

Cities as Engines of Growth: Evidence from a New Global Sample of Cities

Paul Collier

Patricia Jones

Dewi Spijkerman

University of Oxford

28 March 2018

*This work was supported by the Global Research Program on Spatial Development of Cities at London School of Economics and Oxford University, which is funded by the Multi-Donor Trust Fund on Sustainable Urbanisation of the World Bank and by the United Kingdom Department for International Development. The authors may be contacted at patricia.jones@economics.ox.ac.uk.

ABSTRACT

This paper presents the first global analysis of the size and sources of agglomeration economies in the developing world. Using establishment data from more than 12,000 firms in 51 low-income and middle-income countries, we estimate the productivity effects of increased city size, population density, and market potential across 111 cities. To do this, we employ a new measure of city size and density based on LandScan data—a global dataset that estimates the world’s population at approximately 1 km resolution. Importantly, we use the same threshold level of population density (1500 people per square kilometre) to define “urban” areas in each country rather than relying on arbitrary, administrative boundaries. We reveal several important findings about the economic benefits of urbanization. First, we find strong evidence that urban productivity rises with city scale in developing economies. After controlling for country, industry, and year fixed effects, we estimate an output elasticity with respect to city size of between 0.067 and 0.080. This implies that urban productivity increases by about 5% to 6% with each doubling of city population. Second, we find evidence that agglomeration forces play an important role in expanding formal sector employment in cities throughout the developing world. Third, we reveal that African cities are not generating the same level of benefits—in terms of increased productivity, wages, and employment generation—as cities in Asia and Latin America. And, finally, we find evidence that the choice of density measure matters when estimating agglomeration effects.

JEL Codes: D24, O12, O14, O18

I. Introduction

Africa is urbanizing fast. Over the past twenty-five years, the number of urban dwellers throughout the continent has more than doubled, rising from 133 million in 1990 to 360 million today (UN Population Division, 2016). By 2040, more than half of all Africans will be living in urban areas. How will this urbanization process affect the average African worker? If African cities increase worker productivity, average wages should rise as the urban share of the population increases.¹ There is ample evidence from the rest of the world that cities generate many benefits that raise firm productivity (see, for example, the review by Rosenthal and Strange, 2004).² Many of these benefits increase with scale: bigger cities generate larger productive advantages than smaller towns and rural areas.

How do cities do this? Two primary channels are highlighted in the urban economics literature (Duranton and Puga, 2004; Puga, 2010; Brueckner, 2011). First, cities enable firms to operate at a larger scale and take advantage of the productivity benefits that arise from increased division of labour. That is, urbanization fosters *scale economies*. Second, cities facilitate the sharing of resources between firms and workers. Shared labour markets, for example, make it easier for firms to hire new workers without spending a lot of time and money searching for the right candidate. Similarly, shared transport links make it easier for workers to reach jobs and firms to reach their customers. Perhaps the most important economic benefit of cities is that they create learning opportunities for workers. Recent studies reveal that urban workers who live in larger cities learn more on the job than urban workers who live in smaller cities or towns (De la Roca and Puga, 2017). Such benefits—commonly referred to as *agglomeration economies*—arise as a result of the external, urban environment rather than from individual firm or worker characteristics. This creates an opportunity for policymakers to influence city-level productivity and urban incomes through policies that promote agglomeration effects.

¹ Urbanization can have a positive effect on rural incomes as well. There may be backward linkages which increase the demand for agricultural products (Cali and Menon, 2009) and/or rural workers may end up with greater land per person.

² More recent studies include De la Roca & Puga, 2017; D’Costa & Overman, 2014; Mion and Naticchioni, 2009; Combes, Duranton, and Gobillon, 2008; Di Addario and Patacchini, 2008; and Yankow, 2006.

Although numerous studies have examined the size and sources of agglomeration economies in advanced economies, much less is known about the productive benefits of cities in developing countries. Only recently have geo-referenced micro datasets become available enabling researchers to estimate agglomeration effects in developing countries. The evidence from this preliminary research suggests that, in several cases, the earnings premium for urban workers in the developing world is larger than it is for urban workers in advanced economies. While urban workers in the United States, earn about 30% more than their rural counterparts (Glaeser and Maré, 2001), the urban-wage premium is 45% in China, 122% in India, and 176% in Brazil (Chauvin et al, 2016). In Sub-Saharan Africa, the estimated urban wage premium is somewhat smaller. New research from Jones et al (2017) estimates an urban wage premium of 6% in Nigeria, 11% in Tanzania, and 9% in Uganda when a full set of worker controls are included in the wage equation.

If these wage gaps reflect genuine productivity gaps, urbanization is likely to play a crucial role in raising aggregate productivity in developing countries. The important question to policy makers is which policies, if any, are associated with larger agglomeration economies? New evidence on Chinese cities reveals significant heterogeneity in both the size and sources of agglomeration economies (Glaeser and Xiong, 2017). Agglomeration economies are larger in bigger cities, more-educated cities, and in cities with more knowledge-intensive industries. But these characteristics are outcomes, not policies. The economic size of a city, the education level of its workforce, and the composition of its industries are determined by underlying characteristics both of the city, and of the country in which it is located. In this paper, we focus on those city characteristics in low-income and middle-income countries over which policymakers have a degree of control, notably connectivity and market access: to date, there is little evidence on how such factors affect urban productivity in such countries. We partially fill this gap by examining both the size and sources of agglomeration economies in a large number of developing countries. To do this, we use the World Bank Enterprise Surveys (ES) to construct a global sample of more than 12,000 firms in 111 cities. We supplement these data with a variety of spatial variables and city characteristics. These include city-specific measures of average income (estimated using night lights data), road density (estimated using Google maps), market potential (estimated using LandScan data),

urban infrastructure (estimated by the average percentage of firms sales lost due to power outages), and the business environment (estimated by the average percentage of senior management time spent dealing with government regulations).

Using this global sample, we reveal several important findings about the economic benefits of urbanization. First, we find strong evidence that urban productivity rises with city scale in developing economies. As illustrated in Figure 1, there is a clear positive correlation between city size (in terms of population) and value-added per worker. Importantly, this relationship is not explained simply by the fact that wealthier cities have larger populations. After controlling for country, industry, and year fixed effects, our estimated output elasticities with respect to city size range from 0.067 to 0.080 using the LandScan estimates of city population. This implies that urban productivity increases by about 5% to 6% with each doubling of city population³. This bodes well for Africa's productivity potential, since it will reach its peak rate of shift of its aggregate population from rural to urban locations over the next two decades.⁴ Perhaps most interesting is the finding that agglomeration effects in developing countries are about the same size as those estimated for advanced economies. In advanced economies, the estimated output elasticity with respect to city size ranges from 2% to 10% (Duranton, 2015).

Second, we find strong evidence that establishment size (defined as the number of permanent workers per establishment) rises with population density. This is an important finding given the recent evidence that scale economies are intricately linked to both diversification and structural change in developing economies (Buera et al, 2011; Buera and Kaboski, 2012). Our results indicate that establishment size rises by about 20% with each doubling of population density within a city. Importantly, this result is stronger for African cities than it is for cities located in other developing regions, despite the smaller scale of African enterprises. As is well known, African enterprises are significantly smaller than those

³ We estimate the output elasticity with respect to density as $2^{0.084-1} \approx 6\%$ and $2^{0.0461-1} \approx 3\%$.

⁴ The rate of shift of the aggregate population is evidently constrained both when the host urban population is small, and when the supplying rural population is depleted. Hence, *ceteris paribus* it is at its peak when the rural and urban populations are of similar size.

elsewhere, even after controlling for differences in the level of economic development across countries (Iacovone, Ramachandran and Schmidt, 2014). Indeed, in many African countries—like Uganda—the modal size of enterprise is just one.

However, as for nearly all the research on developed economies, our dataset is confined to formal firms: it does not cover informal, micro enterprises such as those that predominate in most of Africa. Hence, we are unable to estimate agglomeration economies for these informal enterprises.⁵ But, as recent research has revealed, many informal enterprises are not meaningfully ‘firms’: organizations with a management that is motivated to use resources productively. More properly, they are default options for workers unable to find opportunities for earning a living by means of wage employment. That is, a large number of informal enterprises are run by ‘reluctant’ entrepreneurs (Banerjee and Duflo, 2005, 2011; Collier, and Jones, 2016).

Third, we examine the importance of connectivity—the ability of firms, consumers, and workers in a city to interact with each other—in raising urban productivity. Well-connected cities increase the scope for both economic and social interactions. Such interactions have the potential to raise urban productivity through a variety of channels, including larger knowledge spillovers (see Duranton, 2008; Duranton, 2015). In this study, we use two broad measures of connectivity: 1) the average road density within a city; and 2) the De la Roca and Puga (2017) measure of market potential. Both measures provide a rough estimate of the ease in which firms, consumers, and workers can interact with each other within a city. Our findings suggest that more connected cities—in terms of both better transport networks and more evenly distributed population density—have higher productivity.

Several recent studies have examined the potential benefits of urbanization in the developing world (Duranton, 2016; Howard, Newman, and Tarp, 2016; Jones et al, 2017; Kriticos and Henderson, 2017). However, these studies focus on just a handful of countries.

⁵ As part of this research project, we are investigating the differential impact of agglomeration on formal and informal sector firms in Uganda. This research is currently in progress.

By contrast, we present new evidence on the cross-country variation in agglomeration economies using establishment data from over 12,000 firms in 111 cities. Our study contributes to the urban economics literature in two respects. First, we employ a new measure of city population and density to measure agglomeration economies. This measure uses the same threshold level of population density (1500 people per square kilometre) to define what is meant by “urban” rather than relying on arbitrary, administrative boundaries. Second, we investigate the impact of urbanization on both establishment level productivity and scale using a unique global data set.

The remainder of the paper is structured as follows. Section 2 describes the different data sources and estimation techniques used for analysis. Section 3 discusses our main empirical finding and their relevance to policy makers in developing countries.

2. Data and Estimation

The estimation of agglomeration economies across multiple cities requires two types of data: 1) standardized microeconomic data that reflect local outcomes (e.g., firm-level output data or individual-level earnings data); and 2) standardized population data at the city-level. Until recently, the combination of such data were not available for many low-income countries. Most census data in low-income countries, for example, contain population numbers but they do not include earnings or income data for individual workers. While household surveys usually do collect such information, these data are not often representative of workers at the city-level. Similarly, most industrial surveys collect production data that are nationally representative but not representative of firms that operate in different cities.

Urban Productivity

Fortunately, there is a set of surveys—the World Bank Enterprise Surveys (ES)—that collects establishment-level data that are representative at the city-level. While the ES was not designed for urban analysis, the fact that economic activity tends to be clustered in urban

centres has resulted in a sampling strategy where the data are usually representative at the city-level.⁶ In addition, the same production module is implemented in each country so the output data are comparable across countries. These two characteristics make the ES data ideal for our purposes. To estimate urban productivity, we chose cities (and firms) based on four criteria. First, the city has a population of at least 300,000. Second, the city is located in either a middle-income or low-income country. Third, the ES program interviewed at least 30 firms in the city. And, finally, the data collected is representative of firms located in the city. Our final sample consists of about 12,000 establishments located in 111 cities. These include 43 cities in Sub-Saharan Africa, 28 cities in Latin America, and 40 cities in Asia. Table 1 provides a complete list of both the cities and countries included in our global sample.

Several steps are taken before estimating the cross-country regressions. First, we remove outliers. To do this, we transform each of the production variables as $\ln(x+1)$ and then group these variables by survey year, country, and sector (broadly defined as “manufacturing” and “services”). “Next, we calculate means and standard deviations of these transformed variables within each group. Observations that are more than three standard deviations away from the mean are then marked as outliers and turned into missing” (World Bank, 2015). Finally, we need to convert all monetary values to a common currency-year. To do this, we first convert the production variables into U.S. Dollars (USD) using the official exchange rate from the World Development Indicators (WDI).⁷ Next, we deflate these values to 2009 values using the GDP deflator for the United States from the relevant reference fiscal year.⁸ All monetary variables are expressed in 2009 US dollars.

Table 2 presents the descriptive statistics of the establishments in our global sample (panel a) as well as descriptive statistics of the cities where they are located (panels b and c). As expected, we find that African firms are smaller than those in other regions. However, given the truncated nature of the ES data (only formal sector firms with 5+ permanent employees

⁶ The ES surveys are stratified to represent the geographical distribution of economic activity within each country. Typically, this stratification coincides with collecting establishment-level data from 2-3 major cities.

⁷ WDI indicator code: PA.NUS.FCRF

⁸ WDI indicator code: NY.GDP.DEFL.ZS

are interviewed), this is not due to contamination by a large informal tail: formal firms in African cities are significantly smaller than those in other developing cities. Evidently, in terms of all establishments, whether formal or informal, the “average” African firm in our sample is much *larger* than the average African establishment. To assess how important this truncation is for our African sample, we use a new dataset by Bento and Restuccia (2017) to calculate average establishment size in Africa when informal enterprises are taken into account.⁹ According to these data, the average enterprise in Africa has only six workers, which is about one tenth the size of the average firm included in the ES sample. This suggests that while our results are representative of *formal sector firms* operating in African cities, we cannot draw inferences on how urbanization might affect informal enterprises. We suggest that far from being a weakness, it makes our results easier to interpret for policy. *Informal enterprises might well proliferate in precisely the conditions in which proper firms struggle to grow.*

Like other studies, we find that African firms tend to be younger and less productive than firms in other developing countries but not necessarily cheaper in terms of labour costs (Iocovone et al, 2017; Lall et al, 2017). As reported in panel a of Table, 2, the average African firm is 21 years old which is half the age of its Asian counterpart and 15 years younger than its competitor in Latin America. This is a disadvantage if there are important learning spillovers associated with entrepreneurship. In fact, the senior management of African firms has about 3 years less experience than senior management in Asian firms and 6 years less experience than the management Latin American firms (see panel b).

The smaller scale of African firms, however, is not simply a byproduct of their youthfulness. Even among older, established firms (i.e., those that have been in operation for at least 40 years), African firms employ three times fewer workers than similarly aged firms elsewhere. As expected, Africa’s smaller scale is associated with lower productivity. African firms are

⁹ The Bento and Restuccia (2017) dataset estimate average firm size for 134 countries based on the full size distribution of firms in each country. To do this, they look at hundreds of secondary sources (e.g., reports on labour force surveys, household surveys, and economic censuses) which document the size distribution of firms.

about half as productive (in terms of value-added per worker) as their Asian counterparts. Despite their lower productivity, African firms tend to pay higher wages, suggesting that African cities may not be a cheap source of international labour. The average (annual) wage in African cities is 3,911 US dollars whereas it is only 3,627 US dollars in Asian cities. Importantly, Africa's productivity gap is not being driven by a lower capital-labour ratio or a higher percentage of unskilled workers in African firms relative to Asian firms. Instead, the descriptive statistics suggest that Africa's lower firm productivity may be related to city and spatial characteristics.

Defining City Boundaries in the Global Sample

The choice of how to define a city's boundaries is important in constructing a consistent measure of population and density across different urban agglomerations. The administrative boundaries of a city—as defined by local political authorities—can vary widely in terms of both average population density and land area. In this study, we use a combination of LandScan data and night lights data to define city boundaries. Essentially, we define urban areas in terms of a threshold level of population density (1500 people per square kilometre). To our knowledge, we are the first to use this methodology to estimate agglomeration economies across a global sample of cities.

LandScan, created by Oak Ridge National Laboratory, is the finest resolution global population distribution data that is currently available. Created for global disaster relief, this dataset estimates the ambient (average over 24 hours) population at approximately 1 km resolution (30" X 30" arc-seconds).¹⁰ It does this by combining satellite imagery (that details the extent of built cover in an area) with population census data (that estimates the population that lives in that area). The ambient population is then derived by an algorithm which essentially assigns people to buildings. The assumption is that people do not necessarily stay in the same place throughout the day where they answered the census questionnaire but, instead, occupy nearby buildings for work and recreation.

¹⁰ See <http://web.ornl.gov/sci/landscan/index.shtml> for data availability and documentation.

As pointed out by Henderson, and Nigmatulina (2016), the main limitation of these “data is that although the built cover information is consistent, since it is derived from satellites, the census sub-units used may vary across countries.” While LandScan does use other sources, like topographic maps or roads, to increase the accuracy of its estimates, these sources may not be available for every city in our sample. Finally, it is unclear from the LandScan website what weights are used to assign the local population to nearby buildings during any 24-hour period. Nevertheless, the LandScan data are the best source of information that we have on global population distribution.

We define urban agglomerations using the same approach as developed by Henderson and Nigmatulina (2016). Several steps are taken. First, the location of each city in our sample is found using a digital map.¹¹ Second, we create a 40 km buffer zone around each city. Within this buffer zone, we use night lights data to locate the brightest pixel and define this location as the city centre. Third, urban agglomerations are defined by merging all areas extending from the city centre which meet a threshold level of population density (1500 people per square kilometer). For six Chinese cities, we apply a slightly higher threshold in order to separate cities that are located near to each other and would merge into one urban agglomeration if we applied our 1500 per square kilometer cutoff.¹² Finally, the unions that generated the boundaries of some cities had to be “clipped” to remove water segments (because the city was located on a coast, lake, or river) or maintain national boundaries (because the city was located near a country border).

We find many examples of urban areas where there is a large mismatch between the city extent as defined by administrative boundaries relative to the threshold level of population density. Some administrative boundaries underestimate the “true” city extent as defined by our LandScan boundaries. In cities like Accra (Fig. 1) and Dhaka (Fig. 2), we find that urban areas defined by administrative boundaries (the areas shaded in red) are much *smaller* than

¹¹ We use the World Cities shapefile from the ArcGis website.

¹² The inclusion or elimination of these cities in our global sample does not change our results in any meaningful way.

those defined using our threshold level of population density (the areas shaded in yellow). The opposite is true in cities like Dar es Salaam (Fig. 3) and Kigali (Fig. 4). In these cities, the urban areas defined by administrative boundaries (the areas shaded in red) are much *larger* than the urban areas defined using our threshold level of population density (the areas shaded in yellow). These examples highlight the potential measurement error that may arise when comparing city population across countries using administrative boundaries.

Estimating City Population and Density

Using our LandScan measure of city boundaries, we next estimate city population and density for each urban agglomeration in our sample. We define city population as the total (ambient) population in the urban agglomeration. Similarly, we define population density as the average ambient population per square kilometer within the urban agglomeration. When comparing our estimates of city population to those compiled by the United Nations (UN) Population Division, the UN and LandScan estimates of city population are highly correlated (the correlation coefficient is 0.89).

Panels b and c of Table 2 presents the descriptive statistics of our population measures. The most important takeaway is that African cities are smaller in terms of overall population but have higher population density than cities in other regions. Africa's increased density is likely due to its large slum population. A recent study by UN Habitat (2016) estimates that 56% of all urban residents in Africa live in slums whereas this percentage is much lower in other regions. For example, 31% of urban residents live in slums in South Asia, 26% live in slums in East Asia, and 21% live in slums in Latin America and the Caribbean (UN Habitat, 2016). Although African cities achieve population density, they have done so without density in residential construction: the region's low-income residential areas consist largely of single-storey structures. Its cities have achieved density of population by extreme sacrifice of floor-space. In effect, African cities have sacrificed livability in order to achieve density.

Further, in many African cities—like Dar es Salaam, Kampala, and Nairobi—large slum neighborhoods are located on prime land near the central business district (CBD). However, this is not because such a location is economically efficient: the sacrifice of livability is not

the price paid for productivity. The explanation for their location is not market forces propelling the spatial allocation of activity towards an efficient distribution, but historical inertia due to the lack of effective land markets. Indeed, despite their apparent proximity to the CBD, the residents of these slum neighborhoods may still be disconnected from the economic heartland of the city in wealthier residential areas and commercial districts due to poor (or expensive) transport links. In other words, there may not be a positive relationship between density and connectivity.

De la Roca and Puga (2017) Measure of Connectivity

The key measure to estimate the degree of connectivity within a city is the De la Roca and Puga (2017) index, hereafter DRP measure. This measure estimates the average interaction (“exposure”) of a person who works and lives in a city to other people within a certain radius (say 5km or 10km) of that person. It is calculated by tracing a circle of radius 10 kilometres around each 1 x 1 kilometre cell in the city, counting the number of people in that circle, and then averaging this count over all cells in the urban area, weighting each cell i by its population, w_i . This yields a measure of the average number of people within 5 kilometres or 10 kilometres of the “average” person in the city. To account for commuting costs that increase with distance, each cell count can be adjusted by a discount factor (say, $e^{-0.5d}$ or e^{-d} where d is the distance to each cell j from the base cell i). The calculation of the DRP measure for each cell i within the city with $e^{-0.5d}$ discount and an exposure radius of 10 kilometres is given in equation (1). The weighted average is given in equation (2).

$$Puga_i = \begin{cases} \sum_j e^{-0.5d_{ij}} * population_j & \text{if } d \leq 10 \text{ km} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$Puga = \sum_i w_i * Puga_i \quad (2)$$

where

w_i = the share of city population that lives in cell i

d = the distance to each cell from cell j from the base cell i

While the DRP measure is highly correlated with a city's population, its main advantage over standard population measures is that it picks up "polycentric" cities—that is, cities with an uneven distribution of population density over space. As pointed out by De la Roca and Puga (2017): "the simple population count for these polycentric urban areas tend to exaggerate their scale, because to maintain contiguity they incorporate large intermediate areas that are often only weakly connected to various centres."

Panel c of Table 2 presents our estimated DRP measures. As expected, these estimates are lower than the regional averages of city population that do not take into account the lumpy nature of economic density over space. Whereas in panel b, mean population density was highest in African cities, the DRP regional averages reported are lowest in African cities. This is consistent with the evidence presented by Henderson and Nigmatulina (2016) who find that, after controlling for total urban population and GDP per capita, urban residents in Africa have about 40% fewer (potential) interactions than urban residents in Asia and Latin America. One reason for this lower connectivity is the common occurrence of "leapfrog" development in many African cities—that is, the development of new parcels of urban land that do not connect with already developed land but, instead, create open, disconnected spaces in the built environment. New research based on satellite imagery reveals that in several major African cities, like Accra, Kigali, Lagos, Nairobi, and Maputo, leapfrog patches account for more than 40% of the increase in built-up area that occurred over the period 2000 to 2010 (Lall et al, 2017).

Such leapfrog development leads to lumpy population density and spatially disconnected cities. Spatial fragmentation matters for three reasons. First, fragmentation makes the provision of public goods and urban infrastructure more expensive because there is more land area to cover. Second, fragmentation reduces the (potential) size of agglomeration economies by making it harder for firms to acquire land near the CBD where the concentration of firms is often highest. And third, fragmentation reduces the size of the market by increasing the physical distance between established firms (often located near the CBD) and new consumers and firms located on leapfrog plots.

Spatial factors can affect access to regional markets as well. Being located near other cities is an advantage to urban firms, especially when urban consumers have substantial purchasing power. As reported in panel c of Table 2, the distance between cities is much shorter in Asia than it is in Africa and Latin America. For Asian firms, the nearest market (i.e., the nearest city with at least 300,000 inhabitants) is about 234 kilometres away. By contrast, the nearest market to African and Latin American firms is over 300 kilometres away. Similar regional differences arise when we estimate the distance from each city to the nearest port. On average, African cities are located 548 kilometres from a port whereas Asian and Latin American cities lie 277 and 464 kilometres away, respectively.

Measuring Agglomeration Effects

While there are several ways in which to measure the economic benefits of cities, we follow the standard approach in the literature by measuring agglomeration effects through changes in establishment-level productivity or wages. Economic theory predicts that, among homogenous establishments, productivity should rise with city population (and density) due to agglomeration effects. To measure the size of these effects, we estimate several variants of the following equation:

$$\log y_{ijc(i)st} = \alpha \log \text{Pop}_{c(i)} + X_{ijt} \beta + A_{c(i)} \gamma + \delta_j + \theta_s + \mu_t + \varepsilon_{ijcst} \quad (1)$$

where the dependent variable, y , is the (natural) logarithm of value-added per worker (or the average wage) of establishment i in industry j located in city c of country s in year t . In this regression, α , represents the elasticity of output (or wages) with respect to city size or density. This coefficient is estimated after controlling for a vector of both establishment-level characteristics, X , and city-level characteristics, A . In addition, the regression controls for industry, δ , country, θ , and year, μ , fixed effects.¹³

Ideally, we would like to compare the productivity of identical establishments randomly distributed across cities that differ only in terms of their population size. Instead, we observe

¹³ We adopt a similar specification as Duranton (2016).

the productivity of establishments that *choose* to locate in different cities that vary in terms of both observable and unobservable characteristics. Following Duranton (2016), we have indexed $Pop_{c(i)t}$ in equation (1) by $c(i)t$ to highlight the fact that firm location is a choice variable.

Indeed, there are two well-known biases that can occur when estimating equation (1). The first type of bias arises when there are unobservable firm characteristics (e.g., managerial ability) correlated with city population. For example, larger cities may attract more talented entrepreneurs than smaller cities because talented entrepreneurs have more to gain from working in larger cities. To address this concern we include several firm “quality” controls (e.g., employment size, capital-labour ratio, firm age, and the proportion of unskilled workers employed by the firm) as well as industry fixed effects.

The second potential bias occurs when some city effects not included in the vector of city characteristics are correlated with city population and establishment-level productivity. For example, larger cities with better amenities may attract more skilled workers than smaller cities with fewer amenities. Econometrically, this type of bias is equivalent to that which arises from reverse causality—that is, cities grow larger because they attract more productive workers rather than larger cities generate higher productivity through increased scale and agglomeration economies.

To address these econometric concerns, we adopt two approaches. First, we include a set of city characteristics that capture the variation across urban areas in the quality of both public infrastructure and the local business environment. Second, we implement an instrumental variable (IV) strategy where we instrument a city’s current population by its historical population in 1930.¹⁴ The exclusion restriction implied by this instrumental variable regression is that historical population levels over 80 years ago have no impact on current firm performance, except through their impact on current population levels. While this is a plausible assumption, it is not difficult to imagine several alternative scenarios where past

¹⁴ This is the earliest date for which population estimates are available for most African cities.

population values can be linked (via path dependence) to current productivity levels. In fact, history reveals several examples of cities where geographical factors—like portage sites¹⁵—have played an instrumental role in a city’s early formation and its subsequent economic development. For example, a city’s climate—particularly if it is located in the tropics—can affect its long-run institutional development (Acemoglu, Johnson, Robinson, 2001) and rate of human capital formation (Gallop et al, 1999), potentially affecting both past and current rates of investment and growth. Due to path dependence, these cities continue to grow (or stagnant) long after their initial geographical advantages (or disadvantages) have become obsolete (see Davis & Weinstein, 2002; Bosker et al, 2007, 2008; and Miguel and Roland, 2011; Bleakley and Lin, 2012). To address these concerns, we include a city’s latitude as an additional instrumental variable.¹⁶

3. Estimation Results

The positive correlation between a country’s rate of urbanization and its per capita income is well established (see, for example, World Bank, 2009; Henderson 2010). While much of the growth-enhancing effects of urbanization are linked to structural change (that is, agricultural workers moving to urban areas in order to work in higher productivity jobs in manufacturing and services), there is now growing evidence that the urban environment can have an independent effect on productivity as well. Cities—through increased economic density—can foster both scale economies and agglomeration economies. How well cities in different regions are able to generate such benefits, however, remains an open question. In this study, we take up this question by examining both the size and sources of agglomeration economies among firms in a global sample of 111 cities.

Tables 3-8 present the main results of our study. In these tables, we report our estimated output and wage elasticities with respect to either city size and population density. Initially, these estimates are derived from simple OLS regressions where the dependent variable is

¹⁵ Portage sites occur where natural obstacles to water transport (like extreme water rapids) cause traders to disembark from their boats and haul their goods overland. The location of many cities in North America coincide with historical portage sites (see Bleakley and Lin, 2012).

¹⁶ We tested other instruments as well—like the distance from a city to the nearest navigable ocean, river, or lake—but none of these variables proved to be valid instruments.

either the logarithm of value-added per worker (defined as plants' annual sales minus raw materials and energy costs) or the logarithm of firm wage (defined as plants' annual labour costs divided by the total number of permanent workers). We estimate various specifications in which we incrementally add controls to our model. First, we control for city size (Table 3) or population density (Table 4). We then add firm and city characteristics (Tables 5 and 6) to control for cross-country variation in the quality of firms, urban infrastructure, and the local business environment. Next, we implement an instrumental variable (IV) strategy where we instrument a city's current population by its population in 1930 and control for additional geographical variables (Table 7). Finally, we investigate the importance of connectivity using two De la Roca and Puga (2017) measures (Table 8).

Table 3 presents the first set of regression results. In panel a, we compare the pooled OLS output elasticities based on our different measures of city size. In columns (1-3), we report the elasticities based on LandScan data and, in columns (4-6), we report the elasticities based on the UN data. All regressions control for country, year, and industry fixed effects. As reported in column (1), the estimated output elasticity for the full sample is 0.080 when we measure city size using the LandScan data, indicating that urban productivity rises by about 6% with each doubling of city population. The positive association between city size and establishment-level productivity holds for all regions, with African establishments (column 2) benefiting slightly more from increased city scale than those in Latin America and Asia (column 3). The coefficients are similar in magnitude when we estimate the output elasticity using the UN data. The estimated output elasticity for the full sample is 0.74, indicating that urban productivity rises by about 5% with each doubling of city population. The only substantial difference between the two sets of results is that the output elasticity is no longer significant for African firms when we measure city size using the UN population estimates.

In panel b, we report the estimated wage elasticities with respect to city size using our two measures of urban population. As reported in columns (1) and (4), the estimated coefficients are similar in magnitude when city size is estimated using either of the two population measures. For the full sample, the estimated wage elasticity is 0.083 (column 1) when population size is estimated using the LandScan data, and 0.086 (column 4) when population

size is estimated using the UN data. In addition, the estimated *wage* elasticities reported in panel b are similar in magnitude to the estimated *output* elasticities reported in panel a. This suggests that agglomeration forces have a similar impact on both wages and productivity across our global sample (as would be expected when labour markets are competitive).

Economic theory predicts that both city size and population density can affect establishment-level productivity. Larger cities coincide with increased market size while greater population density facilitates increased learning, sharing, and matching (see Duranton and Puga, 2004 for a review of this literature). In Table 4, we investigate the impact of *population density*¹⁷ on establishment-level productivity and wages. Surprisingly, none of the estimated output elasticities are significant and many have the wrong sign, particularly when we estimate population density using the UN population data (columns 4-6). In panel b, we report the estimated wage elasticities. Once again, few of the estimated coefficients are significant. The only exception is the estimated wage premium for workers in Asian and Latin American cities. One possible explanation for these weak results is that our density measure is not capturing the uneven nature of population density across many cities in our sample. We explore this possibility later in the paper.

Controlling for Firm & City Characteristics

There are two well-known biases that can occur when estimating equation (1) using an OLS estimator. The first type of bias arises when there are unobservable firm characteristics that are correlated with city population. The second potential bias occurs when some city effects not included in the vector of city characteristics are correlated with city population and establishment-level productivity. To address these econometric concerns, we adopt two approaches. First, we introduce several firm controls intended to capture observable differences across cities in firms' productive capacity. These include firms': 1) capital-labour ratio; 2) employment size; 3) age; and 4) the percentage of workers who are unskilled. Next,

¹⁷ Population density is defined as the average number of people per square kilometre who reside or work within the city. Each city's boundaries are defined using our threshold level of population density (1500 people per square kilometre). This means that the denominator of our density measure is the same whether we use the LandScan or UN measure of population.

we include a set of city characteristics. These include city-specific averages of: 1) the percentage of sales lost by urban firms due to power outages; and 2) the percentage of senior management's time spent dealing with business regulations; and 3) average road density. In addition, we control for two spatial variables: 1) the distance from the CBD to the nearest port; and 2) the distance from the CBD to the nearest market (defined as the nearest city with a population of at least 300,000).

Tables 5 and 6 present the regression results with the full set of firm and city controls. Panel a reports the output elasticities while panel b reports the wage elasticities. Several interesting results emerge from these regressions. First, in panel a, the estimated output elasticity falls by about half—but remains significant—when we control for differences in firm quality across cities. For the full sample, the coefficient on city size falls from 0.80 (Table 3, column 1) to 0.049 (Table 5, column 1). This result is consistent with the view that firms engage in spatial “sorting” across cities—that is, more productive firms choose to locate in larger cities. Second, the coefficient on city size for the Asian and Latin American sample loses significance when we add firm controls (column 3) but then becomes larger and significant once we control for the quality of urban infrastructure and the local business environment (column 6). This suggests that city characteristics in Asia and Latin America actually work to boost firm productivity.

A different pattern emerges for the African sample. Once we control for city characteristics, the output elasticity falls and becomes insignificant (column 5). This suggests that bigger cities are not generating larger agglomeration economies than smaller cities. Some city characteristics, like infrastructure, have a large negative effect on firm performance. On average, African firms lose four times the percentage of annual sales as firms in other developing regions due to power outages. Surprisingly, firms in cities with greater levels of human capital (in the form of increased years of management experience) or better road networks (in the form of increased road density) are not more productive than firms located in cities with lesser amenities. This result holds across all samples.

Table 6 reports the estimated wage elasticities with respect to city size. Once again, there is considerable heterogeneity by region in the estimated elasticities. For the African sample, we find no evidence that workers in larger cities receive a significant wage premium relative to workers in smaller cities. There is, however, one important caveat to this result. Our wage data represent the average wage paid by urban firms—not the individual wage received by urban workers—and, as a result, we are unable to control for a host of worker characteristics that are correlated with wages. How much this omission affects our results depends upon the extent to which worker characteristics are correlated with unobservable firm characteristics. If high-quality workers match with high-quality firms, the potential bias introduced by the omission of worker characteristics is likely to be small. On the other hand, if average wages at the firm-level mask large worker heterogeneity within firms, the potential bias could be large.

With this caveat in mind, it is important to point out that our estimated wage elasticities are similar in magnitude to those estimated for other regions. For example, De la Roca and Puga (2017) find an OLS elasticity of the earnings premium with respect to city size of 0.046 for French cities. Duranton (2016) is one of the few studies that explores this relationship using data from the developing world. Duranton's estimates of the wage elasticity for Colombian cities range from 0.046 to 0.11. While future research is certainly needed, our estimated wage elasticities fall well within the range of those estimated by others.

2SLS Estimation

Much has been written on the potential biases that can occur when estimating the impact of city size on productivity or wages (see Combes and Gobillon, 2015 for a good review of this literature). To address these concerns, we implement an instrumental variable (IV) strategy where we instrument a city's current population by its historical population in 1930. This identification strategy is valid, however, only if a city's past population is correlated with its current population but remains uncorrelated with current productivity levels. As discussed earlier, this assumption is violated when geographical advantages—like climate—lead to path dependence. To address these concerns, we include a city's latitude as an additional instrumental variable. Robustness checks indicate that both variables are valid instruments.

Table 7 reports the results from our two-stage least squares (2SLS) IV estimation. This table replicates Tables 5 and 6 but instruments a city's current population by its historical value in 1930 and its latitude. In panel a, we report the estimated output elasticities for both the full sample (column 1) and the two sub-samples (columns 2-3). For the full sample, the coefficient on log city population is 0.107 which is substantially larger than its OLS counterpart. As reported in Table 5, the OLS estimated output elasticity with respect to city size is 0.060. Both the OLS and 2SLS coefficients are significant at the 1% level. Similar to the OLS analysis, we find significant regional differences. In Asia and Latin America, firms in larger cities have higher productivity than comparable firms in smaller cities. On average, establishment-level productivity rises by 7.6% with each doubling of city size. Nearly identical results are found when we estimate the 2SLS wage elasticities for the full sample (column 4) and non-African cities (column 6).

By contrast, both the OLS and 2SLS results indicate that African firms are not benefitting to the same extent from increased urbanization as firms in other developing regions. Although the coefficient on log city population is significant when we exclude city controls from the regressions (Tables 3 & 4), the addition of these controls wipes out the city-size effect. This suggests that agglomeration forces in larger cities are not larger than those in smaller cities.

Connectivity Measures

As reported in Table 4, we find weak results when we measure the size of agglomeration forces using population density rather than city size. One possible explanation is that our density measure is not capturing the uneven nature of population density across many cities in our sample. We find some evidence to this effect. In Table 8, we replace population density by two De la Roca and Puga (DRP) measures: (1) the DRP measure at 10km; and (2) the DRP measure at 10km with a discount rate of $-0.5d$. The first DRP measure estimates the average number of people within 10 kilometres of the "average" person in the city while the second DRP measure adjusts each cell count by a discount factor $e^{-0.5d}$ where d is the distance to each cell j from the base cell i . Both measures take into account the potential "lumpy" nature of economic density within a city. Surprisingly, none of the DRP measures are

significant in the productivity equations (panel a). By contrast, we find some significant results in the wage equations (panel b).

One possibility is that economic density has an indirect effect on productivity through its effect on firm scale. That is, firms located in cities with greater economic density are able to grow larger due to an increased concentration of demand for their products. Next, we investigate how well cities in developing countries are able to increase firm scale. As reported in Table 9, we find evidence that the level of connectivity within a city—as estimated by one of the DRP measures or average road density—has a significant impact on firm size. This is an important finding given the recent evidence that scale economies are intricately linked to both diversification and structural change in developing economies (Buera et al, 2011; Buera and Kaboski, 2012). Our evidence suggests that both connectivity and the size of the market within a city matters in generating scale. Average firm size is larger in cities with higher average income (as estimated by the luminosity of night lights within the city). Larger firms are more productive, even after controlling for firm characteristics and country, industry, and year fixed effects. However, the effect of city income on firm scale is a little smaller in African cities than in cities elsewhere.

Hence, the relationships between the productivity of firms, and the connectivity and market size of the city in which they operate, is similar in Africa to that in other regions. However, what is radically different is both the average size of those firms, and the number of firms relative to the number of people living in the city. The formal firms found in African cities are smaller, and there are far fewer of them. An explanation for this apparent paradox comes from the analogy with biology. Think of the city as the habitat of the species, ‘firm’. All cities are sufficiently viable habitats that they support some members of the species, but there are large variations in the suitability of these habitats and they are manifested in large variations in the number and size of the firms that are able to survive in them. As with any natural habitat, the species proliferates to the point at which the marginal member of the species is just able to survive, while intra-marginal members make profits. Hence, comparisons of the firms observed in radically different habitats will not reveal large differences in how they respond to variations in the habitat: better connectivity will always increase their viability

to a similar extent. But this is because, offstage, the population of the species, and the size of the representative member, will both be the predominant mechanism by which the species accommodates to these variations. As better connectivity enhances the performance of firms, more firms will enter the habitat, and firm size will increase, until returns are moderated to a similar extent across habitats. Hence, the relationships at the level of the firm are the reduced form of a general equilibrium process in which jobs in formal enterprises are the principal, albeit offstage, endogenous variable.

We explore this issue next. To do this, we create a new, panel dataset using all cities for which we have two years of employment data—that is, cities where two Enterprise Surveys were conducted between 2006 and 2012. In addition, we estimate city population, density, and our two DRP measures at two points in time, 2002 and 2012.¹⁸ We then estimate fixed effects (FE) equations to examine the impact of city size and density on city employment levels. The results from this analysis are reported in Table 10. We find several notable results. First, the DRP measures perform much better in the employment regressions than standard density measures. For the full sample of cities (panel a), formal sector employment more than doubles with each doubling of market potential within 10km of the average person in the city. By contrast, the coefficient on log population density is not significant. Second, increased density is not generating the same employment effects in African cities as it is in other developing regions. As revealed in panel b, none of the density measures are significant for the African sample of cities. Third, Asian and Latin American cities have been more successful at harnessing the positive effects of urbanization than African cities, particularly in terms of creating more formal sector jobs. In Asia and Latin America, all three density measures are significant (panel c) and the estimated size of the coefficients are large.

4. Conclusion

In this paper, we present new evidence on the cross-country variation in agglomeration economies using geo-referenced, harmonized data from over 12,000 firms. Our study

¹⁸ We are unable to estimate road density at two points in time because we only have access to the most recent google maps.

contributes to the urban economics literature in two respects. First, we employ a new measure of city population and density to measure agglomeration economies. This measure uses the same threshold level of population density (1500 people per square kilometre) to define what is meant by “urban” rather than relying on arbitrary, administrative boundaries. Second, we investigate the impact of urbanization on establishment-level productivity and wages using a unique data set that covers establishments in more than 100 cities. Using this global sample, we reveal several important findings about the economic benefits of urbanization. First, we find strong evidence that urban productivity rises with city scale in developing economies. After controlling for country, industry, and year fixed effects, we estimate an output elasticity with respect to city size range from 0.067 to 0.080. This implies that urban productivity increases by about 5% to 6% with each doubling of city population. Second, we find evidence that agglomeration forces play an important role in expanding formal sector employment in cities throughout the developing world. Third, we reveal that African cities are not generating the same level of benefits—in terms of increased productivity, wages, and employment generation—as cities in Asia and Latin America. And, finally, we find evidence that the choice of density measure matters when estimating agglomeration effects.

References

- Acemoglu, Daron, Johnson, Simon, Robinson, James, 2001. The colonial origins of comparative development: an empirical investigation. *American Economic Review* 91, 1369–1401.
- Banerjee, A.V. and E. Duflo. 2005. “Growth Theory through the Lens of Development Economics,” in Steve Durlauf and Philippe Aghion, (eds.), *Handbook of Economic Growth*, Elsevier Science Ltd.-North Holland: 2005, Vol. 1A, pp. 473-552.
- Banerjee, A.V. and E. Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. NY: Public Affairs.
- Bento, P. and D. Restuccia. 2017. “Misallocation, Establishment Size, and Productivity,” *American Economic Journal: Macroeconomics* 9(3): 267-303.
- Bleakley, H., and J. Lin. 2012. “Portage and Path Dependence.” *The Quarterly Journal of Economics*, 127(2): 587–644.
- Bosker, M., S. Brakman, H. Garretsen, and M. Schramm. 2007. “Looking for Multiple Equilibria when Geography Matters: German City Growth and the WWII Shock.” *Journal of Urban Economics*, 61(1): 152–169.
- Bosker, M., S. Brakman, H. Garretsen, and M. Schramm. 2008. “A Century of Shocks: The Evolution of the German City Size Distribution 1925-1999.” *Regional Science and Urban Economics*, 38(4): 330–347.
- Brueckner, J. 2011. *Lectures on Urban Economics* (Boston: MIT University Press).
- Buera, F.J. and J. P. Kaboski. 2012. “Scale and Structural Change,” *Journal of Economic Theory* 147: 684-712.
- Buera, F.J., J. P. Kaboski, and Y. Shin. 2011. “Finance and Development: A Tale of Two Sectors,” *American Economic Review* 101: 1964-2002.
- Cali, M. and C. Menon. 2009. “Does Urbanization Affect Rural Poverty? Evidence from Indian Districts.” SERC Discussion Paper No. 14.
- Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. “What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States,” *Journal of Urban Economics* 98: 17-49.
- Collier, Paul and Patricia Jones. 2017. “Transforming Dar es Salaam into a City that Works,” in Adam, C., P Collier and B. Ndulu (eds.), *Tanzania: the Path to Prosperity*. Oxford: Oxford University Press, 2017.

- Combes, P.P., Duranton, G. and L. Gobillon, 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 68: 723-742.
- Combes, P.P., Duranton, G. and L. Gobillon, 2015. "The Empirics of Agglomeration Economies." In: Duranton, G., Henderson, J.V. and Strange, W.C. (eds.), *Handbook of Regional and Urban Economics*, vol. 5. (Amsterdam: North-Holland).
- Davis, D. R., and D.E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5): 1269–1289.
- D'Costa, S. and H. Overman. 2014. "The Urban Wage Growth Premium: Sorting or Learning," *Urban Economics*, 48: 168-179.
- De la Roca and Puga. 2017. "Learning by Working in Big Cities," *Review of Economic Studies* 84(1):106-142.
- Di Addario, S. and E. Patacchini, 2008. "Wages and the City: Evidence from Italy," *Labour Economics* 15:1040-1061.
- Duranton, G. 2008. "Viewpoint: From Cities to Productivity Growth in Developing Countries," *The Canadian Journal of Economics* 41(3): 689-736.
- Duranton, G. 2015. "Growing through Cities in Developing Countries," *World Bank Research Observer* 30(1): 39-73.
- Duranton, G. 2016. "Agglomeration Effects in Colombia," *Journal of Regional Science* 56(2): 210-238.
- Duranton, G. and D. Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In: Henderson, V, and Thisse, F (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. (Amsterdam: North-Holland).
- Gallop, John L., Sachs, Jeffrey D., Mellinger, Andrew D., 1999. Geography and Economic Development, Harvard Center for International Development Working Paper No. 1.
- Glaeser, E. and D. Maré. 2001. "Cities and Skills," *Journal of Labour Economics* 19:316-342.
- Glaeser, E. and W. Xiong. 2017. "Urban Productivity in the Developing World," NBER Working Paper 23279.
- Henderson, J. V. 2010. "Cities and development." *Journal of Regional Science*, 50 (1). 515-540.

- Henderson, J.V. and S. Kriticos. 2018. "The Development of the African System of Cities." *Annual Review of Economics*, forthcoming.
- Henderson, J.V. and D. Nigmatulina. 2016. "The Fabric of African Cities: How to Think about Density and Landuse," unpublished paper. London School of Economics.
- Howard, E., C. Newman, and F. Tarp. 2016. "Measuring Industry Coagglomeration and Identifying the Driving Forces," *Journal of Geography* 16:1055-1078.
- Iocovone, L., V. Ramachandran, and M. Schmidt. 2014. "Stunted Growth: Why Don't African Firms Create More Jobs?" Centre for Global Development Working Paper #353.
- Jones, P, O. D'Aoust and L. Bernard. 2017. "The Urban Wage Premium in Africa," (with Olivia D'Aoust and Louise Bernard) in S. Johnson-Lans (ed.), *Wage Inequality in Africa*. NY: Palgrave, 2017.
- Lall, Somik, J. Vernon Henderson, and Anthony J. Venables (with Juliana Aguilar, Ana Aguilera, Sarah Antos, Paolo Avner, Olivia D'Aoust, Chyi-Yun Huang, Patricia Jones, Nancy Lozano Gracia, and Shohei Nakamura). 2017. *Africa's Cities: Opening Doors to the World.* World Bank.
- Miguel, E. and G. Roland. 2011. "The Long-Run Impact of Bombing Vietnam." *Journal of Development Economics*, 96(1): 1-15.
- Mion, G. and PE. Naticchioni, 2009. "The Spatial Sorting and Matching of Firms," *Canadian Journal of Economics*, 42: 28-55.
- Puga, D. 2010. "The Magnitude and Causes of Agglomeration Economies," *Journal of Regional Science* 50(1): 203-219.
- Rosenthal, S. and W. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies". ch. 49, p. 2119-2171 in Henderson, J. V. and Thisse, J. F. eds., *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam: Elsevier.
- Salop, S. 1979. "Monopolistic Competition with Outside Goods," *Bell Journal of Economics* 10(1): 141-156.
- Syverson, C. 2004a. "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy* 112(6) : 1181-1222.
- Syverson, C. 2004b. "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics* 86(2): 534-550.

- UN-Habitat. 2016. *Slum Almanac 2015/2016: Tracking Improvement in the Lives of Slum Dwellers*. Nairobi: UN-Habitat.
- United Nations, Department of Economic and Social Affairs, Population Division. 2016. *World Population Prospects: the 2016 Revision*, New York.
- World Bank. 2015. "Firm Level Productivity Estimates: Methodological Note," World Bank Enterprise Surveys.
- Yankow, J. 2006. "Why do Cities Pay More? An Empirical Investigation of Some Competing Theories of the Urban Wage Premium," *Journal of Urban Economics* 60: 139-161.

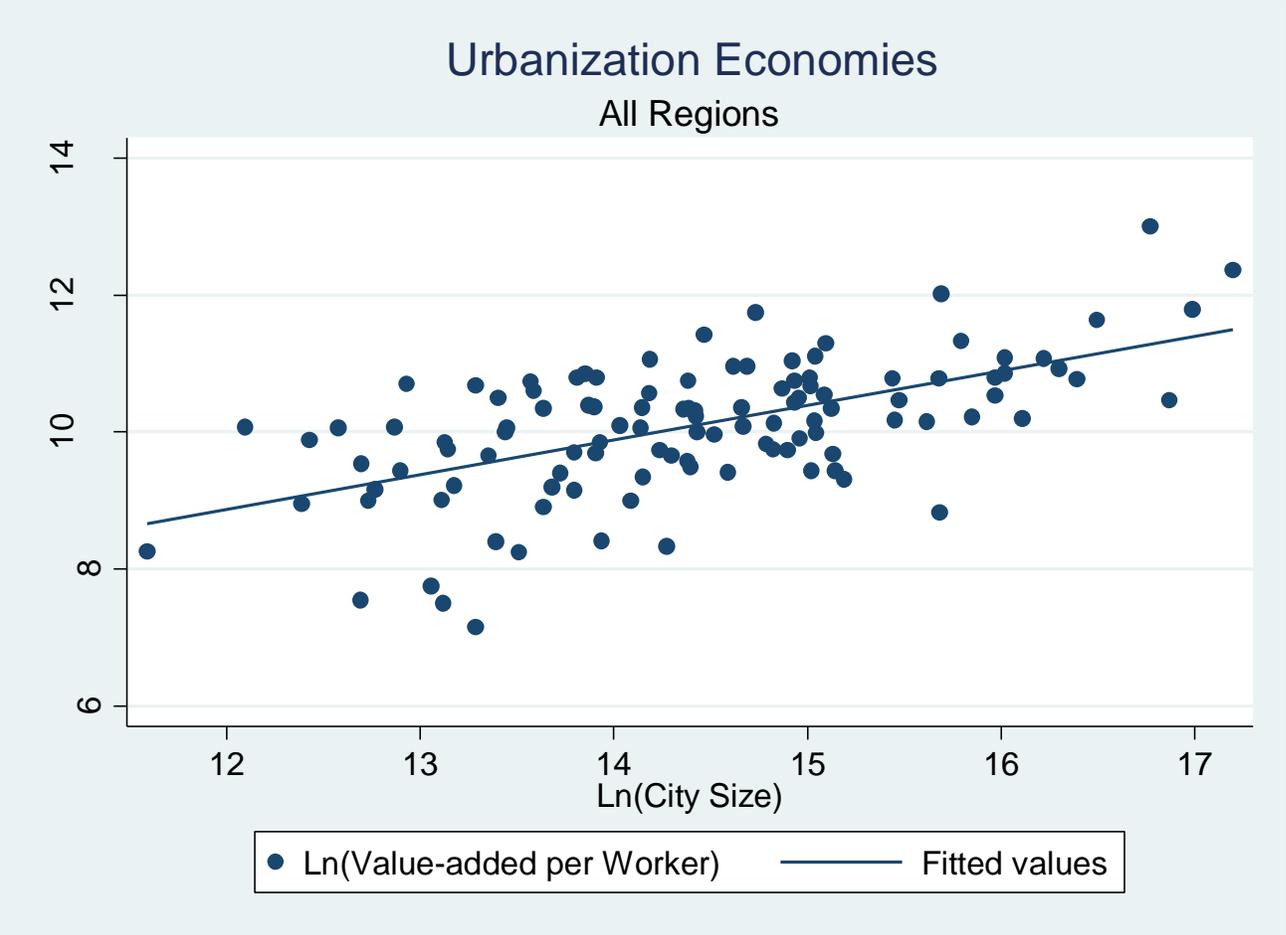


Figure 1: Urbanization Economies



Figure 2: Accra



Figure 3: Dhaka



Figure 4: Dar es Salaam

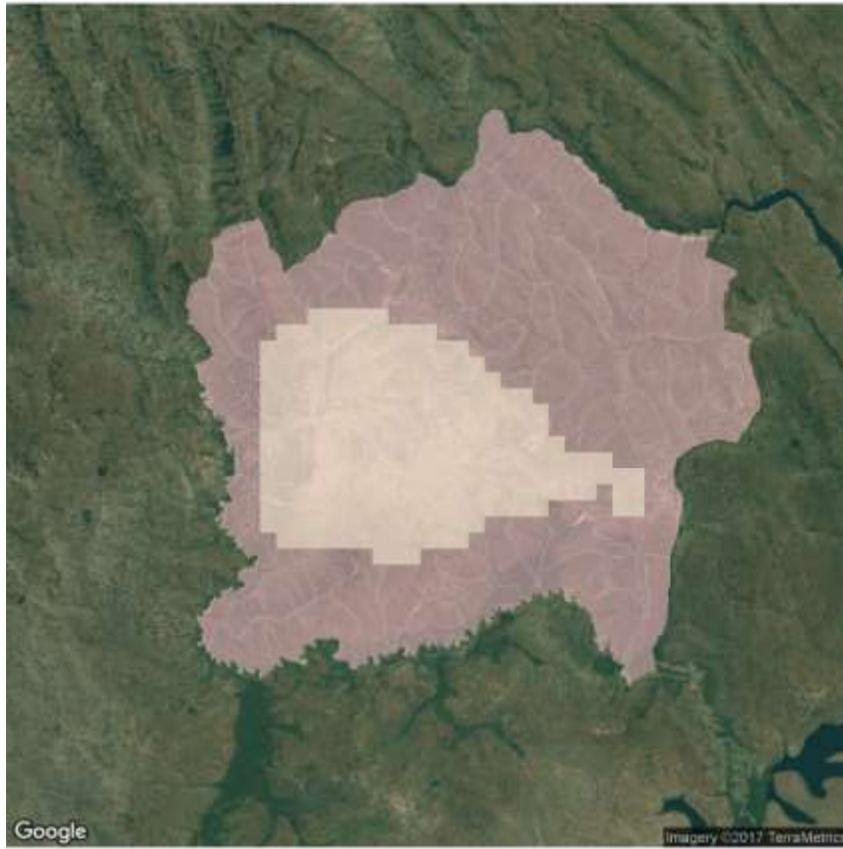


Figure 5: Kigali

TABLE 1: Sample of Global Cities

AFRICA (N=43 Cities)		LATIN AMERICA & CARIBBEAN (N=28 Cities)		ASIA (N=40 Cities)	
Abidjan (CIV)	Kano (NGA)	Arequipa (PER)	San Pedro Sula (HND)	Bago (MMR)	Nantong (CHN)
Abuja (NGA)	Khartoum (SDN)	Asunción (PRY)	San Salvador (SLV)	Bangkok (THA)	Ningbo (CHN)
Accra (GHA)	Kigali (RWA)	Barranquilla (COL)	Santa Cruz (BOL)	Beijing (CHN)	Phnom Penh (KHR)
Addis Ababa (ETH)	Kitwe (ZMB)	Bogotá (COL)	S. Domingo (DOM)	Cebu (PHL)	Qingdao (CHN)
Antananarivo (MDG)	Lagos (NGA)	Brasília (BRA)	Tegucigalpa (HND)	Chengdu (CHN)	Shanghai (CHN)
Arusha (TZA)	Lilongwe (MWI)	Buenos Aires (ARG)	Veracruz (MEX)	Chittagong (BGD)	Shengyang (CHN)
Bamako (MLI)	Lusaka (ZMB)	Chiclayo (PER)		Dalian (CHN)	Shenzen (CHN)
Banjul (GMB)	Maputo (MOZ)	Cochabamba (COL)		Delhi (IND)	Shijiazuang (CHN)
Beira (MOZ)	Mbeya (TZA)	Córdoba (ARGO)		Dhaka (BGD)	Suzhou (CHN)
Blantyre (MWI)	Mombasa (KEN)	Guatemala City (GTM)		Dongguan (CHN)	Tangshan (CHN)
Bobo-Dioulasso (BFA)	Mwanza (TZA)	Kingston (JAM)		Foshan (CHN)	Ulan Bator (MNG)
Bujumbura (BDI)	Nairobi (KEN)	La Paz (BOL)		Hangzhou (CHN)	Vientiane (LAO)
Cape Town (ZAF)	Nakuru (KEN)	Lima (PER)		Hefei (CHN)	Wenshou (CHN)
Conakry (GIN)	Nampula (NGA)	Managua (NIC)		Islamabad (PAK)	Wuhan (CHN)
Dakar (SEN)	Ndola (ZMB)	Medellin (COL)		Jakarta (IDN)	Wuxi (CHN)
Dar es Salaam (TZA)	Ouagadougou (BFA)	Mendoza (ARG)		Jinan (CHN)	Yangon (MMR)
Durban (ZAF)	Port Elizabeth (ZAF)	Mexico City (MEX)		Khulna (BGD)	Yantai (CHN)
Enugu (NGA)	Sokoto (NGA)	Panama City (PAN)		Luoyang (CHN)	Zhenshou (CHN)
Johannesburg (ZAF)	Takoradi (GHA)	Puebla (MEX)		Mandalay (MMR)	
Juba (SSD)	Toamasina (MDG)	Rio de Janeiro (BRA)		Manila (PHL_)	
Kaduna (NGA)	Windhoek (NAM)	Rosario (ARG)		Monywe (CHN)	
Kampala (UGA)		San José (CRI)		Nanjing (CHN)	

TABLE 2: Descriptive Statistics

Panel a:				
Firm Characteristics	All	Africa	Asia	Latin America
Firm Size (# of workers)	98.4 (260.1)	55.3 (133.1)	173.1 (399.8)	93.3 (230.4)
Firm Age	31.9 (157.8)	21.1 (106.0)	41.7 (219.3)	36.1 (153.1)
Average Wage	5,101 (10,046)	3,911 (11,291)	3,627 (6,611)	7145 (10,204)
Value-added per Worker	27905 (16,438)	20,591 (162,054)	37,777 (192,299)	28778 (73,082)
Capital-Labour Ratio	16,438 (89,550)	16,822 (106,018)	10,448 (31,450)	19,773 (96,238)
% Unskilled workers	25.9 (24.6)	25.7 (25.0)	26.2 (23.4)	25.9 (25.0)
Number of Firms	12,747	4,726	3,063	4,958
Panel b:				
City Characteristics	All	Africa	Asia	Latin America
Population, 2010 (UN)	3510.1 (4422.74)	1828.534 (2086.971)	5264.524 (5298.984)	3586.185 (4825.774)
Population, 2012 (LandScan)	3869.139 (5576.756)	1766.577 (1935.908)	6474.67 (7597.993)	3375.886 (4487.475)
Population, 1930	323.765 (555.2976)	87.263 (104.550)	466.423 (602.860)	357.870 (612.338)
Avg. management experience, years	17.6 (4.5)	14.6 (3.6)	17.4 (3.23)	20.8 (2.88)
Avg. % sales lost due to power outages	6.00 (5.43)	8.39 (5.40)	3.51 (5.29)	5.25 (4.14)
Avg. % management time spend on regulations	8.55 (7.39)	8.38 (5.88)	3.09 (4.49)	16.64 (5.26)
Night lights luminosity	737,042.1 (185327)	639,829.4 (178,738.8)	756,033.5 (198,643.5)	859,202 (46,636)
Number of Cities	111	43	28	40
Panel c:				
Spatial Characteristics	All	Africa	Asia	Latin America
Population Density	5,481 (2,388)	6,251 (2,678)	4,551 (1,933)	5,628 (2,102)
DRP 5km	594,673 (306,593)	496,287 (293,134)	721,799 (308,410)	564,157 (267,546)
DRP 5km with $e^{-0.5}$ discount	11,1976 (49,698)	103,854 (52,524)	126,057 (45,822)	104,333 (47,735)
DRP 10k	1,533,826 (1,007,719)	1,147,123 (843,602)	2,032,188 (1,072,801)	1,415,746 (864,412)

TABLE 2 cont'd: Descriptive Statistics

Panel c: Spatial Characteristics	All	Africa	Asia	Latin America
DRP 10km with $e^{-0.5}$ discount	124,745 (68,204)	100,484 (63,201)	155,623 (68,972)	117,889 (58,890)
Urban area (km ²)	773 (1,185)	273 (293)	1450 (1651)	573 (732)
Avg. (primary) road density	1.23 (0.70)	0.93 (0.49)	1.21 (0.70)	1.71 (0.74)
Distance to nearest port	429 (409)	548 (445)	277 (343)	464 (383)
Distance to nearest city	287 (220)	314 (238)	234 (229)	319 (162)
Latitude	7.73 (20.88)	-3.97 (14.5)	27.64 (10.42)	-2.77 (19.27)
Climate Index	0.51 (0.50)	0.70 (0.46)	0.28 (0.45)	.57 (0.50)
Number of Cities	111	43	28	40

Sources: Author's calculations based on World Bank Enterprise Surveys (May 2017), UN Population & Statistics Division (2017), LandScan (2012), and Mitchell (1998). All currency values in 2010 USD. Urban population in 1000s. Standard deviations in parentheses. Value-added defined as plants' annual sales minus raw material costs and energy costs. Wage defined as plants' annual labour costs divided by the total number of permanent workers. Slightly smaller samples for historical population estimates: 68 global cities with of which 19 are in Africa, 26 in Asia, and 23 in Latin American & the Caribbean.

TABLE 3: Agglomeration Economies, OLS Estimates
LandScan Population versus UN Population Estimates

	LandScan Population Estimates			United Nations Population Estimates		
Panel a: Output Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(value-added per worker)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(City Population)	0.080** (0.024)	0.092* (0.045)	0.072* (0.029)	0.074** (0.025)	0.087 (0.048)	0.067* (0.029)
Number of Firms	12,747	4,726	8,021	12,747	4,726	8,021
Number of Cities	111	43	68	111	43	68
R-squared	0.343	0.331	0.305	0.343	0.330	0.305
Panel b: Wage Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(wage)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(City Population)	0.083*** (0.021)	0.028 (0.041)	0.111*** (0.020)	0.086*** (0.022)	0.040 (0.040)	0.109*** (0.023)
Number of Firms	12,755	4,728	8,027	12,755	4,728	8,027
Number of Cities	111	43	68	111	43	68
R-squared	0.503	0.447	0.486	0.503	0.447	0.485

Notes: Author's calculations based on LandScan (2012) data and World Bank Enterprise Surveys (May 2017). Geographic boundaries of cities defined using population density threshold of 1000 people per square kilometer. In panel a, the dependent variable is logarithm of value-added defined as plants' annual sales minus raw material costs and energy costs. In panel b, the dependent variable is logarithm of wage defined as plants' annual labor costs divided by the total number of permanent workers. OLS regressions with country, industry, and year fixed effects. Robust standard errors clustered at the city level in parentheses. Standard errors in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 4: Urbanization Economies
LandScan Density versus UN Density Estimates

	LandScan Population Estimates			United Nations Population Estimates		
Panel a: Output Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(value-added per worker)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(Population Density)	0.042 (0.105)	-0.002 (0.141)	0.099 (0.155)	-0.112 (0.079)	-0.161 (0.138)	-0.059 (0.083)
Observations	12,747	4,726	8,021	12,747	4,726	8,021
R-squared	0.342	0.329	0.304	0.342	0.330	0.304
Panel b: Wage Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(avg. wage)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(Population Density)	-0.002 (0.106)	-0.215 (0.161)	0.193* (0.077)	-0.049 (0.065)	-0.131 (0.105)	0.024 (0.075)
Observations	12,755	4,728	8,027	12,755	4,728	8,027
R-squared	0.501	0.448	0.483	0.501	0.447	0.482

Notes: See Table 3.

TABLE 5: Urbanization Economies
LandScan Population Estimates

Output Elasticities:	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(value-added per worker)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(City Population)	0.049** (0.018)	0.066* (0.032)	0.041 (0.023)	0.060** (0.020)	0.047 (0.029)	0.061** (0.022)
Ln(K/L)	0.221*** (0.023)	0.248*** (0.015)	0.202*** (0.037)	0.208*** (0.025)	0.243*** (0.015)	0.176*** (0.042)
Ln(Firm Size)	0.161*** (0.024)	0.185*** (0.026)	0.146*** (0.031)	0.191*** (0.024)	0.181*** (0.027)	0.185*** (0.032)
Ln(Firm Age)	0.063*** (0.013)	0.061* (0.026)	0.058** (0.017)	0.059*** (0.014)	0.061* (0.025)	0.057** (0.017)
Ln(% unskilled)	-0.182** (0.058)	-0.166 (0.094)	-0.206** (0.073)	-0.199*** (0.056)	-0.183* (0.089)	-0.225** (0.071)
Ln(Management Experience)				-0.116 (0.176)	-0.125 (0.183)	-0.151 (0.278)
Ln(Road Density)				0.045 (0.094)	0.169 (0.124)	-0.007 (0.149)
Ln(% Sales Lost due to Outages)				-0.141*** (0.028)	-0.405*** (0.101)	-0.107*** (0.028)
Ln(Distance to Nearest Port)				0.004 (0.014)	0.000 (0.018)	0.003 (0.019)
Ln(% Mgmt. time spent dealing with government regulations)				-0.009 (0.081)	0.048 (0.103)	-0.017 (0.109)
Observations	12,656	4,678	7,978	11,024	4,678	6,346
R-squared	0.417	0.432	0.365	0.401	0.438	0.297

Notes: See Table 3.

TABLE 6: Urbanization Economies
LandScan Population Estimates

Wage Elasticities:	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(wage)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
Ln(City Population)	0.066*** (0.018)	0.017 (0.033)	0.092*** (0.019)	0.036 (0.023)	-0.051 (0.037)	0.072*** (0.019)
Ln (K/L)	0.160*** (0.009)	0.178*** (0.016)	0.147*** (0.010)	0.158*** (0.010)	0.173*** (0.017)	0.143*** (0.011)
Ln(Firm Size)	0.091*** (0.014)	0.122*** (0.030)	0.076*** (0.015)	0.113*** (0.012)	0.119*** (0.031)	0.103*** (0.010)
Ln(Firm Age)	0.051*** (0.013)	0.053* (0.023)	0.046** (0.017)	0.056*** (0.015)	0.053* (0.023)	0.056** (0.020)
Ln(% unskilled)	-0.147** (0.048)	-0.148* (0.061)	-0.162* (0.068)	-0.165** (0.049)	-0.151* (0.063)	-0.187* (0.073)
Ln(Management Experience)				-0.039 (0.125)	-0.017 (0.154)	-0.003 (0.157)
Ln(Road Density)				0.165 (0.092)	0.385** (0.140)	0.154 (0.086)
Ln(% Sales Lost due to Outages)				-0.074** (0.028)	-0.317*** (0.088)	-0.046 (0.030)
Ln(Distance to Nearest Port)				0.014 (0.009)	0.022 (0.016)	0.015 (0.011)
Ln(% Mgmt. time spent dealing with government regulations)				0.040 (0.049)	0.057 (0.104)	0.066 (0.068)
Observations	12,663	4,679	7,984	11,029	4,679	6,350
R-squared	0.554	0.523	0.528	0.548	0.529	0.468

Notes: See Table 3.

TABLE 7: Urbanization Economies
2SLS Estimation

Panel a: Output Elasticities	(1)	(2)	(3)
Dependent Variable: Ln(value-added per worker)	All Cities	African Cities	Asian & Latin American Cities
Ln(City Population)	0.106** (0.027)	0.219 (0.177)	0.096** (0.024)
Number of Firms	8,713	3,394	5,319
Number of Cities	56	19	37
p-value	0.021	0.007	0.064
F-statistic	82.57	133.9	66.80
Partial R-squared (first-stage)	0.85	0.87	0.85
Panel b: Wage Elasticities	(1)	(2)	(3)
Dependent Variable: Ln(avg. wage)	All Cities	African Cities	Non-African Cities
Ln(City Population)	0.102** (0.026)	0.048 (0.069)	0.096** (0.020)
Number of Firms	8,718	3,395	5,323
Number of Cities	56	19	37
p-value	0.040	0.758	0.042
F-statistic	82.56	60.63	66.79
Partial R-squared (first-stage)	0.85	0.86	0.85

Notes: 2SLS regressions with historical population and city latitude as instrumental variables. Explanatory variables include full set of firm and city controls plus country, year, and industry fixed effects. Robust standard errors clustered at the city level in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 8: Urbanization Economies
De la Roca Puga (DRP) Density Estimates

	DRP 10km			DRP 10km with $e^{-0.5}$ discount		
Panel a: Output Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(value-added per worker)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
De la Roca-Puga measure	0.087 (0.052)	0.070 (0.079)	0.104 (0.068)	0.086 (0.067)	0.063 (0.092)	0.116 (0.095)
Observations	12,747	4,726	8,021	12,747	4,726	8,021
R-squared	0.342	0.329	0.398	0.342	0.329	0.305
Panel b: Wage Elasticities	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Ln(avg. wage)	All Cities	African Cities	Asian & Latin American Cities	All Cities	African Cities	Asian & Latin American Cities
DRP 10km	0.088 (0.050)	-0.023 (0.078)	0.180*** (0.048)	0.075 (0.065)	-0.044 (0.093)	0.207*** (0.056)
Observations	12,755	4,728	3,065	12,755	4,728	8,021
R-squared	0.501	0.447	0.487	0.501	0.447	0.487

Notes: All regressions include only county, year, and industry fixed effects.

TABLE 9: Scale Economies
All Connectivity, Density & Market Size Measures

Panel a: All Cities	Employment Elasticity	Number of firms	R-squared
DRP 10km	0.194** (0.064)	12,755	0.192
DRP with e ^{-0.5} discount	0.182* (0.078)	12,755	0.190
Road Density	0.205 (0.108)	12,755	0.189
Population Density	0.269* (0.109)	12,755	0.190
Night lights	0.738** (0.223)	12,755	0.191
Panel b: African Cities	Employment Elasticity	Number of firms	R-squared
DRP 10km	0.141 (0.092)	4,728	0.272
DRP with e ^{-0.5} discount	0.147 (0.107)	4,728	0.271
Road Density	0.224 (0.188)	4,728	0.271
Population Density	0.321* (0.149)	4,728	0.274
Night lights	0.734** (0.224)	4,728	0.277
Panel c: Asian & Latin America Cities	Employment Elasticity	Number of firms	R-squared
DRP 10km	0.242** (0.081)	8,027	0.128
DRP with e ^{-0.5} discount	0.223* (0.098)	8,027	0.126
Population Density	0.238 (0.138)	8,027	0.125
Road density	0.222 (0.124)	8,027	0.125
Night lights	0.938 (0.548)	8,027	0.125

Note: Each column represents a separate regression where the dependent variable is the (natural) log of total formal sector employment at the city-level.

TABLE 10: Urban Employment
FE Regressions, All Cities

Panel a: All Cities	Employment Elasticity	Number of Cities	Overall R-squared
Population Density	0.247 (0.317)	104	0.545
DRP 10km	0.649* (0.317)	104	0.366
DRP 10km with $e^{-0.5d}$ discount	0.655+ (0.371)	104	0.361
Panel b: African Cities	Employment Elasticity	Number of Cities	Overall R-squared
Population Density	0.114 (0.183)	46	0.484
DRP 10km	0.083 (0.353)	46	0.453
DRP 10km with $e^{-0.5d}$ discount	0.218 (0.394)	46	0.504
Panel c: Asian & Latin America Cities	Employment Elasticity	Number of Cities	Overall R-squared
Population Density	1.210+ (0.618)	58	0.602
DRP 10km	1.580** (0.522)	58	0.390
DRP 10km with $e^{-0.5d}$ discount	1.515* (0.681)	58	0.335

Notes: Each column represents a separate regression where the dependent variable is the (natural) log of total formal sector employment at the city-level.