		0.00947^{**} (0.00433)		0.00598^{**} (0.00157)	
uneven	~	-0.000212^{***}	~	-0.000266*** (0.0000)	~	-0.000233^{***}	~	-0.000248^{***}
ln dist sea	0.00148^{**}	0.00104	0.000156	3.99e-05	0.00214	0.00233	0.000570	(0.00311)
ln pop 2000	(0.000122) - 0.0164^{***} (0.00481)	(0.000139) -0.0153*** (0.00426)	(0.00104) - 0.0180^{***} (0.00433)	(0.00183*** -0.0183*** (0.00417)	(0.00252*** -0.0252*** (0.00856)	(0.00403) -0.0171*** (0.00403)	(0.00103^{+4}) -0.0193*** (0.00394)	-0.0188*** -0.0188*** (0.00367)
Observations	1,070	1,070	1,070	1,070	1,070	1,070	1,070	1,070
Controls			IGS	IGS	\mathbf{Yes}	Yes	Yes	${ m Yes}$
First stage F	14.19	15.77	44.80	34.00	6.03	15.61	40.62	31.56

Table 7: IV using distances as instrument. d centre measures road density in the city centre,; uneven is the unevenness index. The First stage F test statistic reports the multivariate F test of the excluded instruments. The set of *controls* consists of variables measuring ruggedness, elevation, minimum temperature, maximum temperature, average temperature, precipitation, distance to the big lakes, distance to the coast and climatic conditions for malaria. Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

	(1)	(2)	(3)	(4)	(5)	(9)	(2)	(8)
	$\Delta^{20\dot{1}5}_{2000}Pop$	$\Delta^{2015}_{2000}Pop$	$\Delta^{2015}_{2000} Pop$	$\Delta^{20\dot{1}5}_{2000}Pop$	$\Delta^{2015}_{2000}Pop$	$\Delta^{2015}_{2000}Pop$	$\Delta^{20\dot{1}5}_{2000}Pop$	$\Delta^{20\dot{1}5}_{2000}Pop$
d centre	0.00228^{**}		0.00358^{*}		0.00361^{***}		0.00251	
	(0.000888)		(0.00209)		(0.00114)		(0.00198)	
uneven		-0.0000		-0.00019^{*}	e.	-0.0000*		-0.0001
		(0.0000)		(0.0000)		(0.000)		(0.000)
ln dist sea	0.00113^{**}	0.000919^{*}	9.38e-05	-0.000357	0.00113	0.00112	0.000491	0.00142
	(0.000494)	(0.000522)	(0.000682)	(0.000838)	(0.00127)	(0.00130)	(0.00136)	(0.00159)
$\ln pop 2000$	-0.00893^{***}	-0.00718^{***}	-0.0127^{***}	-0.0148^{***}	-0.0130^{***}	-0.00915^{***}	-0.0124^{***}	-0.0128^{***}
	(0.00253)	(0.00257)	(0.00449)	(0.00509)	(0.00309)	(0.00257)	(0.00416)	(0.00396)
Observations	1,412	1,412	1,412	1,412	1,412	1,412	1,412	1,412
R-squared	0.052	0.045	0.096	-0.198	0.035	0.078	0.177	0.102
Country fe			$\mathbf{Y}_{\mathbf{es}}$	Y_{es}			Y_{es}	${ m Yes}$
Controls					\mathbf{Yes}	$\mathbf{Y}_{\mathbf{es}}$	${ m Yes}$	${ m Yes}$
First stage F	46.76	19.45	4.34	2.57	24.55	9.98	3.89	2.25

Table 8: Robustness checks I. Each column of this table reproduces the results from another table and illustrates the robustness of our main results when changing the sample. We focus on our two main independent variables: d center (in columns (1), (2) and (3)) and uneven (columns (4), (5) and (6)). For the two variables we report the main coefficient using either OLS (columns (1) and (4) for each variable, respectively), age IV (columns (2) and (5)) and distance IV (columns (3) and (6)). Panel A uses our baseline results and thus displays the reproduction of the OLS results in col (1) (which is also col (1) in table 4) and col (4) (also col (2) table 4); age IV in col (2) (also col (3) table 6) and col (5) (also col (4) in table 6); and distance IV in col (3) (also col (3) table 7) and col (6) (also col (4) in table 7). Panel B includes measures of Market Access. Panel C replaces country fixed effects by province fixed effects. Panel D includes capital cities in the sample. Panel E includes cities from Nigeria. Panel E uses road measure at the city level (as opposed to centre only).Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

		(1) $\Delta^{2015}_{2000} Pop$	(2) $\Delta^{2015}_{2000} Pop$	$\begin{array}{c} (3) \\ \Delta^{2015}_{2000} Pop \end{array}$	(4) $\Delta^{2015}_{2000} Pop$	(5) $\Delta^{2015}_{2000} Pop$	(6) $\Delta^{2015}_{2000} Pop$
Panel A: Baseline	d centre uneven Obs. 1st stage F:	0.00110*** (0.000267) 1.412	$\begin{array}{c} 0.00635^{***}\\ (0.00172)\\ 1.070\\ 44.80\end{array}$	$\begin{array}{c} 0.00358^{*}\\ (0.00209)\\ 1.412\\ 3.872\end{array}$	-3.30e-05*** (8.00e-06) 1.412	-0.000266*** (7.00e-05) 1.070 34	$\begin{array}{c} -0.000190^{*} \\ (9.91e\text{-}05) \\ 1.412 \\ 2.630 \end{array}$
Panel B: Mark. Acc.	d centre uneven Obs. 1st stage F:	$\begin{array}{c} 0.00106^{***}\\ (0.000262)\\ 1.412\end{array}$	$\begin{array}{c} 0.00636^{***}\\ (0.00177)\\ 1.070\\ 42.05 \end{array}$	$\begin{array}{c} 0.00334\\ (0.00220)\\ 1.412\\ 3.225\end{array}$	-3.23e-05*** (7.95e-06) 1.412	$\begin{array}{c} -0.000268^{***} \\ (7.28e-05) \\ 1.070 \\ 31.52 \end{array}$	$\begin{array}{c} -0.000190^{*} \\ (9.91\text{e-}05) \\ 1.412 \\ 2.630 \end{array}$
Panel C: Province	d centre uneven Obs. 1st stage F:	0.00126*** (0.000319) 1.412	$\begin{array}{c} 0.00549^{***} \\ (0.00127) \end{array}$ $\begin{array}{c} 1.070 \\ 37.08 \end{array}$	$\begin{array}{c} 0.000581 \\ (0.00380) \end{array}$ $\begin{array}{c} 1.412 \\ 1.127 \end{array}$	-3.23e-05*** (7.95e-06) 1.412	-0.000250*** (6.23e-05) 1.070 18.99	$\begin{array}{c} 2.36\text{e-}06\\ (0.000112)\\ 1.412\\ 0.982\end{array}$
Panel D: Capitals	d centre uneven Obs. 1st stage F:	$\begin{array}{c} 0.00117^{***} \\ (0.000271) \\ 1.450 \end{array}$	$\begin{array}{c} 0.00653^{***}\\ (0.00176)\\ 1.107\\ 44.19\end{array}$	$\begin{array}{c} 0.00513^{**}\\ (0.00253)\\ 1.450\\ 4.191 \end{array}$	-3.22e-05*** (8.00e-06) 1.450	$\begin{array}{c} -0.000278^{***} \\ (7.40e-05) \\ 1.107 \\ 32.88 \end{array}$	-0.000281^{**} (0.000129) 1.450 1.976
Panel E: Nigeria	d centre uneven Obs. 1st stage F:	0.00106*** (0.000225) 1.817	$\begin{array}{c} 0.00475^{***} \\ (0.00119) \end{array}$ $\begin{array}{c} 1.428 \\ 72.53 \end{array}$	$\begin{array}{c} 0.00326^{*}\\ (0.00187)\\ 1.817\\ 5.162\end{array}$	-2.74e-05*** (6.57e-06) 1.817	-0.000259*** (7.01e-05) 1.428 28.83	$\begin{array}{c} -0.000175^{**} \\ (8.75e-05) \\ 1.817 \\ 4.408 \end{array}$
Panel F: City-wide	d city Obs. 1st stage F:	$ 0.00182^{***} \\ (0.000449) \\ 1.412 $	$ \overline{\begin{array}{c} 0.0265^{***} \\ (0.00827) \\ 1.070 \\ 16.61 \end{array} $	$\begin{array}{c} 0.0206^{*} \\ (0.0117) \\ 1.412 \\ 1.068 \end{array}$			

Table 9: Robustness checks II. Panels A and D present the coefficient of interest from OLS estimation for d centre and uneven, respectively. Panels B and E report the respective age IV results instead; and panels C and F the distance IV. Colmun (1) reproduces baseline results, while outcome is changed for: 1975-2015 growth in column (2), 2015 population level in column (3), 2000-2015 growth as in GPW data in column (4) and in Africapolis data in column (5). tars denote significance at 10 (*), 5 (**) and 1 (***) percent.

		(1) Baseline	(2) Λ_{1075}^{2015} Pop	(3) $\log \text{Pop}^{2015}$	(4) GPWv4 data	(5) Africapolis data
		Daseille	-19751 op	108 1 09	GI WYI data	
Panel A:	d centre	0.00110^{***}	0.00159^{***}	0.0154^{***}	0.000353^{***}	2.18e-05
015	Obs.	1.412	1.339	1.412	1.412	1.003
Panel B: age IV	d centre	0.00635^{***} (0.00172)	0.00577^{***} (0.000817)	0.0888^{***} (0.0241)	0.00123^{**} (0.000581)	0.000188 (0.00230)
	Obs.	1.07	1.024	1.07	1.07	1.003
	1st stage F:	44.80	65.74	44.80	43.71	6.357
Panel C:	d centre	0.00358*	0.00126	0.0501*	-0.000115	0.00211
aist. IV	Obs	(0.00209) 1 419	(0.00145) 1 330	(0.0295) 1 412	(0.000875) 1 419	(0.00138) 1.003
	1st stage F·	4.345	2.952	4.345	2.676	4.097
	150 500go 1 .	1.0 10	2.002	1.010	2.010	1.001
Panel D:	uneven	-3.30e-05***	-4.02e-05***	-0.000463***	-8.54e-06***	-1.25e-05*
OLS		(8.00e-06)	(4.46e-06)	(0.000112)	(3.11e-06)	(6.89e-06)
	Obs.	1.412	1.339	1.412	1.412	1.003
Panel E:	uneven	-0.000266***	-0.000238***	-0.00372***	-4.95e-05**	-7.60e-06
$age \ IV$		(7.00e-05)	(3.39e-05)	(0.000980)	(2.36e-05)	(9.35e-05)
	Obs.	1.07	1.024	1.07	1.07	1.003
	1st stage F:	34	55.20	34	31.28	5.064
Panel F:	uneven	-0.000190*	-7.61e-05	-0.00265*	9.92e-06	-0.000154**
dist. IV		(9.91e-05)	(6.42e-05)	(0.00139)	(5.10e-05)	(7.62e-05)
	Obs.	1.412	1.339	1.412	1.412	1.003
	1st stage F:	2.569	1.733	2.569	1.244	2.048



Figure 1: This figure presents two cities with stylised road layouts. Both cities have the same road density in the centre by construction. The city on the right has a shorter distance from a random point to the nearest road, a property we call 'evenness'.



Figure 2: We identify 2,779 cities from 40 countries.



Figure 3: City extent and centre are defined using nightlight luminosity (in late - 2008-2012 - and early - 1994-1996 - period, respectively).



Figure 4: Our measure of road layout is based on OSM roads in a 1km-radius circle around the city centre.



Figure 5: Evenness is computed from 12 within-centre points.



Figure 6: The edge bearings histograms (right) summarise the orientation of city centre edges (left). The central panel is a polar representation of the histogram, illustrating the construction of our measure. The first city (top) is Umm Ruwaba (Sudan) and has a typically high Herfindahl index (0.48). Below is Mohale's Hoek city (Lesotho), which is representative of the lowest values in our sample (0.05)

Data Appendix

A Data Sources

A.1 Cities, location and boundaries

• Lights: NOAA's DMSP-OLS sattleites record images of earth at night (every day, at 20:00 to 21:30 local). Composite images are built combining daily series in order to remove noises, cloud effects and temporary lights. The data is a 30x30 arc-second (around 1sq.km.) grid with pixels taking a value between 0 (no lights) and 63 (top coded value). Data is available from 1992 to 2013. We use the 2008-2012 period to build city boundaries and the 1994-1996 to find city centres.

A.2 Road Layout

- Road Layout: Open Street Map is a volunteered geographic information project (which information comes from over 2 millions users) mapping, among others, cities' streets. Our analysis is based on the shapefiles from GeoFabrik.
- Basic network analysis: we use G. Boeing's OSMNX python package to perform simple network analysis of road layouts (counting "nodes", i.e. intersections, measuring street network orientation, etc.)

A.3 Population

• European Commission, Global Human Settlement:

GHS-Population Grid (LDS) (2016) is a 250m x 250m grid depicting the distribution of population for years 1975, 1990, 2000 and 2015. The data provided the CIESIN, Gridded Population v4, was disaggregated based on the distribution and density of built-up mapped by the GHS-Built-up Grid (that detects built-up presence from Landsat image collections: GLS1975, GLS1990, GLS2000, and ad-hoc Landsat 8 collection 2013/2014)

• Oak Ridge National Laboratory, LandScan:

LandScan High Resolution global Population Data Set provides another population distribution at a 30 x 30 degree scale grid. We accessed the dataset for 2012 only.

A.4 Others

- Coasts and Lakes: Natural Earth is a public domain map dataset that we use to clip coasts and lakes when necessary (scale: 1:10m).
- Borders: we define countries using the borders from the Global Administrative Areas dataset GADM (v2)
- Nightlights growth: we test the robustness of our results using nighttime luminosity as the dependant variable; when this is the case, we also rely on NOAA's DMSP-OLS imagery.

B Construction of the Variables of interest

B.1 Defining cities

In order to define a consistent sample of cities for many countries in Sub-Saharan Africa, we rely entirely on nightlight activity. We follow a two-steps procedure in which we first define *city boundaries* and then find the *city centre*. This procedure identifies 2,779 cities in Sub-Saharan Africa (excluding South Africa). The remainder of this sub-section details the GIS procedure.

B.1.1 Cities boundaries

In a first step, we use nightlights picture to define city boundaries. Specifically:

1. Contiguous lit areas are aggregated to form *union* of pixels;

Step 1 is based on nightlights activities during the 2008-2012 period captured as by satellite imagery from DMSP-OLS (cf. A.1). Any pixel being lit (i.e. value being greater than or equal to 1) at least twice over the period is kept at this stage. Unions so defined are then being "clipped" to coasts (sea and lakes sometimes reflect lights) and borders so as to add natural and cultural borders that divide them.

This procedure follows Fetzer et al. (2016). When working with countries from the Gulf of Guinea, we directly rely on their extended sample. This is because Fetzer et al. manually corrected for gas flares issues that are numerous in this area of the world. Note that, for the same reason, we exclude Nigeria from our main specifications.

At this stage, we observe 6,382 union in Sub-Saharan Africa (excluding South Africa).

B.1.2 City centre

In order to determine the location of each city centre, and in the absence of complete dataset of city-centres' coordinates, we rely on nightlights. Our measure of city centres follow three steps:

- 1. First, we compute city-specific maximum value of light and keep only pixels of this value;
- 2. Second, we aggregate contiguous pixels into unions and keep only the largest union(s);
- 3. Third, we take the geographic (within-)centroid of these unions.

This definition ensures that only one centre is defined in each city (referring to the monocentric assumption we make); and that this centre falls within one of the brightest pixels of the city.

Because we are interested in historical centres, we rely on early lights information rather than most recent images. We average pixel value over the three years 1994, 1995 and 1996 (earliest years for which we can build a 3-year-long panel relying on only one satellite). Another advantage of taking early observations is that top-coding of the data is less likely to be an issue (few pixels actually reach the value of 63 in mid-90's Africa).

From the 6,382 *unions* defined in the previous step, only 2,779 displayed sufficient nightlights over the 1994-1996 period. These 2,779 constitutes our sample of *cities*.

B.2 Road Layout

We construct road layout information by extracting data from *Open Street Map* (OSM). OSM (cf. A.2) is a volunteered geographic information project which information comes from over 2 millions users. Anyone can contribute to the construction of this database. Although not the most complete, OSM has the great advantage of being publicly available.

Technically, we used two alternative ways to gather road network information for each city. First, OSM dataset is digitised by *GeoFabrik* from which we take roads information for 40 African countries. In particular, at the time we impored these data, Burundi and Somalia OSM maps were not available. The OSMNN roads are often (although not always) classified between primary, secondary and residential roads. Data potential extent is the totality of roads within each country; but the actual extent is what people contributed to at the date of our download (last: December 2016).

In a second step, we use OSMNX python package to extract basic network descriptive statistics such as number of nodes (i.e. crossroads), network orientation, etc. This package was developed by G. Boeing (cf. A.2) for urban planning studies and is open-access. (Last download: July 2018.)

The two data extraction sometimes conflict with each other (some road may be missing in one but not the other, etc.) This may come from the updates that are made to OSM dataset between our two extractions. In any case, we restrict our analysis to cities for which both data exist. Our results are robust to using one or the other dataset (e.g. the correlation of total road length in centres is 0.85) as well as to the restriction on cities having both information.

For any city, we gather two types of information regarding the road network: centre road layout and whole city layout. We call city centre a circle of 1,000m around the centre point (as defined in B.1.2). Note that we "clip out" any part of the circle that would cross the coastline or a border (so it can happen that a centre is not a full circle) or that would fall outside the city boundary (ensuring that the area of the city centre is always less than or equal to the city total area).

B.2.1 About Open Street Map Precision

Being user-generated, the quality and coverage of the OSM data set is often questioned. There exists a few studies comparing OSM data to other (authoritative) datasets and generally pointing to the high and increasing quality of OSM. Most of these studies nevertheless focus on developed countries. G. Boeing (2017) present a quick survey on the topic. Recently, Barrington-Leigh and Millard-Ball (2017) however show that, as of 2016, OSM was already of high coverage in many countries of the world, including in Africa. Although the region does not rank highest in the world, a large number of sub-Saharan countries score really high degrees of completeness.

In this appendix, we re-estimate our main 6 models on the sub-sample of countries with more than 80% coverage. Barrington-Leigh and Millard-Ball set this threshold at the country level with two important consequences. First, there is a strong negative correlation between country size and completeness. Which implies that the sub-sample is made of smaller countries. Indeed, we observe only 384 cities from 16 countries). Second, qualitative evidence shows that cities are more likely to be covered by OSM than smaller, rural places. Using a country-level threshold here can thus be thought of as a lower bound of the OSM completeness. Table 10 compares our 6 baseline estimates in the whole sample and in the high-quality sample only. Again, qualitative message remains: signs of coefficient do not change and although magnitudes differ slightly, significance remains similar. Finally, not that instrumental variables approach we use in the paper corrects measurement bias arising from classical errors in OSM.

B.3 Population

We extract population information from the European Commission GHS-Population Grid (GHSP, cf. A.3). GHS is a geo-coded 250m by 250m grid providing estimates of the number of inhabitants, based dis-aggregation and interpolations from national censuses. The GHS population uses information from the presence of built-up cover as derived by the GHS-Built-up Grid, in order to spatially allocate population data. The population estimates are provided for four target years: 1975, 1990, 2000 and 2015. Technically, we extract the total number of inhabitants whose coordinates fall within the boundaries of our cities sample.

There exists alternative data set one may use to compute population at the city level. In particular, the GHS dataset is partly built upon the Gridded Population of the World (GPW) in its 4th version, which provides population mapped into a regular grid. We however think that the GHS data improves on this initial source by its use of built-up presence day-time satellite imagery (derived from Landsat). Because the GHS data only goes back to 1975, we also take advantage of the OECD Africapolis (2019) dataset which collects official population sources into a unified dataset. Even if our main interest in this data was its ability to distinguish cities of a certain size in 1950, we also consider it as a potential source for outcome measurement. We note that all of these datasets give population information that correlates fairly positively. The coefficient of correlation for log population in the year 2000 from GHS with log population from Africapolis is 0.72; and is 0.56 with log population from GPW version 4. This correlation is particularly high for larger cities, while the disagreement comes mainly from smaller cities and towns. While most of our results in this paper use the GHS population as outcome variable, we show that the main results are qualitatively robust to using these alternative sources.

C Sample Selection

The construction of our sample implies various steps where choices had to be done. We summarise these steps here, making clear when and what decision were taken.

When using nightlights over 2008-2012, we find 6,382 "unions" in Sub-Saharan Africa. Of these unions, only 2,779 did emit enough light in 1994-1996 for us to define a city centre. We restrict the sample to those units. From these, we keep only places that met the two size criteria (cf. C.1). First, we require the total city area to be greater than 2 square kilometres; and second, we require that cities have more than 2,000 inhabitants as measured in 2015. We add such steps in order to discard very small places for which the grid would not play any importance. This leaves us with mainly cities and towns. Some of those may still be small, but we think are large enough for the grid in the centre to affect interaction of people. With these two criteria, we reduce our sample to 2,204 observations.⁹

We expect capital cities to follow unique development paths that are influenced by factors of politics and international aid. Since we are trying to identify broad economic trends free from such intervention, we drop capital cities throughout, and focus on the other towns and cities. This represents a further reduction to 2,164 observations. We provide robustness analysis of this decision in the results.

We also exclude several countries, for different reasons listed in C.2. First, we excluded South Africa from the start, since we consider it too different from the rest of the sample. Second, Burundi and Somalia data were not available and we thus drop their cities. At this stage we observe 2,141 cities. Third, we exclude Nigeria and Madagascar. Nigeria is a large country and is also a place where measurement from imagery could be noisy. We fear that including Nigeria would have our result depends to a large degree on one noisy set of observations. We do however verify that results are broadly similar when we include Nigeria.¹⁰ We also exclude Madagascar throughout this paper, since the reasoning of our instrument does not apply there given its geography and history.

Finally, as explained in C.3 we exclude cities for which we did not find roads when extracting OSM data. These represent around 13.4% of the 2,141 cities sample. In doing so, we obtain our total sample of 1,850 cities.

C.1 Minimum size restriction

In order to ensure our results are not driven by the large number of very small places, we exclude from our analysis any observation that does not meet one of the two following criteria:

- 1. Minimum area: cities must be at least greater than 2sq.km.
- Minimum population: cities must have at least 2,000 inhabitants in 2015, as observed in GHS population¹¹.

These restrictions set boundaries so low that only fairly rural places get excluded. There are still quite small towns left in the main dataset after this exclusion.

C.2 Excluded countries and Capital cities

We focus on Sub-Saharan Africa, from which we excluded South Africa (because of its specific history) as well as non-continental areas.

 $^{^{9}}$ We use the population estimates of 2015 in order to increase the extent of our definition of a city. We also verify that our main results are robust to the inclusion of these small places.

¹⁰In fact, typically statistical significance increases with the larger sample that includes Nigeria.

 $^{^{11}\}mathrm{We}$ also exclude 5 cities that had 0 inhabitants in 2000, according to GHS data.

In particular, the exclusion of Madagascar follows from the incompatibility of its geography with our historical IV.

We also exclude Nigeria from our main specification because we want to ensure it does not drive the results alone (Nigeria does represent a big share of the total potential sample). Moreover, the presence of gas flares in the Gulf of Guinea makes the delimitation of cities from nightlights imagery particularly noisy in this country; although our data-set was manually corrected for this.

In the case of these last two countries, our choice was guided by specifications issues; we still include them in the following tables as one could be interested in the robustness of our results. In fact, our main results are also robust to the inclusion of one or both of these countries in the sample.

As mentionned, when using OSM data has exported by GeoFabrik (cf. B.2) we must also exclude Burundi and Somalia, for which there is not data available. For consistency we always exclude these two countries from our specifications.

Finally, we also exclude the capital city of each country. The reason is that capital cities in Africa often have unique economic developments and histories, and may deviate from general trends observed elsewhere. But here again, the inclusion of these few cities does not change the qualitative results.

C.3 Missing observations

We then extract OSM road networks for our 2,141 cities (i.e. the sample including Madagascar and Nigeria). We are nevertheless unable to observe roads (in at least one of our extraction) in 291 of these cities.

These missing information could be caused either by an error in our definition of cities (i.e. declaring an area to be a city while it is not) or by the incompleteness of OSM data. A quick comparison allows us to say that the missing observations represent 13.59% of the total sample and affect some countries in particular (like Ghana, Sierra Leone, Sudan, Swaziland, etc.), cf. Table 11. Table 12 also shows that the cities with no road observation are, on average, smaller and less populated than the average city in the rest of the sample.

References

- Barrington-Leigh, Christopher, and Millard-Ball, Adam, (2017). "The world's usergenerated road map is more than 80% complete", PLOS ONE, 12(8), 1-20.
- Boeing, Geoff. (2017). OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. Computers Environment and Urban Systems. 65. 126-139. 10.1016/j.compenvurbsys.2017.05.004.

Table 10: OSM Quality. Panel A is our main baseline results as in Table 8 Panel B includes measures of Market Access. Panel B reports the same estimates on the sub-Sample of countries with high OSM completeness. Stars denote significance at 10 (*), 5 (**) and 1 (***) percent.

		(1) $\Delta^{2015}_{2000} Pop$	(2) $\Delta^{2015}_{2000} Pop$	$\begin{array}{c} (3) \\ \Delta^{2015}_{2000} Pop \end{array}$	$\begin{array}{c} (4) \\ \Delta^{2015}_{2000} Pop \end{array}$	(5) $\Delta^{2015}_{2000} Pop$	(6) $\Delta^{2015}_{2000} Pop$
Panel A: Baseline	d centre uneven	$\begin{array}{c} 0.00110^{***} \\ (0.000267) \end{array}$	$\begin{array}{c} 0.00635^{***} \\ (0.00172) \end{array}$	0.00358^{*} (0.00209)	-3.30e-05***	-0.000266***	-0.000190*
	Obs. 1st stage F:	1.412	$\begin{array}{c} 1.070\\ 44.80\end{array}$	$1.412 \\ 3.872$	(8.00e-06) 1.412	(7.00e-05) 1.070 34	$(9.91e-05) \\ 1.412 \\ 2.630$
Panel B: OSM HQ	d centre	0.00201^{***} (0.000740)	0.00506^{***} (0.00185)	0.00674 (0.00244)			
0,0112,024	uneven	(0.0001-0)	(0.00200)	(0.00111)	-3.44e-05* (1.81e-05)	-0.000278*** (0.000110)	-0.000350^{***} (0.000130)
	Obs. 1st stage F:	384	$301 \\ 35.91$	$\frac{384}{5.5}$	384	$301 \\ 13.95$	384 3.55

Table 11: Missing OSM Data

	OSM	Missing
Angola	80.49	19.51
Benin	84.38	15.63
Botswana	87.88	12.12
Burkina Faso	96.43	3.57
Cameroon	95.83	4.17
Central African Republic	90.91	9.09
Chad	85.71	14.29
Congo, Dem Republic of	98.11	1.89
Congo, Republic of	88.24	11.76
Côte d'Ivoire	64.53	35.47
Djibouti	100.00	0.00
Equatorial Guinea	100.00	0.00
Eritrea	100.00	0.00
Ethiopia	99.12	0.88
Gabon	89.29	10.71
Gambia	100.00	0.00
Ghana	79.28	20.72
Guinea	95.83	4.17
Kenya	89.89	10.11
Lesotho	100.00	0.00
Liberia	100.00	0.00
Madagascar	97.06	2.94
Malawi	95.00	5.00
Mali	90.91	9.09
Mauritania	100.00	0.00
Mozambique	93.94	6.06
Namibia	91.30	8.70
Niger	85.00	15.00
Nigeria	88.96	11.04
Rwanda	100.00	0.00
Senegal	98.18	1.82
Sierra Leone	85.71	14.29
South Sudan	100.00	0.00
Sudan	76.47	23.53
Swaziland	77.78	22.22
Tanzania, United Rep of	94.34	5.66
Togo	95.65	4.35
Uganda	92.68	7.32
Zambia	88.89	11.11
Zimbabwe	80.52	19.48
Total	86.41	13.59

	OS	SM	Mis	sing	Diffe	erence
	Mean	Sd	Mean	Sd	Diff.	\mathbf{t}
Ln.Area (sq.km)	3.77	(1.10)	3.39	(0.92)	0.38**	*(5.90)
GHS Population Data						
Ln.Pop. 2000	10.38	(1.37)	9.05	(1.30)	1.32^{**}	*(16.03)
Ln.Pop. 2015	10.68	(1.37)	9.40	(1.06)	1.28^{**}	*(18.39)
Av. Growth Rate $(2000-15)$	2.04	(2.99)	2.30	(4.92)	-0.26	(-0.87)
Observations	1850		291		2141	

Table 12: Samples Comparison