Growth in the African Urban Hierarchy

J. Vernon Henderson*
Cong Peng†
Anthony J. Venables‡

July 15, 2022

Abstract

What regularities emerge as countries develop a pattern of built settlement? This paper uses satellite data to trace the evolution of some 50,000 built areas in Sub-Saharan Africa between 1975 to 2014, a period in which total built area increased by a factor of 2.4 due to growth and merger of settlements and the birth of new settlements. The median growth rate of settlements in the smallest initial size bin was twice that of settlements in the largest (of five) bins, rejecting Gibrat's law. Settlements of different size generally specialise in different activities, and we model this by supposing three settlement types: agricultural, agro-processing, and manufacturing/ service based. In the presence of many dispersed agricultural settlements the model predicts regular spacing of fewer and larger agro-processing settlements, and few large manufacturing/ service settlements. This pattern of spacing arises as settlements of the same type are in a competitive relationship with each other (competing for inputs and for sales of output), while settlements of different types are in a complementary relationship (because of input-output relationships). We confirm this empirically by grouping settlements into three size classes and regressing each settlement's growth on its proximity to settlements in the same and other size classes. A fast growing neighbour of similar type reduces growth, while proximity to fast growing settlements of a different type increases growth.

Keywords: urban development, urban hierarchy, built settlement, Africa

JEL Codes: R1, O1

*J.V.Henderson@lse.ac.uk, The London School of Economics
†cpeng@hks.harvard.edu, Harvard Kennedy School of Government
‡tony.venables@economics.ox.ac.uk, University of Manchester and Monash University

1 Introduction

The paper studies the evolution of settlements in the 43 countries of sub-Saharan Africa from 1975 until 2014, excluding South Africa due to the role of apartheid in affecting urban geography. This was a period of enormous change in which total built area increased by a factor of 2.4 as new settlements developed and existing settlements grew and sometimes merged. Our empirical work is based on satellite imaging from which we identify built areas as small as 900 sq meters. There were about 47,500 such areas in 1975, rising to about 111,500 in 2014. We refer to these built areas as settlements throughout the paper. The growth factor of the median 1975 settlement was 2.13 but, because of entry of small settlements, median settlement size was approximately constant through the period.

Our principal interest is the geography of settlements and how this is shaped by patterns of competition and complementarity between settlements of different sizes and types. In the main section of the paper, we characterize an urban hierarchy and investigate the spatial relationships between types of settlement, developing an analytical model which we then estimate. We find that settlements of the same type are in competition, such that increased growth of one detracts from the growth of neighbours of the same type. However increased growth of a settlement benefits neighbours of other types, so settlements of different types complement each other. This is the first evidence of these competition versus complementarity effects set in space that we know of.

We proceed in several stages. First we describe the data and some of the technical issues involved in its analysis (Section 2). Then before turning turning to the main Sections 5 and 6, in Sections 3 and 4, we look at sub-Saharan Africa from the point of view of the literature on urban size distributions and growth which includes papers by Gabaix (1999), Eeckhout (2004), Harris Dobkins and Ioannides (2001), Black and Henderson (2003), Duranton (2007), Desmet and Rappaport (2017), Bosker and Buringh (2017), de Bellefon et al. (2020), Jedwab and Storeygard (2021), and Düben and Krause (2021) to name a few. Different papers use different data sources, initially most using population data, although more recently, some papers focus on the built environment and use of satellite data, as we do. We use satellite data on built area largely because of the low quality of sub-national population data for much of sub Saharan Africa. Our intent in the first part of the paper is to provide a sense of the data and some basic facts about African urban growth. We show that findings for Africa tend to differ from those found for the USA and Europe today, likely because of the low level of economic development and rapid urbanization of the sub-continent.

In Section 3, we analyze the statistical distribution of settlements of different sizes. We review Zipf's law and examine log-normality of the city size distribution. In Section 4, we give an overview of the growth in built area between 1975 and 2014, looking both at growth of 1975 settlements and at births during the period. There is enormous upward mobility, but growth rates of settlements and the variance in that growth rate are lower for settlements which are relatively large initially. As such, Gibrat's Law is rejected. Births since 1975 dominate the 2014 number of settlements, but births show little evidence of growth from 1975 to 1990 or 1990 to 2000. Larger 2014 settlements were all present in 1975 and many grew both internally and by overrunning smaller neighbours, with evidence of substantial churning.

After these descriptive results we turn to the main body of the paper. We were initially inspired by recent papers on growth in the shadows of urban giants, such as Cuberes et al. (2021), Beltràn Tapia et al. (2017), and Bosker and Buringh (2017). These papers explore whether smaller cities benefit or suffer from being near larger cities, given opposing competition and market potential effects, with Cuberes et al. (2021) arguing that which effect dominates the other changes over time in the USA. In exploring this notion for Africa for the period 1975-2014, we concluded that something much more fundamental was at work, that invokes notions of an urban hierarchy, as proposed in Lösch (1954) with modeling in Fujita et al. (1999) and Tabuchi and Thisse (2011), drawing on the new economic geography literature (Krugman, 1991).

In section 5, we develop a theoretical model of an urban hierarchy in which each settlement can perform three possible functions. Some settlements are agricultural, using land and labour to produce output which is costly to transport. Some are 'market/agro-processing' towns which use agricultural inputs and labour to produce consumption goods that have lower transport costs than unprocessed agricultural output. The final type are manufacturing/service cities which use labour to produce easy-to-ship final or intermediate goods. Urban labour is perfectly mobile, and settlements can develop at any place on the geographical space. Starting from a uniform distribution of population, a structure of settlements evolves with many small agricultural settlements, a smaller number of larger and approximately equally spaced market/agro-processing towns; and a still smaller number of large, and approximately equally spaced, manufacturing/ service cities. This spatial pattern emerges as settlements of the same type are in a competitive relationship with each other whereby a growth shock to one reduces growth in near neighbours of the same type, while settlements of different types are in a complementary relationship whereby a growth shock in one increases growth in a near neighbor of a different type.

We investigate these patterns of competition and complementarity in the data. In Section 6, our main results are based on classifying settlements into three size classes, where the relative distribution across the three types is country specific, using a statistical criterion. As predicted by the model, settlements generally grow more slowly if close to neighbours in the same size class that are fast growing, and, grow faster if close to fast growing settlements in one of the other size classes.

2 Data

Our primary data are based on the European Union's Global Human Settlements built cover data set, GHS-BUILT, that defines built surface derived from Landsat 30-meter resolution satellite data for different dates: 1975, 1990, 2000, and 2014. There is a companion data set on the spatial distribution of population from Gridded Population of the World [GPWv4], but we work with built cover in preference to population data for several reasons. First is accuracy. The GHS population data allocates administrative unit census population to the built cover data, smearing population into commercial or industrial buildings and roads (impermeable surface), as well as residential buildings. Smearing across areas of built cover is a crude procedure to determine where people actually live. Accuracy is particularly poor if administrative units are large, as is typically the case in Africa. Of the 12.9 million input population polygons worldwide in GPW, 10.5 million are in the United States, so accuracy for the USA is much higher than in Africa. Equally compelling for some African countries, census population numbers are outdated and of questionable accuracy. While population numbers may be poorly and inconsistently measured across countries and time, built cover is more consistently measured. Thus, we use built cover rather than trying to ascertain where residential population lives.

In recording built cover, a 30x30 m pixel is built or not, and built area is simply the area covered by the built 30x30m pixels. In working with 30x30m resolution data, for computational purposes we aggregate to 210x210m size super-pixels, summing from 30x30 m built pixels to get the built area of the super-pixel.

The key decision, given all the built pixels in a country, concerns what comprises a settlement. Implicit is the idea that built pixels could be randomly located (rural) bits such as huts or hamlets within a country, but some subset are agglomerations or clusters that define a settlement. Settlements have high density values compared to a counterfactual: higher than the expected intensity of clustering, beyond what one would find on a "dartboard" (Ellison and Glaeser, 1999).

To proceed, we follow de Bellefon et al. (2020). As described in Appendix A, we use a smoothed surface to capture disconnected parts of built pixels within a settlement, so each super-pixel has a smoothed density from the surrounding area and itself. Super-pixels further from the own-pixel are discounted by distance up to a maximum of 2.3 km. Most super-pixels will have zero share of built area for themselves and surrounding pixels, meaning that the pixel is deemed not built. The normalized maximum is 1, in which case everything around the super-pixel and itself is built.

Details are in the Appendix.

Next, we need a density cut-off or threshold to determine what is a significantly high degree of density. For this we must pose a counterfactual. Based on a suggestion in de Bellefon et al. (2020), we use a fixed large square area divided into pixels which we treat as a hypothetical country. We randomly allocate built pixels across this area. The number of allocated built pixels and counterfactual distribution varies by country, according to the share of actual built pixels within each country's total count in 2014. Then for each counterfactual built-up density, we calculate its smoothed built-up density, as we did for the real spatial distribution. We bin the smoothed built-up in 10000 bins (many bins being 0). We repeat this process 500 times and sum up the counts in each bin to generate a stable distribution by leveraging the law of large numbers. The threshold we use for each country is the 95 percentile in the counterfactual built-up distribution. These cut-offs for each country are shown in Figure A2 in the Appendix. They range from a value close to 0 to about 0.006 for the smoothed share density threshold value. While that may seem small even at the upper end, in the Appendix we show that it makes a huge difference to what are defined as the extents of settlements.

Finally, we need to define settlement areas. For each country, we take all contiguous super-pixels with smoothed density above its threshold and agglomerate them into a shell that defines the boundaries of the settlement. For coastal settlements there is some infill for non-built super-pixels on the coast surrounded by built surface. Then for this shell, the size of the settlement is the actual built area within the shell, based on the sum of all 30x30 m built pixels within the shell. Figures on numbers and size of settlements by country are given in Appendix Table A1.

Table 1 depicts the a summary of our 2014 data for Africa as a whole. Table A2 in the Appendix shows the corresponding data for 1975. Table 1 divides the data into 10 bins of (almost) equal share of total built area. Shares and total built area (in sq. km.) in each bin are given in columns 7 and 4 respectively. Because settlements are an integer count it is not possible, especially at the upper end with large cities, to get exactly 10% in each bin. Size categories and the minimum and maximum settlement sizes in the each bin are given in columns 1-3. The largest city in the sample (1370 sq.km.) is about 52% of its bin total built area (2649 sq. km).

There are two sets of notable facts in the table. First, smaller settlements can be really small, just one 30x30m grid square of built area (size: 0.0009 sq km). In the first bin the maximum size is just 0.25 sq km of built area. Those small settlements account for 94.5% of the total 111469 settlements in the sample in 2014. Second, shell area, or the land area of settlements within their boundaries as defined in the algorithm used to characterise settlements is enormous, compared to the area of actual built pixels. In the first bin, shell area is 171 times built area; while, in bin 10, shell area is 4.9 times larger than built area. These multiples tend to fall as we move to bins with larger sizes. In the sections that follow we will look at all settlements, noting that some settlements that were tiny in 1975 do grow at remarkable rates into medium size settlements in 2014, so should not be ignored. However because we want to make sure statistical results are dominated by the volume of tiny settlements, in robustness checks, we ran our basic results on larger settlements beyond a minimal size defined below. Qualitatively, results are very similar to the full sample.

3 The Distribution of Settlements

We start by looking at the statistical (rather than geographical) distribution of the cross-section of the 111469 settlements in 2014. Do these cities follow a Pareto distribution as in Gabaix (1999) and approximate the rank size rule, or is the distribution more log normal as in Eeckhout (2004)? Since we cover the entire size distribution not just the upper tail, as in Eeckhout we expect the latter to be more likely. Indeed, in Figure A3a for 2014, a plot of rank versus size throughout the distribution show a pattern that strongly deviates from the rank size rule. OLS estimates, although they are biased, all show coefficients that deviate from the rank size coefficient of 1. All are well under 1 for every country in the sample, with an estimate for a pooled sample with country fixed effects of a very low -0.58. We repeated the exercise in A3a for the 47519 settlements in the 1975 cross-section, with a similar deviation from the rank size rule.

For a further perspective, we look at the size distribution of settlements in 1975 and in 2014 for the 1975 settlements which still exist in 2014. The Kolmogorov-Smirnov test for log-normality does not reject overall log-normality for either year, noting, however, that the test is known to be weak. In the Appendix, we show that the 2014 distribution in Figure A3b looks log normal with a few bumps in the left tail, while the 1975 distribution has a jagged left tail. [1] As usual, Gabaix's (1999) Pareto shape in just the right tail of size distributions appears to hold.

For reference in Section 6, while we estimate our model for all settlements, The Appendix Figure A3b for 1975 might suggest trying a cut-off on the left that excludes tiny settlements, where most tiny settlements in 1975 never emerge as full-fledged significant size settlements above, say, an area of built surface that is 0.011 sq km. We rerun our models in Section 6 using this cut-off in logs of -4.5 for 1975 in robustness checks.

[1]In graphing the left tail, the main issue is that there are size gaps: we are moving in logs of settlements from 1 30x30 m grid square, rendered in logs of sq km, to 2 to 3 and so on grid squares, with huge concentrations at two grid squares (about -6.3 in log points). The gaps produce the spikes and peaks in the very left tail.

4 The Evolving Pattern of Settlement

We draw out some descriptive facts about the way in which the city system has evolved, building on work, for example, in Harris Dobkins and Ioannides (2001), Black and Henderson (2003), and Desmet and Rappaport (2017). We will focus on growth of the 47519 settlements that were present in 1975. We look at the growth of settlements by size class. We show Gibrat's law, underlying Zipf's law in Gabaix (1999), is violated. Then, we look at the past growth of settlements by their final size, accounting for the birth and merger of built areas, hence showing how the evolution from 47519 settlements in 1975 to 111469 in 2014 occurs. Finally we look at the growth in urban shadows hypothesis reviewed and analyzed in Cuberes et al. (2021).

4.1 The Growth of 1975 Settlements

How did settlements of different initial sizes grow? Patterns are summarized in the "violin" Figure 1, based on 1975 share bins (Table A2) collapsed from 10 bins into 5 with a 20% share of built each, to give bigger settlement counts to upper-level bins. The figure gives median growth rates and the dispersion of growth rates of settlements in 1975 that survive to 2014. Recall that survive means that they have not been absorbed by bigger nearby settlement as those settlements expanded, and survivors may have absorbed nearby smaller settlements over time.

Figure 1 indicates slower growth of settlements that were initially large, relative to those in the lower initial size bins. The median settlement amongst those that started in the lowest bin grows by 170 % (= 100 exp(0.99)-1), while the median growth rate in the highest bin is 72 % (=100 exp(0.54)-1). The dispersion of growth rates as depicted by the violin graph in the lower end bin is enormous. In this bin the fatter part of the violin is centred below the median and there is a long handle of large growth rates where cities increase in size up to 3000-fold. Although cities in bin 1 are initially tiny, (all less than 1.35 sq km and 95% less than 0.22 sq km), some grow so fast that by 2014 the biggest from this bin is over 18 sq km. As we move up the hierarchy in Figure 1, the spread experiencing really high growth rates shrinks. Settlements in the upper 2 bins (the top 31 cities in 1975 accounting for 40% of 1975 built area) have growth rates much more concentrated around the median.

These differential patterns of growth rates of 1975 settlements can be viewed through the lens of Gibrat's Law which says growth rates are independent of their initial sizes. The violin graph suggests this is not the case. Appendix Table A3 and Figure A4 support that rejection. Table A3 shows results from estimation of a regression of 1975 to 2014 growth on 1975 size for the 37007 of 47519 1975 settlements which survive to 2014. The table shows specifications with and without country fixed effects, as well as with bin fixed effects. Results yield a highly significant negative coefficient of about -0.09 in all cases. The corresponding Figure A4 shows how growth rates vary by initial size. As with the violin figure, smaller settlements have comparatively higher growth rates while larger ones grow more slowly.[2]

This result is very different than what the literature suggests for, say, American cities, where Gibrat's Law is often not rejected, as in Eeckhout (2004). We think this occurs because we are looking at a sample and time frame in Africa of enormous expansion in numbers of urban areas and rapid growth of all cities. This degree of motion and churning in the urban system are well beyond what developed countries have experienced in at least the last half century. In the largely stagnant/stable city systems in developed countries, after 1970 there is little entry of new cities and cross sections in different years can look quite similar. Here that is not the case, as we explore further next.

4.2 Transitions

In this section we trace the transition of settlements from their position in the 1975 size classes to their position in 2014 size classes (Table 1). We add to this births of settlements, since many 2014 settlements are not present in 1975, and also add exits since many 1975 settlements were absorbed into bigger neighbours as those neighbours expanded.

Table 2 shows a transition matrix for 1975 settlements going to 2014, with a column for exits and a row for births. Panel (a) shows counts, while Panel (b) shows shares. Rows are the 10 outcome bins, or states for settlements in 1975, using 1975 bin cutoffs. Columns are 2014 states, based on 2014 bin cut-offs from Table 1. Thus, an element in row i and column i is the probability that a settlement transitioned from 1975 state i to 2014 state j.

Reading the transition matrix, 69.9% of 1975 settlements in state 1 remain in state 1, while 7.3% advance to state 2 in 2014 and 0.4% advance to state 3. The diagonal of 1975 state i to 2014 state i is the probability that settlements remain in the same state and, as usual, dominates. But the domination is quite limited. Some diagonal elements are well under 0.50. In higher states there is enormous motion with both high upward and downward mobility. For example, starting in state 5, there is a 21.4% chance of moving up one state and a 8.9% chance of moving up two. Or from state 8, there is a 33% chance of moving up a state and starting from state 9, there is a 20.0% chance of moving up. There is also churning in relative sizes, whereby settlements move down states, especially at the upper end. Starting from states 8, 9, and 10, there are respectively a 33.3%, 40.0% and a 44.4% chance of moving down a state. Unlike American data as in Harris Dobkins and

[2]The error bands are pretty tight up to a 1975 size of about about 7 sq km, but even at a size of 55 sq km or more they are significantly less than growth rates for small settlements.

Ioannides (2001) and Black and Henderson (2003) for example, there is huge downward mobility in relative rankings as very fast growing settlements in this massive urbanization period in sub-Saharan Africa surge and displace other settlements that were in the upper ranks.

Births and exits add motion. The next to last column shows exits, which are settlements that merge with, or are overrun by a bigger neighbour. There are a large number of exits from the first 4-5 states in 1975, with 22.3% of state 1 settlements being absorbed into other places. Even in states 4 and 5, 13.0% and 7.1% respectively of settlements from 1975 are absorbed. The bottom row shows net births that occur between 1975 and 2014 (net births are births that survive: are not merged into a bigger settlement). Births are almost always in state 1 in a given observed year (1990, 2000, and 2014) and 99.4% of the 74462 net births remain in 2014 state 1. Of the 47519 settlements in 1975, 10512 exit and 37007 survive to 2014. These 37007 survivors plus the 74462 births make up the 111469 settlements in 2014, noting the vast majority are births. Moreover, for the 2014 stock, 105366 are in state 1 and of those 70% are births since 1975. While in the 8 top states in 2014 there are no births, only settlements that were present in 1975.

4.3 Urban Shadows

Recent work has focused on the urban shadows hypothesis (e.g. Cuberes et al. (2021); Beltràn Tapia et al. (2017); Bosker and Buringh (2017)), whereby being in the shadow of a giant city detracts from growth. In Appendix B, we examine a version of this hypothesis. We find that, for settlements in general, having more other settlement activity very nearby (0-50km) detracts from growth, potentially a competition effect. Having more activity nearby but not immediate (50 - 150 km) on the other hand offers market potential and improves growth performance as an average effect (as in Jedwab and Storeygard, 2021 for Africa). However the extent to which activity is centered in bigger settlements detracts from growth, consistent with the urban shadow hypothesis. What stopped us from pursuing this line of inquiry was that these effects varied by place in the urban hierarchy. The biggest cities benefit from more proximate 1975 activity (probably because they overrun and absorb immediate towns), and they are indifferent or perhaps somewhat divorced from activities at 50 - 150 km. That led us to really try to grapple with the notion of urban hierarchies, both in theory and in empirical implementation.

5 The Geographical Pattern of Settlements: a Model of Urban Hierarchy

We are interested in the spatial distribution of settlements, as well as the statistical distribution of their size and growth. We observe geographical spacing and differential growth patterns of settlements of different sizes. The remainder of the paper is devoted, using theory matched with empirical work, to understanding these growth patterns. The central idea is that growth patterns of settlements arise as a consequence of a pattern of competing and complementary interactions between places. Competing, as places may supply similar outputs (goods that are close substitutes), and compete for similar primary inputs. Complementary, as places may produce quite different products which are supplied to households in nearby locations, and are also used by firms as intermediate inputs. Demand from neighbours creates a demand or backwards linkage, and access to supply of intermediate inputs constitute a cost or a forwards linkage.

We capture this in a model in which there are three different sectors which correspond broadly to activities in developing countries, in which primary sectors of production employ a large part of the labour force. Sector 3 is agriculture, using land and labour to produce goods that go both to final consumption and further processing, but are costly to ship (bulky or prone to rapid deterioration). Sector 2 is agro-processing, or more broadly activities that support the agricultural sector; it uses sector 3 output as an input, and its output (such as processed food products) is less costly to ship. Sector 1 is manufacturing and modern services, outputs that are also relatively easy to ship between places.

A starting question is, where do these sectors locate? We suppose that there are many possible locations, each ex ante identical (endowed with the same amount of land and level of technology) and that labour is perfectly mobile between places and sectors. Starting from a position in which all places are identical we show how a pattern of settlement emerges, with settlements becoming of different types, i.e. specialising, at least partially, in different sectors. There is regularity in the spacing of settlements of different types, with types having different sizes and spatial frequencies. Along the path to this outcome settlements grow faster if they are near to settlements of different types and remote from settlements of the same type.

5.1 Model Structure:

We set up the model for a general input-output structure and geographical space. Results come from simulation and details of implementation and parameters are given in the following sub-section.

There are N points (or places) in a geographical space, and these places are labelled with subscripts. The distance between two places i, j is d_{ij}, and this underpins the costs of shipping goods and services around the space. Each place has a fixed endowment of land. There are three sectors of production, as outlined above, indexed by superscripts $s, r = 1, 2, 3$. There is place and firm specific product differentiation, represented by CES modelling of differentiation.

Sectoral demand: The price index for sector s products in place i is P_i^s,

$$P_i^s = [\sum_j n_j^s (t_{ji}^s p_j^s)^{(1-\sigma^s)}]^{1/(1-\sigma^s)}, \quad s = 1, 2, 3. \tag{1}$$

This is the usual CES aggregator, with p_j^s and n_j^s respectively the price and number of varieties produced in place j, t_{ji}^s the iceberg trade cost factor in shipping from j to i, and σ^s the elasticity of substitution. All these variables and parameters are sector specific.

The value of demand for sector s output in place j is E_j^s, so total demand (across all locations) for a sector s variety produced in place i is

$$x_i^s = (p_i^s)^{-\sigma^s} \sum_j E_j^s (P_j^s)^{(\sigma^s-1)} (t_{ij}^s)^{(1-\sigma^s)}, \quad s = 1, 2, 3. \tag{2}$$

Production: Production uses primary factors (labour and, in sector 3, also land) and intermediates with Cobb-Douglas technologies, so has unit cost functions (equal to price)

$$p_i^s = (w_i^s)^{(1-a^{1s}-a^{2s}-a^{3s})} (P_i^1)^{a^{1s}} (P_i^2)^{a^{2s}} (P_i^3)^{a^{3s}}, \quad s = 1, 2, 3. \tag{3}$$

The exponents a^{rs} are the value share of sector r in production of sector s and w_i^s is the place i sector s price of primary factors. This allows for all sectors to be used as input to all other sectors, although we will set some of these input-output coefficients to zero in what follows. A key link is a^{32}, the input of primary to agro-processing, sector 3 to sector 2.

Sectors 1 and 2 are monopolistically competitive, with an endogenously determined number of firms each producing a distinct variety which breaks even when producing and selling one unit of output, so [3]

$$x_i^s = (p_i^s)^{-\sigma^s} \sum_j E_j^s (P_j^s)^{(\sigma^s-1)} (t_{ij}^s)^{(1-\sigma^s)} = 1, \quad s = 1, 2. \tag{4}$$

Labour is the only primary factor used in these sectors, so equations (3) and (4) can be thought of as defining a wage equation, i.e. giving the value of w_i^s at which firms break even, as a function of

[3]Equation (3) is average cost at unit scale of production.

price indices, numbers of varieties, and expenditure levels throughout the economy.

Sector 3 is agriculture, and we give it a slightly different and simpler treatment. Each place is endowed with the same quantity of land and uses land and labour with fixed coefficients to produce a fixed quantity of a single place specific variety.[4] In equations (3) and (4) this means that n_i^3 = 1, and x_i^3 takes fixed and uniform value x^3. The level of sector 3 employment is therefore the same everywhere, L^3. However, since demand and the input prices vary across places, so too does the market clearing price of each place's agricultural variety, p_i^3, and hence also w_i^3, the return to primary factors, labour and land. This return could be divided between a wage and a rent component but, since much African land is operated by family farms under traditional communal land tenure, we leave it as a combined return to labour and land. We assume that the return is large enough to retain L^3 units of labour in each place.

Income and expenditure: Wage bills in each sector and place are the share of labour in the value of output,

$$w_i^s L_i^s = (1 - a^{1s} - a^{2s} - a^{3s}) n_i^s p_i^s x_i^s, \quad s = 1, 2, 3, \tag{5}$$

where, as noted above, in sector 3 this takes the form $w_i^3 L^3 = (1 - a^{1s} - a^{2s} - a^{3s}) p_i^3 L^3$, with $w_i^3 L^3$, interpreted as a combined return to land and labour. Summing across sectors, total income in each place is given by

$$Y_i = w_i^1 L_i^1 + w_i^2 L_i^2 + w_i^3 L^3. \tag{6}$$

Expenditures in each place i on products of sector s come from final and derived demands and are

$$E_i^s = \mu^s Y_i + \sum_{r=1,2,3} a^{sr} n_i^r p_i^r x_i^r, \quad s = 1, 2, 3, \tag{7}$$

where consumer preferences are Cobb-Douglas, with sector shares μ^s. The consumer price index in each place, P_i, and per worker utility in each place-sector pair, u_i^s, are therefore

$$P_i = (P_i^1)^{\mu^1} (P_i^2)^{\mu^2} (P_i^3)^{\mu^3}, \qquad u_i^s = (w_i^s)/P_i. \tag{8}$$

The total labour force is fixed at L, of which NL^3 workers are engaged in agriculture, and the remaining $L - NL^3$ are perfectly mobile between sectors 1 and 2 and all places. It follows that all places that have employment in either sector 1 or 2 have equal values for u_i^s, s = 1,2, with utility less than or equal to this in places-sectors where there is no employment in these sectors.

Before moving to implementation of the model a few further comments are in order. First, all places

[4]This is 'Armington' product differentiation, in contrast to the firm-specific differentiation of sectors 1 and 2.

– including large settlements with manufacturing or agro-processing – also have agriculture, sector 3. This is partly for simplicity, but also supported by evidence on widespread agricultural output produced in African urban areas (Henderson and Kriticos 2018). Second, land is not explicitly modelled, except as an input to agriculture, where it is combined in fixed proportions with labour. Neither rent, nor amenity or congestion enter consumer utility. This is for simplicity, although it also reflects the difficulty of modelling African land tenure across the range of settlements we study. Third, product differentiation and variety effects create agglomeration and spatial structure in this model, exactly as in the basic core-periphery model (Krugman 1991, and its extension to intermediate products in Fujita et al. 1999). The model is isomorphic to one in which the number of varieties is fixed and replaced by technological agglomeration externalities.

5.2 Implementation

We use numerical simulation to track the evolution of the system of settlements, focusing on several stylised cases. The simplest, and that which yields the greatest symmetry is to assume that places are located on the circumference of a circle – the racetrack economy – with radius of unity and distance between places measured around the circumference. For a richer picture we also show results for places on a hexagonal lattice set on a (near) circular disk. This has the advantage of being a two-dimensional space, but is complicated by having an edge (i.e. not being a featureless natural geography). Transport costs are assumed to be exponential in distance, $t_{ij}^s = exp(-t^s d_{ij})$. For clarity, we present results only for the case in which just one of the input-output linkages is switched on, that of agricultural supply to agro-processing, so $a^{32} > 0$. Agricultural products from each place are assumed to be close substitutes ($\sigma^s = 20$) with high transport costs, such that shipping just 6 degrees around the circumference of the circle loses 50 percent of output. Elasticities of substitution and trade costs are lower in the other two sectors, and a full list of parameter values is given in Appendix C.

Our main experiment is to start this model from an equilibrium in which all places are identical – economic activity is uniformly distributed across space. This is an equilibrium which is stable if trade costs are all very high, making each of the places autarkic. Spatial reorganisation is initiated by (a) reducing trade costs to a point at which this equilibrium is unstable, and (b) perturbing the equilibrium by a small random redistribution of the labour force, and having labour move in response to utility differences between sectors and places, u_i^s.

The ensuing long-run equilibrium is illustrated in Figure 2. The top panel is the racetrack economy, and has $N = 600$ places on the horizontal axis (the two ends connecting around the circle) and employment in sectors 1 plus 2 is on the vertical. This is shown as employment relative to mean employment in settlements and measured in log units, so 0 is this mean value. Agricultural employment takes place everywhere, at value L^3, and is not shown on the figure. Sectors 1 and 2 both concentrate in a subset of places which we will refer to as type-1 and type-2 settlements, noting that type-1 settlements contain sector 1 and sector 2 employment, while type-2 settlements are exclusively sector 2. The reason for concentration is the usual home-market effect, as consumers are attracted to places with a large supply of locally produced varieties, and firms are attracted to the large market created by these consumers. Sector 1 operates in just four evenly spaced places and these are relatively large, the spikes in the figure. Sector 2 operates in these places and in the 24 places indicated by the smaller spikes. The relatively large number of these type-2 places arises because sector 2 uses sector 3 output as an input to production, and sector 3 output is produced everywhere and is particularly costly to transport. Since it operates in more places, type-2 settlements are (ceteris paribus) significantly smaller than type-1 settlements. In short, there are many 'market towns' (type-2 settlements) fairly evenly spread since they are supplied by dispersed agriculture, and fewer but larger manufacturing cities.

Figure 2 was generated by a small random perturbation of employment which caused the system to evolve away from a uniform distribution of activity. Different simulations with the same parameters but different initial (small) perturbations all produce a very similar outcome for reasons first expounded (in a different context) by Turing (1952) and applied to the spatial context by Fujita et al. (1999). However, Figure 2 displays an extreme degree of regularity that does not hold generally. Type-1 and type-2 settlements form at different frequencies, and the example in Figure 2 is constructed such that they mesh together in a regular way (24 type-2 divided by 4 type-1 is an integer).

The bottom panel of Figure 2 is a similar equilibrium, constructed with the same parameters but now with the geographical space being an entire disk, rather than just its circumference. The largest settlements (containing type-1) are yellow shading to orange, smaller ones (type-2) light-blue, and sector 3 (agriculture only) everywhere (dark-blue). A clear structure of settlements has emerged, with a central large type-1 settlement, a further 8 such settlements further out with regular spacing (and rational symmetry of order 4) and 'market towns' (18 type-2 settlements) interspersed between them.

Varying parameters used in the simulation changes the spatial equilibrium in complex ways, and results are detailed in Appendix C. Looking first at transport costs, lower transport costs for primary output (sector 3) leads to fewer and larger type-2 settlements as the benefits of locating agro-processing very close to agriculture are reduced. Lower transport costs for sector 2, agro-processing, have the opposite effects, increasing the number type-2 settlements as it tips the transport cost balance in favour of these settlements locating close to agriculture inputs rather than consumers in type-1 settlements. Lower transport costs for sector 1 reinforces agglomeration and, for 20 percent lower costs, reduces the number of type-1 settlements in the racetrack economy from 4 to 3. A 20 percent reduction in transport costs across the board has no effect on the number of type-1 settlements and reduces the number of type-2, the driving force being the reduction in the need for these settlements to be close to agriculture. Similar experiments can be conducted with other parameters. Reducing σ^s in all sectors facilitates agglomeration, reducing the number and increasing the size of settlements, as is usual in models of this type. The input-output linkage between sectors 2 and 3 is important, and reducing this below a critical point causes sectors 1 and 2 to co-locate, so there are no distinct type-2 settlements.

An alternative comparative static experiment is to take an equilibrium such as those illustrated in Figure 2, and perturb productivity or employment in a subset of places. For example, raising productivity in a single type-2 settlement, holding employment in all other places constant, has the effect of reducing utility in nearby type-2 settlements, and raising it in nearby type-1, the competing and complementary effects we expect. Letting employment in other places change in response to these utility differences creates waves or ripples of activity. Nearby type-2 places contract, their contraction causing places further away to expand, and so on. Since there is no sunk capital in the model, changes of this type may well cause settlements to move, i.e. some places empty out completely, and new settlements form.

To link the model to the following empirical analysis we revert to our core simulations, in which the pattern of settlement evolves away from an initial uniform distribution of activity. We now add some noise to these simulations by giving places randomly and independently different productivity levels in sectors 1 and 2 (i.e. shifts in the cost function, equation 3). Figure 3 (top panel) illustrates an equilibrium with these place-sector productivity differentials. It is conceptually similar to the bottom panel of Figure 2, but the rotational tidiness of Figure 2 is disturbed. Crucially, it retains a pattern of few (8) well-spaced type-1 settlements interspersed with type-2 (of which there are 22).

What are the spatial relationships between different types of settlements in these equilibria? The bottom part of Figure 3 (4 panels) uses repeated simulations to construct scatter plots of sector s employment in each place as a function of proximity to employment in the same and other sectors, r. We measure place i's proximity to sector r employment in other places j as $\sum_{j\neq i}\theta_{ij}L_j^r$ where $\theta_{ij} = exp(-Kd_{ij})$. The scatter plot in the upper-left panel has employment in sector $s = 1$ (in places where it is positive) on the vertical axis, and proximity to other places' sector $s = 1$ activity on the horizontal. The plots combine 5 separate runs of the model, each with different random draws of productivity levels in each place, so the number of data points is the sum of all type-1 settlements in the 5 runs, or 40 size and proximity pairs. The negative relationship illustrates the competing relationships that we expect between settlements of the same type. The two scatter plots on the right give the effect on employment in sector 2 of proximity to employment in each sector, so the upper-right plot gives sector $s = 2$ employment on proximity to sector $r = 1$ employment, and indicates a positive or complementary relationship. Competing and complementary interactions show up with negative same-sector effect and positive cross-sector effect. Simple regressions of the form $L_i^s = \alpha_s + \beta_{ss} \sum_{j \neq i} \theta_{ij} L_j^s + \beta_{rs} \sum_{j \neq i} \theta_{ij} L_j^r + u_i^s$, $r, s = 1, 2$, on numbers from this example yield significant negative own effects, β_{ss}, and significant positive cross effects, β_{rs}, in each case, indicating the competing same-type and complementary cross-type effect. The β's are represented by the slopes of what would be the best fit lines through the scatters in each panel in Figure 3.

The empirical work in the following section takes the African data to regressions of similar form to that above. We look both at cross-sectional regressions and at growth regressions, i.e. asking whether a settlement grows relative fast (slowly) if it is close to fast growing neighbours of different (similar) type.

6 The Geographical Distribution of Settlements: Empirics

In this section we describe and report results from an empirical estimation of the model of Section 5. We work with three types of cities, which is the maximum differentiation we are comfortable with empirically, and also follows the model. We first present the empirical specification and then results.

6.1 Specification

For each country, divide settlements into 3 groups by the size of their built area. Group 1 are settlements with the largest built areas; group 2 are middle size, group 3 are the smallest. We discuss the criteria that define these groups below. For settlements in each group, we want to know the effect of being close to other settlements of same type (i.e. in the same size group) and of proximity to settlements in other size groups. Looking first at group 1, b_i^1, is the log of built area for settlement i in this group, and $b_i^1 = 0$ for settlements not in group 1. Effects depend on the proximity of this settlement i to the built area of settlements in each of the groups, where θ_{ij} is a measure of the proximity of settlement i to j, so $\sum_{j \neq i} \theta_{ij} b_j^G$ captures the proximity to i weighted sum of log built area in group G where G can take values from 1 to 3. Going across the 3 groups, estimating equations are;

For i in group 1:

$$b_i^1 = \alpha_1 + \beta_{11} \sum_{j \neq i} \theta_{ij} b_j^1 + \beta_{21} \sum_{j \neq i} \theta_{ij} b_j^2 + \beta_{31} \sum_{j \neq i} \theta_{ij} b_j^3 + controls \tag{9}$$

For i in group 2:

$$b_i^2 = \alpha_1 + \beta_{12} \sum_{j \neq i} \theta_{ij} b_j^1 + \beta_{22} \sum_{j \neq i} \theta_{ij} b_j^2 + \beta_{32} \sum_{j \neq i} \theta_{ij} b_j^3 + controls \tag{10}$$

For i in group 3:

$$b_i^3 = \alpha_1 + \beta_{13} \sum_{j \neq i} \theta_{ij} b_j^1 + \beta_{23} \sum_{j \neq i} \theta_{ij} b_j^2 + \beta_{33} \sum_{j \neq i} \theta_{ij} b_j^3 + controls \tag{11}$$

The coefficients of interest are the nine β coefficients. The estimated coefficients β_{GI} capture the effect of proximity to built area of type G on built area of type I, where $I, \; G = 1, \; 2, \; 3$. For the impact of city j of type G on city i of type I, they are elasticities mediated by the θ_{ij}'s.

From the model we expect β_{11}, β_{22}, and $\beta_{33} < 0$. This would be evidence of competition effects, or that settlements of the same type are in a competing relationship with each other. We expect $\beta_{IG} > 0$ for $I \neq G$, so that settlements of other types complement each other: a positive shock to G generates greater demand for settlement I type products and thus enhanced size.

The exact proximity measures are $\theta_{ij} = \exp(-K d_{ij})$, where d_{ij} is Euclidean distance between settlements in 100's of km. The parameter K measures a rate at which the negative impact of bilateral distance diminishes. We set $K = 0.25$. So at 100km, a neighbour's level of activity is discounted by 22% while at 200km, the discount is 39% and at 500 km it is 71%. We further explore this parameter, increasing and decreasing K in robustness checks.

The system above is described for a single country. We estimate it separately for each country in our African sample, yielding different α, β coefficients for each country, and assuming that spatial interactions occur only within country, so $\theta_{ij} = 0$ for i, j, settlements in different countries.

In implementation and estimation, the definition of groups is country-specific. For each country we first rank all settlements by built-up in the initial period from largest to smallest, then obtain the cumulative built-up. We define groups based on the cumulative share of built-up, using fractions k_1 and k_2. For example, for the first group, $k_1 = 0.6$ sets the largest settlement in the group to be the cut-off settlement such that the accumulated share of built up is strictly equal or less than 0.6. The second group includes the next ranked settlements up to that at which the cumulative share of built-up reaches k_2. The last group includes the rest of the settlements. Given starting k_1 and k_2, we run the regressions as shown in the equations 9 to 11. We then sum the residual sum of squares [RSS] from each of the three equations to obtain the total RSS for the set of k_1 and k_2. To choose the best k_1 and k_2, we iterate k_1 from 0.6 to 0.9 with 0.05 as the interval and k_2 from 0.65 to 0.95 with 0.05 as the interval. We choose the k_1 and k_2 that gives the minimum RSS. Appendix Figure D1 illustrates what are the cut-offs which minimize the RSS in six cases given in Figure 6 analyzed below.

In estimation we include controls for terrain ruggedness, distance to the nearest harbor, distance to large lakes, distance to rivers, elevation, distance to the coast, the Ramankutty land suitability index, temperature and precipitation as discussed in Henderson, et al. (2018). We also drop the largest city in each country which is in all cases the de facto political capital as of 1990. This is based on findings in Ades and Glaeser (1995) and Davis and Henderson (2003) that the size and growth of these political capitals are less governed by market forces than by political forces. In robustness checks we examine what happens if we add the capital back in. Finally, to be consistent with growth specifications, the sample for each country is the 1975 settlements which survive to 2014.

Following the model of section 5, we first estimated the levels system (9) - (11). That is, we estimated a simple cross section of level built area of settlement i, as a function of contemporaneous level $\sum_{j\neq i}\theta_{ij}b_j^G$, for 2014 data, for 1975 settlements which survive to 2014. We present the results for 33 countries, dropping counties which have less than 200 settlements. Results are in Figure 4. Figures are designed so each column corresponds to an equation for each type. So column 1 is for type-1, the largest types of settlement as in Equation (9). Then each row reports on the type of neighbour variable, so row 2 column 1, β_{21} is the effect of type-2 neighbour settlements on type-1 cities in Equation (10). Row 3 column 2, β_{32}, is the effect of the smallest settlements, type-3, on type-2 neighbouring settlements. And so on. Note that degrees of freedom for each of the 3 column estimating samples are given at the bottom of each figure. The implied sample sizes for each of the 3 types are endogenous depending on the k_1 and k_2 which minimize the RSS. (If that minimization point is not at a division where all equations have positive degrees of freedom, the country is dropped, resulting in a loss of either 1 or 2 countries depending on the specification.) Coefficients are given by the error bars about their estimated value.

In Figure 4 the diagonals report the own effects for each type of settlement. The theory suggests all these should be negative, and that is generally what we find. For example, for type-1 settlements in the top left panel, the own type neighbour effects are all negative, except for one insignificant one. All insignificant coefficients are marked in red with an "X" added. In row 2 (middle panel) for type-2 settlements the own effects are significantly negative in 19 cases and insignificant in the other cases. However, in row 3 (last panel) the own effects for the smallest settlements are all over the place, with only 9 significantly negative. The first 2 diagonal elements support our priors: the idea of substitutability, or competition effects.

Off diagonal elements give the cross-effects. We expect these all to be positive. For proximity of type 2 to type 1, type 1 to type 2 and type 3 to type 2, in 50% of the cases, coefficients are significantly positive. However in the rest taken together, proximity of type 3 to type 1, type 1 to type 3 and type 2 to type 3, about 1/3, 1/4, and 1/2 are negative, positive and insignificant respectively. That is, there is no clear pattern. However, this cross-section specification is flawed, so we do not go into further detail.

Why is the specification flawed? At the most basic level, settlements will have time invariant unobservables (beyond our observed controls), such as general socio-economic status of the settlement which affect size and growth (Moretti, 2004), as well as other geographic and institutional factors.

To address this first problem, we run a growth specification of eqs (9) to (11), to remove the impact of time persistent unobservables. For this growth on growth formulation, variables b_i^1 in equations (9) - (11) become log of built area in final period relative to initial period values, i.e. log of proportionate change for each city. We leave in geographic controls which in principle are first differenced out, because one might think of growth formulations in which they might affect growth as well as levels.

In estimation we look at two different growth episodes: 1975-2014 and 2000-2014. Estimation still of course has issues: results are still subject to missing variables that may be correlated with contemporaneous parts of b_i^1. For example, while the model is simple without frictions, settlements could be subject to time varying but persistent (correlated) productivity shocks which enhance their growth rates. These shocks then spill over through trade and migration interactions to their neighbours and hence are correlated with the neighbour based treatment effects. To try to deal with this in the estimation for growth from 2000 to 2014, we insert variables for own 1975 and 1990 sizes, which helps control for these influences (Duranton et al., 2014). But the problem of contemporaneous shocks affecting own growth and neighbours through regional shocks and interactive feedback effects remains. The classic way (Arellano and Bond, 1991) to deal with this is to instrument with lagged values, in this case 1975 and 1990 values of the neighbour-settlement type variables. While these are strong instruments in our context, they do not solve the problem. Even the weak tests for exogeneity of such instruments cited in most empirical work simply fail in our context.

Instead we accept that there is bias but focus on the direction of bias. As noted, the key issue concerns local regional shocks which affect the own settlement and neighbouring settlements. Then the bias direction is positive: the effects of a positive shock on the own settlement are also experienced by neighbours, so the neighbour coefficient picks up part of the own settlement effect of this shock and is biased upwards. For results for own-type settlements that are in competition where we expect substitution effects, our negative estimates will be, in absolute terms, lower bounds on the effect. However for complements in the hierarchy, the positive effects may truly exist but they are biased and are an upper bound.

6.2 Results on Growth Formulations

Results for the 1975-2014 growth specification are shown in Figure 5 and the 2000-2014 growth specification in Figure 6. For the latter besides the geographic variables, as noted, we also control for own, predetermined sizes in 1975 and 1990. The 1975-2014 specification can't include lagged own values, but it gives a nice long difference. As we will see, patterns of the two sets of results are similar.

In Figure 5 for growth from 1975 to 2014, considering the own type of neighbour effects in the diagonal panels, for type-1 cities, all but 4 are significantly negative and the 4 are insignificant. The β_{11}'s are centred around about -0.75 or so. For the elasticity of city j of type 1 affecting the own city i of type 1, the country specific β_{11} is mediated (or multiplied) by θ_{ij}, which for 400km apart, for example, would be 0.37. For type-2 settlements (middle panel in row 2), all but 2 of 31 coefficients are significantly negative. Finally for type 3 settlements, in the last panel, all own effects are significantly negative, except 4 which are insignificant. Overall, there is strong evidence of competition effects and substitutability, especially given these are lower bound estimates on how negative the effects can be.

For the off-diagonal and expected complementarity effects in Figure 5, we see evidence now of positive effects. For type-1 settlements, in column 1 row 2, 17 coefficients are significantly positive and only one significantly negative. In column 2 for the effects of type-1 and type-3 settlements on type-2, 41 of 62 are significantly positive and only 2 are significantly negative. In column 3, for the effect of type-2 settlements on type-3 we have 22 significant positives and no significant negative. This is fairly strong evidence of complementary effects, albeit from biased coefficients. Where this weakens is when types are "far apart": the effects of type-3 settlements on type-1 and vice versa. There, of 62 cases in the bottom left and top right panels, only 22 are significantly positive, while 15 are significantly negative. This is as might be expected, as theory suggests the effect is a composite rather than a direct effect, i.e., type-1 complementary with type-3 is intermediated through type-2 rather than a direct linkage effect. Overall, the evidence suggests cross-type coefficients that are positive and significant, except for the effects of type-1 on type-3 and vice versa.

Similar patterns hold for Figure 6 where we look at growth for 2000 to 2014. Here the patterns are stronger and clearer, we think because the results are better founded econometrically. Here, we have past sizes (in 1975 and 1990) as covariates as in Duranton and Turner (2014) to better control for any influences of time invariant variables on settlement growth. It is for this set of regressions that we conduct various robustness checks and show the full set of results.

In Figure 6, on the diagonal across all 3 panels, all but 3 of 96 coefficients are significantly negative. This is even stronger evidence of substitutability effects from own types of neighbours. On complementarity from non-own type neighbors, patterns are similar to Figure 5 but a stronger and clearer pattern emerges. Types near each other complement each other. For the effects of type-2 settlements on type-1, the effects of type-1 and type-3 settlements on type-2 and for the effects of type-2 settlements on type-3, of the 128 cases, 99 have significant positive coefficients and only 5 significant negative. However, for types-1 and -3 which more distant from each other in the hierarchy, there is lack of clear evidence of complementarity. In the bottom left and top right panels, as above, coefficients are pretty evenly split among being positive, negative and insignificant. In Appendix E, Tables E1 to E32 we present the full regression results that correspond to Figure 6, so as to see results on all controls and the numbers for the variables of greater interest.

6.3 Robustness Checks

We conduct 2 types of robustness check reported in Appendices. Additional to those we did experiment with altering the spatial decay parameter K, trying values different than 0.25, given there are no trade data between between our cities available to estimate the parameter. Changing K has some quantitative impacts, but the pattern of results in Figure 6 is unaffected.

The first formal robustness check involves the fact that we dropped the capital city as of 1990 in each country, where in 1990 that was always the primate city. Figure F1 adds back in the primate city. Compared to Figure 6, the diagonal results of own type neighbours on oneself are again virtually all significantly negative. Again, the vast majority of type-1 and -2 and then type-2 and -3 interactions are complementary, while evidence on type-1 and -3 interactions is much more mixed. Qualitatively, results are unchanged relative to Figure 6.

Second, we examine what happens if drop very small 1975 settlements, those that are below 0.011 square km, or -4.5 in log scale in Appendix Figure A3b. From the full sample of 33338 1975 settlements which survive to 2014, this cut reduces the base sample to 16160 settlements.

However, that cut leaves us only 22 countries in the estimating samples with settlement counts over 200. Figure F2 in the Appendix shows the results. Again on the diagonals for complementarity effects, all coefficients but two are significantly negative. For the effects of type-1 settlements on type-2 and vice versa, and for type-2 on type-3 and vice versa, 76% of coefficients are significantly positive, presenting strong evidence of complementarity. As in other specifications presented above, for the effect of type-1 settlements on type-3 and vice versa, results are mixed. Again, qualitatively, results are the same as in Figure 6.

7 Conclusion

Sub-Saharan Africa experienced greater than doubling of its built area in the period 1975-2014, this putting in place a hierarchy of settlements that is likely to shape future development for decades – if not centuries – to come. This paper has describes this process and provides insights for some of the factors shaping this emerging hierarchy. Growth has been far from uniform, particularly for smaller settlements which had widely differing growth rates. On average, smaller settlements grew faster than settlements that had achieved scale by 1975, this somewhat reducing the relative position of larger settlements and in contrast to experience of higher income countries as captured by Gibrat's law.

We show that the relative growth performance of settlements is strongly dependent on their relationship to neighbouring settlements. Settlements grow more slowly if they are close to fast growing settlements of similar size, and faster if close to fast growing settlements that are either much larger, or much smaller than they are. This growth process generates a somewhat regular pattern of spacing of settlements of similar sizes. We rationalise this in terms of a theoretical model in which settlements perform different functions – primary and agriculture, primary-processing, and manufacturing and services. Patterns of complementarity and competition between these functions generates the growth performance and emergent urban hierarchy that we see in the data.

References

Arellano, Manuel and Stephen Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 04 1991, *58* (2), 277–297.

Black, Duncan and Vernon Henderson, "Urban evolution in the USA," *Journal of Economic Geography*, 10 2003, *3* (4), 343–372.

Bosker, Maarten and Eltjo Buringh, "City seeds: Geography and the origins of the European city system," *Journal of Urban Economics*, 2017, *98*, 139–157. Urbanization in Developing Countries: Past and Present.

Cuberes, David, Klaus Desmet, and Jordan Rappaport, "Urban growth shadows," *Journal of Urban Economics*, 2021, *123*, 103334.

Davis, James C. and J. Vernon Henderson, "Evidence on the political economy of the urbanization process," *Journal of Urban Economics*, 2003, *53* (1), 98–125.

de Bellefon, Marie Pierre, Pierre Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin, "Delineating urban areas using building density," *Journal of Urban Economics*, 2020, (October 2018).

Desmet, Klaus and Jordan Rappaport, "The settlement of the United States, 1800–2000: The long transition towards Gibrat's law," *Journal of Urban Economics*, 2017, *98*, 50–68.

Düben, Christian and Melanie Krause, "The Emperor's Geography - City Locations, Nature, and Institutional Optimization," *SSRN Electronic Journal*, 2021.

Duranton, Gilles, "Urban Evolutions: The Fast, the Slow, and the Still," *American Economic Review*, March 2007, *97* (1), 197–221.

__ , Peter M. Morrow, and Matthew A. Turner, "Roads and Trade: Evidence from the US," *The Review of Economic Studies*, 04 2014, *81* (2), 681–724.

Eeckhout, Jan, "Gibrat ' s Law for (All) Cities Gibrat ' s Law for (All) Cities By," 2004, (1949), 1429–1451.

Ellison, Glenn and Edward L. Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *American Economic Review*, May 1999, *89* (2), 311–316.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables, *The Spatial Economy: Cities, Regions, and International Trade*, The MIT Press, 1999.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 1999, *114* (3), 739–767.

Harris Dobkins, Linda and Yannis M Ioannides, “Spatial interactions among U.S. cities: 1900–1990,” *Regional Science and Urban Economics*, 2001, *31* (6), 701–731.

Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil, “The Global Distribution of Economic Activity: Nature, History, and the Role of Trade1,” *The Quarterly Journal of Economics*, 09 2017, *133* (1), 357–406.

Jedwab, Remi and Adam Storeygard, “The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010,” *Journal of the European Economic Association*, 06 2021. jvab027.

Krugman, Paul, “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 1991, *99* (3), 483–499.

Lösch, August, “The Economics of Location, Jena, Germany: Fischer,” *English translation, Yale U. Press, New Haven*, 1954.

Tabuchi, Takatoshi and Jacques-François Thisse, “A new economic geography model of central places,” *Journal of Urban Economics*, 2011, *69* (2), 240–252.

Tapia, F. Beltràn, A. Díez-Minguela, and J. Martínez-Galarraga, “The Shadow of Cities: Size, Location and the Spatial Distribution of Population in Spain,” Cambridge Working Papers in Economics 1749, Faculty of Economics, University of Cambridge November 2017.

Tables

Table 1: The data in the 2014 cross-section

	Size category	Min size	Max size	Total built	Shell area	Count	Share built
1	[0.0009,0.251]	0.0009	0.2511	2608	447165	105366	0.1000
2	(0.251,1.44]	0.2520	1.4445	2607	112957	4716	0.1000
3	(1.44,5.85]	1.4454	5.8536	2606	62547	954	0.0999
4	(5.85,18.2]	5.8554	18.1845	2595	41014	261	0.0995
5	(18.2,43.6]	18.3942	43.6311	2586	26117	91	0.0991
6	(43.6,95.5]	43.6869	95.4792	2645	34056	41	0.1014
7	(95.5,173]	96.8661	173.4880	2494	25710	20	0.0956
8	(173,313]	173.6270	313.4980	2486	17023	10	0.0953
9	(313,544]	347.6650	543.6480	2804	20616	7	0.1075
10	(544,1.37e+03]	634.6690	1370.4000	2649	12997	3	0.1016

Note: The table depicts the a summary of our 2014 data for Africa as a whole. It divides the data into 10 bins of (almost) equal share of total built area. Shares and total built area (in sq. km.) in each bin are given in columns 7 and 4 respectively.

Table 2: Transition matrix

Panel (a)

	1	2	3	4	5	6	7	8	9	10	exit	rowsum
1.0	31315	3276	188	6	0	0	0	0	0	0	10005	44790
2.0	5	1039	557	46	0	0	0	0	0	0	421	2068
3.0	0	0	197	141	16	4	0	0	0	0	66	424
4.0	0	0	0	61	39	5	2	0	0	0	16	123
5.0	0	0	0	4	31	12	5	0	0	0	4	56
6.0	0	0	0	0	5	18	2	2	0	0	0	27
7.0	0	0	0	0	0	2	7	4	1	0	0	14
8.0	0	0	0	0	0	0	4	2	3	0	0	9
9.0	0	0	0	0	0	0	0	2	2	1	0	5
10.0	0	0	0	0	0	0	0	0	1	2	0	3
birth	74046	401	12	3	0	0	0	0	0	0	0	0

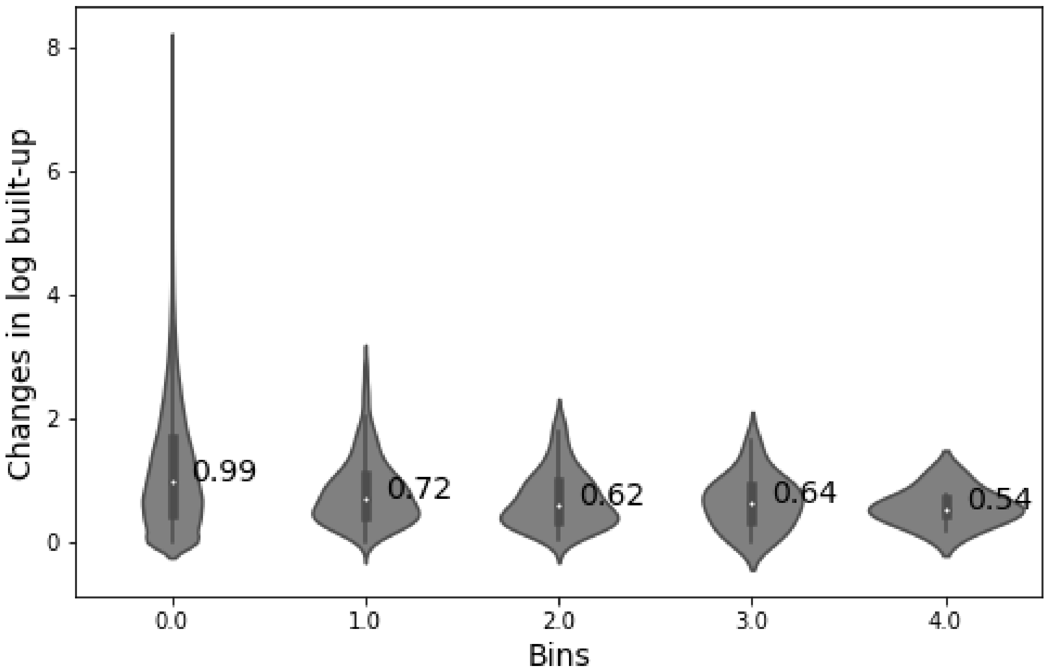
Panel (b)

	1	2	3	4	5	6	7	8	9	10	exit	rowsum
1.0	0.699	0.073	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.223	1.000
2.0	0.002	0.502	0.269	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.204	1.000
3.0	0.000	0.000	0.465	0.333	0.038	0.009	0.000	0.000	0.000	0.000	0.156	1.000
4.0	0.000	0.000	0.000	0.496	0.317	0.041	0.016	0.000	0.000	0.000	0.130	1.000
5.0	0.000	0.000	0.000	0.071	0.554	0.214	0.089	0.000	0.000	0.000	0.071	1.000
6.0	0.000	0.000	0.000	0.000	0.185	0.667	0.074	0.074	0.000	0.000	0.000	1.000
7.0	0.000	0.000	0.000	0.000	0.000	0.143	0.500	0.286	0.071	0.000	0.000	1.000
8.0	0.000	0.000	0.000	0.000	0.000	0.000	0.444	0.222	0.333	0.000	0.000	1.000
9.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.400	0.400	0.200	0.000	1.000
10.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.667	0.000	1.000
birth	0.994	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Table 2 shows a transition matrix for 1975 settlements going to 2014, with a column for exits and a row for births. Panel (a) shows counts, while Panel (b) shows shares. There are 10 bins, or states. For columns, states are based on 2014 bin cut-offs from Table 1. For rows, the first 10 rows in column 1 are states for cities in 1975, using 1975 bin cutoffs. State 1 is for 1975 cities in the bottom 1975 bin. State 2 is for 1975 cities in the second bin based on 1975 cut-offs, and so on.

Figures

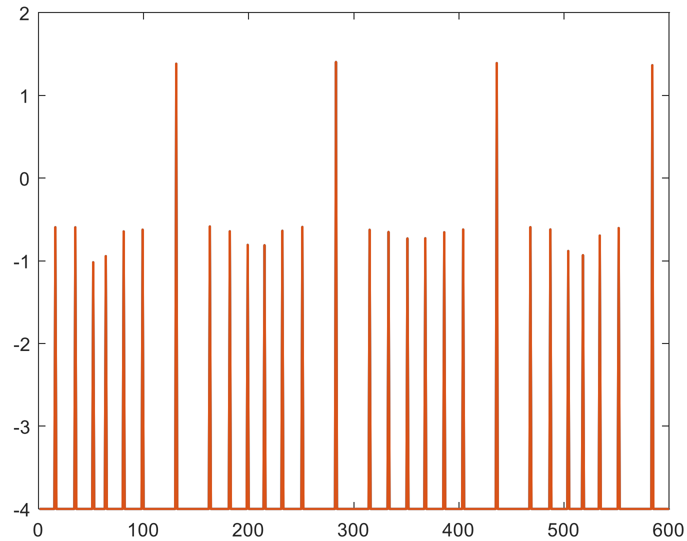
Figure 1: Growth distribution by size



Note: The sample excludes settlements that exit (merged by other bigger neighbours by the end period.). Each bin has about equal amount of built-up in year 1975. The numbers show the median growth rate.

Figure 2: Model simulation: employment

Employment around a circle (600 locations)



Employment on a disk (hexagonal lattice, 1147 locations)

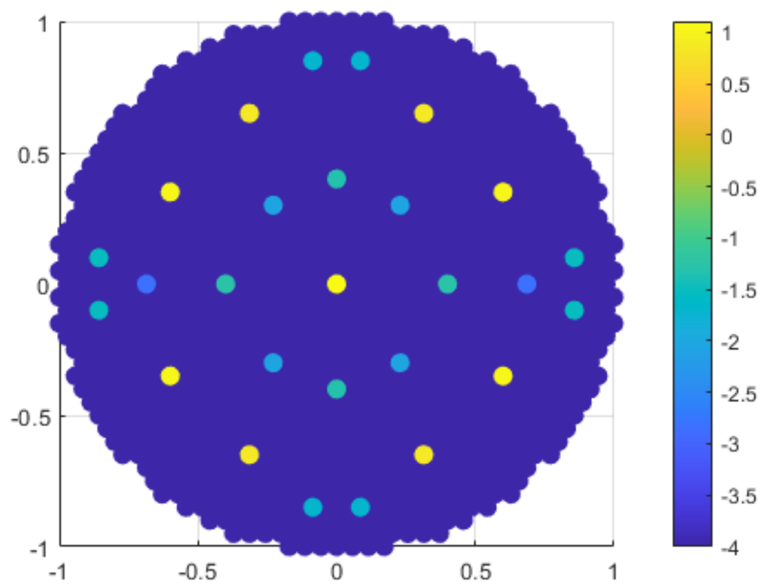
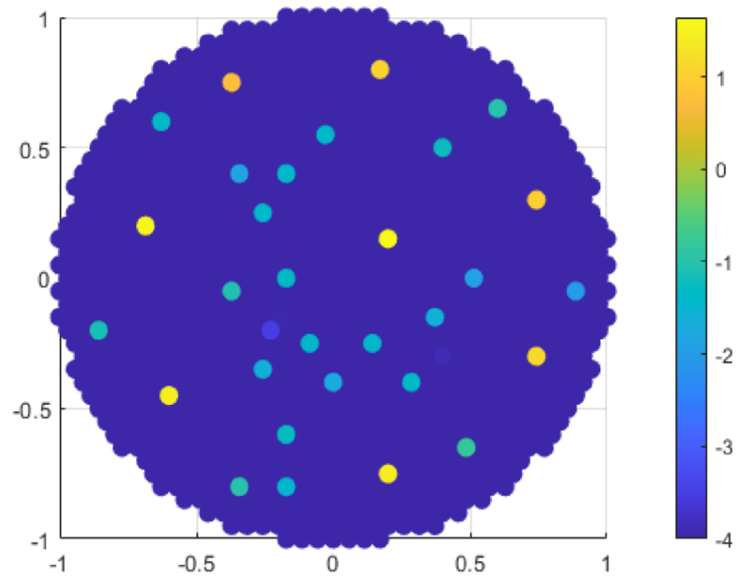


Figure 3: Employment and own- and cross-type effects

Employment on a disk with productivity variation



Proximity and employment; own - and cross-sector effects

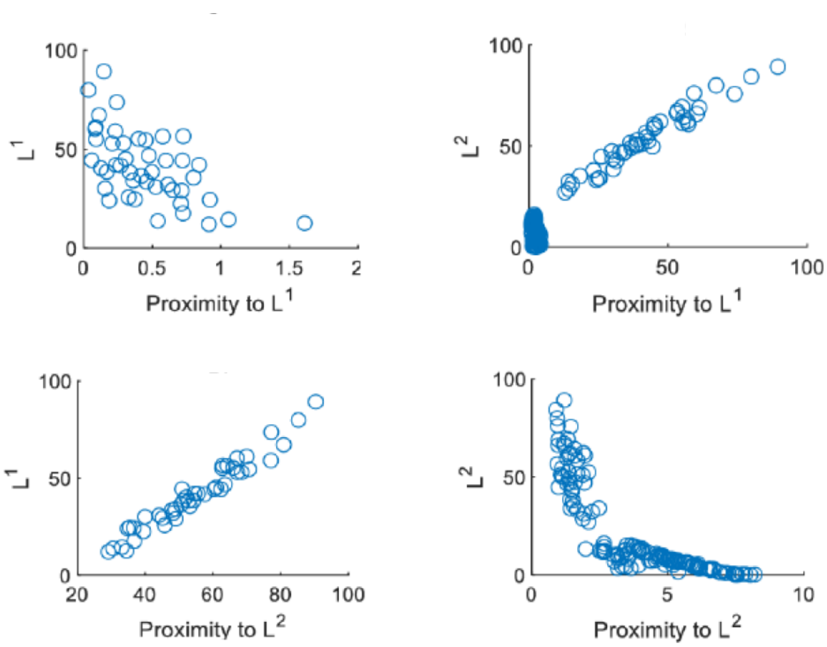
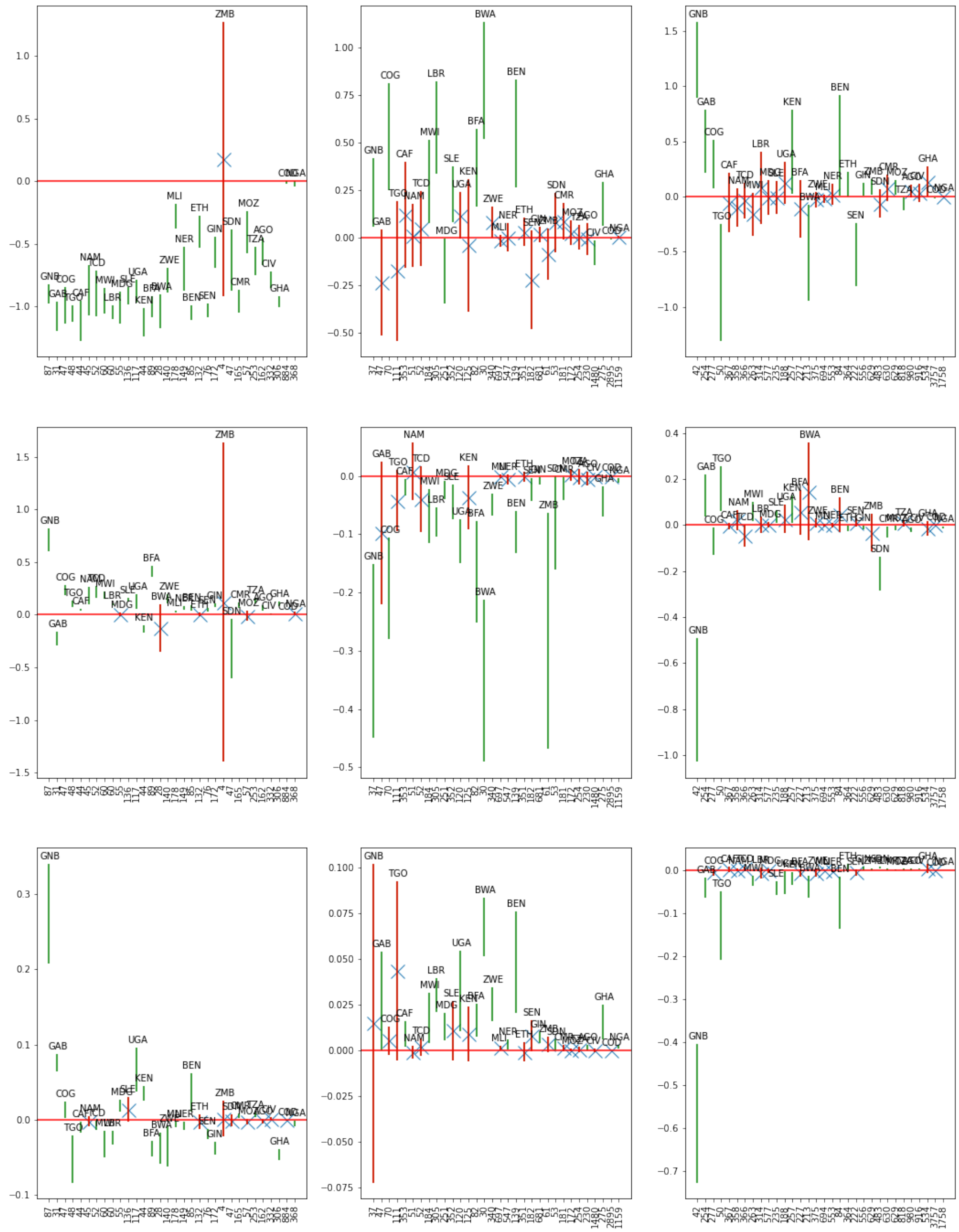
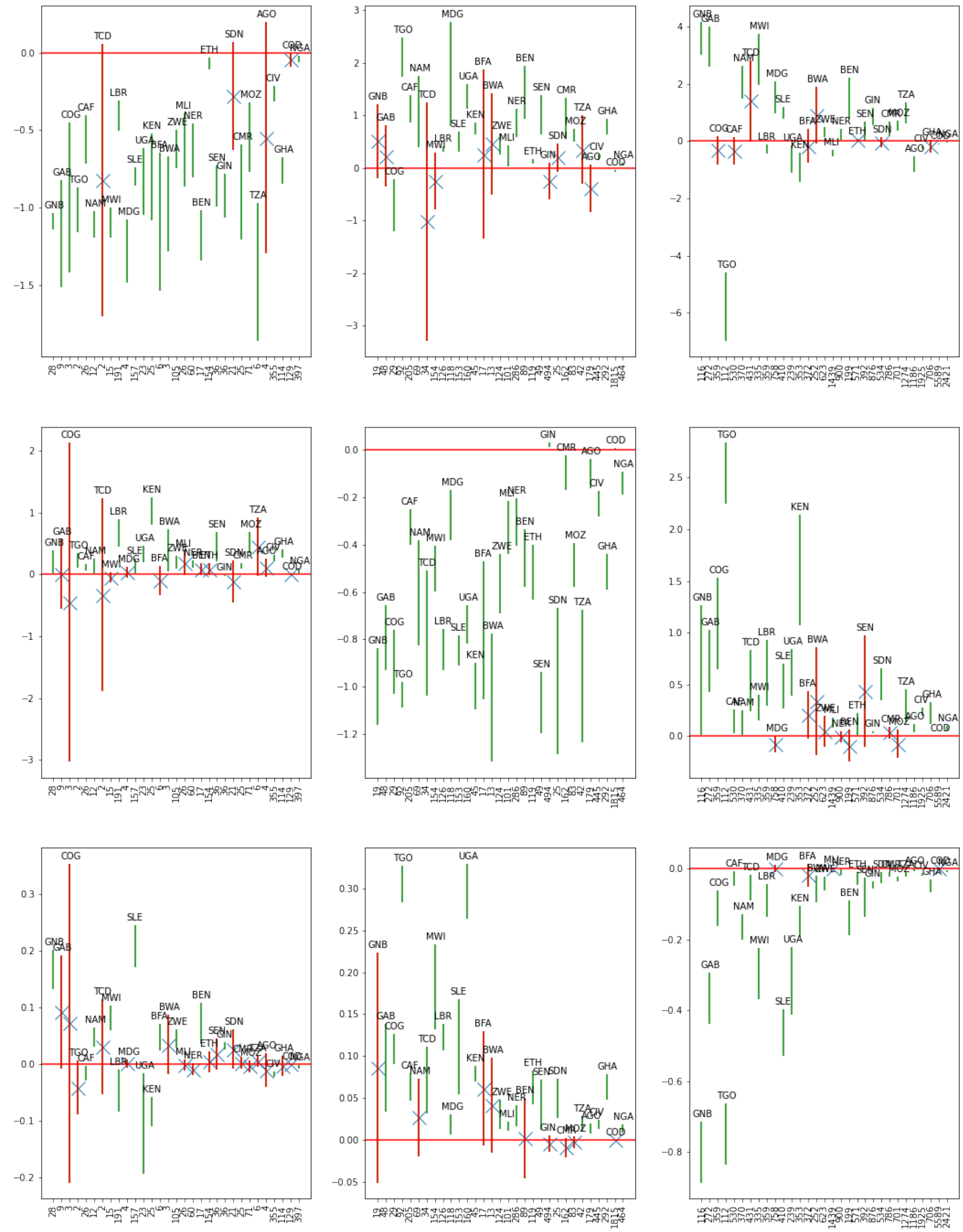


Figure 4: Cross sectional results 2014. Effects by neighbours by type (OLS). No initial levels



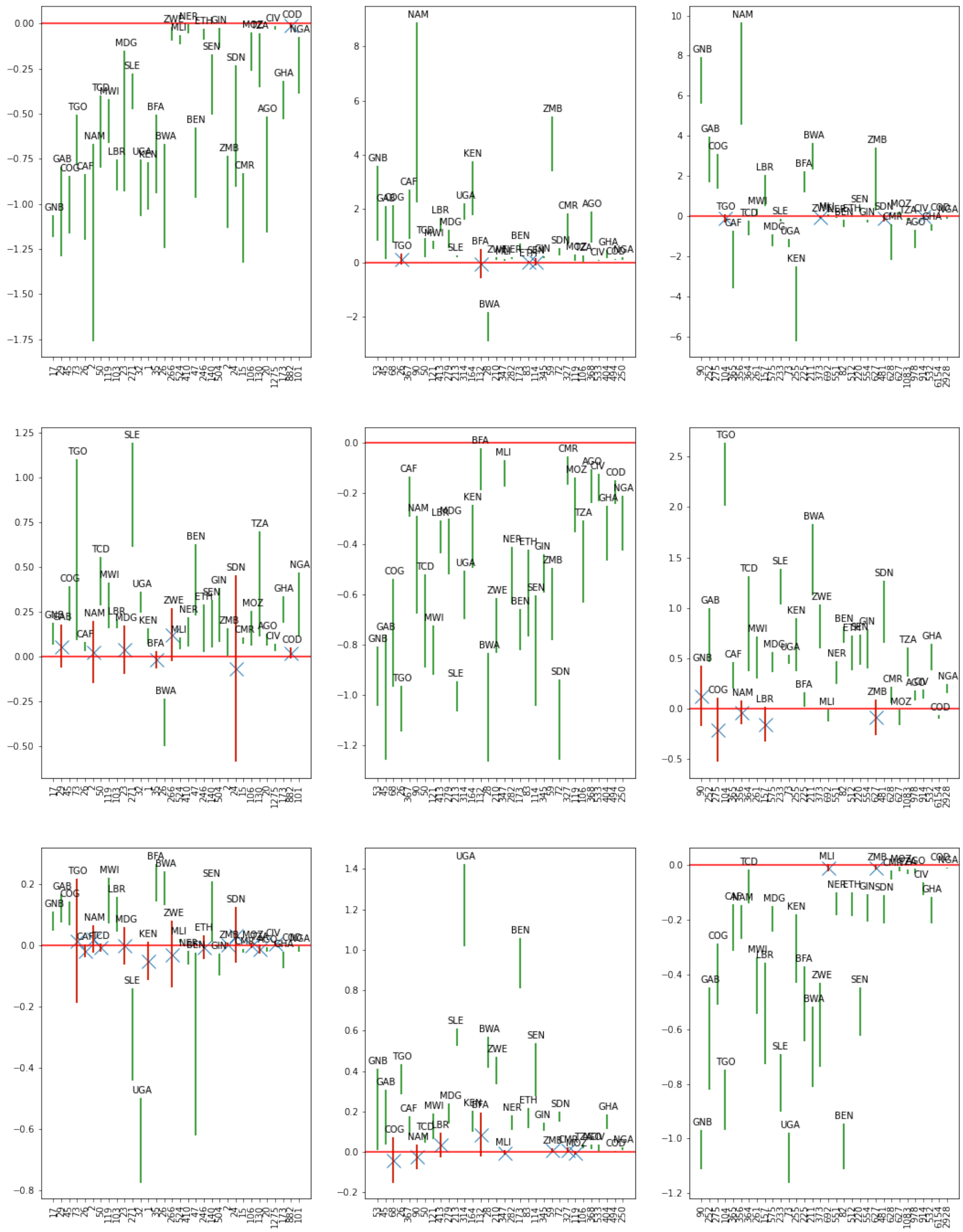
Note: The figure shows the error bars of the OLS estimates by countries. The location of the error bars on the x-axis is based on the ranking of the total built-up area, with larger countries shown to the right of the figure. The positions of panels correspond to the positions of the coefficients in equations 9 to 11, with each equation being a column and each row a city type. In searching for the optimal hierarchy of cities (bin split), we minimize the sum of squared residuals (RSS). In the search, geographic controls are included.

Figure 5: Growth 1975-2014. Effects by neighbours by type (OLS)



Note: The figure shows the error bars of the OLS estimates by countries. The location of the error bars on the x-axis is based on the ranking of the total built-up area, with larger countries shown to the right of the figure. The positions of panels correspond to the positions of the coefficients in equations 9 to 11, with each equation being a column and each row a city type. In searching for the optimal hierarchy of cities (bin split), we minimize the sum of squared residuals (RSS). In the search, geographic controls are included.

Figure 6: Growth 2000-2014. Effects by neighbours by type (OLS). Controlling 1975 and 1990 initial built-up levels



Note: The figure shows the error bars of the OLS estimates by countries. The location of the error bars on the x-axis is based on the ranking of the total built-up area, with larger countries shown to the right of the figure. The positions of panels correspond to the positions of the coefficients in equations 9 to 11, with each equation being a column and each row a city type. In searching for the optimal hierarchy of cities (bin split), we minimize the sum of squared residuals (RSS). In the search, geographic controls are included.