WEB APPENDIX: NOT FOR PUBLICATION

A Model Appendix

A1. Setting Up the Basic Model of an Individual City

Set-Up. Following a long tradition extending from Muth (1969) and Mills (1972) through the summary and review by Brueckner (1987), a monocentric city with homogenous households having one worker per household is housed around an employment center at a radius k, ranging from $[0, k^*]$. k^* is the edge of the city, and the territory beyond the edge is the alternative location. This subsection abstracts from the fact that there may be multiple cities of different sizes.

A fraction θ of the land area L is available for housing due to factors including topography and land required for non-housing uses. Land available for housing is given by:

$$L = \int_0^{k^*} \theta k dk \text{ with } 0 < \theta < 2\pi. \tag{1}$$

Households choose a location in a city based on earnings, y, the price of a composite commodity, p_c, and the rental price of housing at distance k, $r(k)$, so that their utility is identically equal to u^*, which is the utility associated with the location outside the city/countryside based on wages and house prices in that alternative location. Households located at distance k experience transportation costs which can include out-of-pocket or time costs (reductions in earnings) equal to tk. Housing suppliers rent to the highest bidder and hence maximize housing rent, subject to the iso-utility constraint imposed by u^*. This makes the landlord's pricing problem in a city:

$$\text{Max } r(k) = (y - tk - p_c c)/h(k) \text{ subject to } u^* - u(h(k), c(k)) = 0 \tag{2}$$

Where c is the composite commodity and h is housing space consumed per household. It is well known that the solution to this problem is Muth's equation (Muth, 1969):

$$dr(k)/dk = -t/h(k) \tag{3}$$

or dividing both sides by $r(k)$:

$$[dr(k)/dk]/r(k) = d\, log\, r(k) = -t/(r(k)h(k)) \tag{4}$$

Housing Price Gradient. Assume that the real income constant own price elasticity of demand for housing is equal to unity, $(r(k)h(k))$ = constant = v, where v is total expenditure on housing. In simpler terms, as the rent decreases with distance from the central business district (CBD), housing consumption increases proportionally, so that the homogenous workers spend a constant share of their equal income net of transportation cost on housing across all distances from the CBD. Following Brueckner (1982) and Kim and McDonald (1987), this is a necessary assumption to produce a negative exponential population density function.[1] Equation 4 implies a negative exponential *housing price gradient* in each city:

$$r(k) = r_o e^{-(t/v)k} \text{ or } ln\, r(k) = ln\, r_o - (t/v)k \text{ or } dr(k)/dk = -t/v \tag{5}$$

[1] See Zhao (2017) for a recent application of this approach including empirical validation for U.S. cities.

Equation 5 indicates that housing price is a negative exponential function with a decay rate equal to $-t/v$. Thus, housing prices decrease with distance from the CBD, and this gradient is flatter if housing expenditure v is higher or commuting costs t are lower. Workers want to consume more housing as income increases, so they are willing to live farther away, which flattens the gradient, even more so if commuting costs are low.

Housing, which should be interpreted as interior space used for housing, is produced using structure and land inputs according to a standard Cobb-Douglas production function:

$$H = AL^{\beta}S^{(1-\beta)} \tag{6}$$

where H is interior space produced, L is land input and S is structure inputs (i.e., non-land inputs). Structure inputs determine the height or density of housing space per unit land.[2] Structure inputs have the same price, p_s, everywhere as does the housing productivity parameter, A. β indicates the substitutability between land and building heights in the construction of housing space. A lower β suggests that land is less constraining in terms of housing production.

Land has a rental price of $R(k)$ which varies across locations within the city. Given that housing is produced by a perfectly competitive industry, the structure/land ratio is:

$$S/L = [(1-\beta)R(k)]/[\beta p_s] \tag{7}$$

Land Rent Gradient. The *land rent gradient* is based on the derived demand for land:

$$R(k) = [A\phi r(k)^{1/\beta}]/[p_s^{(1-\beta)/\beta}] \text{ or } ln\, R(k) = ln\, A\phi + (1/\beta) ln\, r(k) - [(1-\beta)/\beta] ln\, p_s \tag{8}$$

where $\phi = (1-\beta)^{(1-\beta)}\beta^{\beta}$ so that, as expected, land rent varies directly with the productivity of the housing sector reflected in A, and with housing price $r(k)$, while it varies inversely with the price of structure inputs, p_s. Differentiating equation 8 with respect to distance k, the nature of the land rent gradient is revealed to be:

$$dln\, R(k)/dk = (1/\beta) dln\, r(k)/dk = (1/\beta)(-t/v) = -t/(v\beta) \tag{9}$$

Equation 9 shows that land rent is a negative exponential function with a slope that is steeper than housing price (see equation 5) by a factor of $(1/\beta)$. In other words, land rents decrease with distance from the CBD, even more so than housing prices. A smaller share of land required in housing production, i.e. a lower β, implies a steeper land rent gradient.

Structure Density Gradient. The perfectly competitive housing industry takes $r(k)$ and $R(k)$ as given and supplies housing at a density of:

$$(H/L)(k) = A\phi r(k)^{[(1-\beta)/\beta]}/p_s^{(1-\beta)/\beta} \text{ or } ln(H/L)(k) = ln\,(A\phi) + [(1-\beta)/\beta] ln\, r(k) - (1-\beta)/\beta ln\, p_s \tag{10}$$

Equation 10 implies that structure density is an increasing function of productivity in construction, A, and housing rent $r(k)$ while it varies inversely with the price of structure inputs p_s and the effect of β is ambiguous, depending on the relative prices of structure and land inputs.

As is traditionally the case with the SUM discussed in Brueckner (1987), the model does not

[2]Housing services are sometimes interpreted to include attributes of the unit such as appliances, air conditioning, etc., but for the purposes of this discussion, the focus is on basic space.

deal explicitly with the spatial distribution of land used for non-housing purposes including commercial activity. However, only a fraction of land at each radius is available for housing. This fraction is based on the parameter θ (see equation (1)). It follows from the logic of the model that the land market forces non-housing activity to compete for land at each distance based on the land rent function (see equation (8)) and hence that structure densities in non-housing will follow the same negative exponential characterizing housing as structure is substituted for land when land price rises. Indeed, space for commercial purposes is produced using a process that is quite similar to residential space and it is not uncommon to find both activities in the same physical structure. For purposes of the SUM it is not necessary that all or even most employment be located in a CBD at the city center. It is simply necessary that employment is more spatially concentrated toward the CBD so that there is a premium on commuter access to the center.

Differentiating equation 10 with respect to distance gives the slope of the *housing density gradient*:

$$dln(H/L)/dk = [(1-\beta)/\beta]dr(k)/dk = [(1-\beta)/\beta](-t/v) = -t(1-\beta)/(v\beta) \tag{11}$$

Equation 11 makes it apparent that the slope of the housing density gradient is steeper than housing prices (see equation 5) but not as steep as the land rent gradient (see equation 9).

Population Density Gradient. Population density is simply the quotient of housing density from equation 10 and housing consumption per household. If the real income constant own price elasticity of housing demand is unity, housing consumption throughout the city is simply equal to housing expenditure divided by the housing price: $h(k) = v/r(k)$. Dividing both sides of equation 10 by this expression yields the *population density gradient* function:

$$D(k) = (H/L)(k)/h(k) = A\phi r(k)^{1/\beta}/(vp_s^{(1-\beta)/\beta}) \text{ or } lnD(k) = ln\,(A\phi)+(1/\beta)ln\,r(k)-ln\,v-(1-\beta)/\beta ln\,p_s \tag{12}$$

$$\text{and } dln\,D(k)/dk = (1/\beta)dln\,r(k)/dk = -t/(\beta v) \text{ or } D(k) = D_o e^{-(t/\beta v)k} \tag{13}$$

Combining this equation with the rest of the model results in an open version of the SUM where income is exogenous and, along with preferences and technology, determines the right hand side components of the density function. Accordingly, rising income from sources such as agglomeration economies determines city size along with preferences and technology. Note that the slope of the population density function is steeper than the housing density function (see equation 11). Indeed, $t/(\beta v) > t(1-\beta)/(\beta v)$. The intuitive reason for this is that housing space consumption rises with distance from the CBD because total expenditure is constant while price falls with distance. The slope of the population density function is identical to that of the land rent function (see equation 9). This is the artifact of assuming Cobb-Douglas housing production and real income constant own price elasticity of housing. If either of these assumptions are relaxed, the slope of the land rent function is steeper.

Taken together, equations 5, 9, 11 and 13 imply that for a given city, housing price, structure density, population density, and land rent are characterized by a series of negative exponential functions of increasingly steep slope. Figure A5 illustrates the relations among these variables that characterize the spatial pattern of land use within the city. The city limit at k^* is then determined by the point where the land rent function falls to the opportunity cost of land for other purposes, commonly called the agricultural reservation price, R_A.

Total Population. Having established the nature of the population density function in terms of the model parameters, the next task is to generate the total population of the city, which also depends on central density and not just the gradient. Recalling that total land available for household residences is given by $L = \int_0^{k^*} \theta k dk$, it follows that household population in the city is the integral of land area times density in equation 12 or:

$$N = \int_0^{k^*} \theta D(k)dk = \int_0^{k^*} \theta D_o e^{-(t/\beta v)k} dk \tag{14}$$

Thus far, no assumptions regarding the shape of the city have been made. If city development is unrestricted, it will develop as a circle, and $\theta = 2\pi\times$ the fraction of land available for housing. It is convenient to let the absolute value of the population density gradient be equal to $\lambda = t/(\beta v)$. Then, the definite integral in equation 14 can be written as:

$$N = (\theta D_0/\lambda^2)[(-\lambda k^* e^{-\lambda k^*}) - (e^{-\lambda k^*}) + 1] \tag{15}$$

Consider that, for large cities, which is the focus of our empirical analysis, k^* is large and $e^{-\lambda k^*}$ approaches zero. This implies that total households can be written simply as:

$$N = \theta D_0/\lambda^2 \tag{16}$$

$$ln\, N = ln\, \theta + ln\, D_0 - 2ln\, \lambda = ln\, \theta + ln\, D_0 - 2ln\, (t/(\beta v)) \tag{17}$$

The message of equation 17 should not be a surprise. The elasticity of population with respect to central density is $dlnN/dlnD_0 = 1$, as a given percentage rise in central density raises densities throughout the city by the same percentage. If the slope of the density gradient depends largely on parameters t, β and v, then the city population growth rate must be approximately proportional to the rise in central density, D_0. Note that, holding constant central density, factors that flatten the gradient, like housing expenditure v (and thus earnings) and the importance of land in housing β increase population while higher transportation cost t steepens the gradient and hence lowers population.

Having followed the literature on the SUM to establish the relation among general characteristics for a single city, it is possible to vary city earnings to alter population, and establish the relation among earnings or population and other characteristics.

A2. Using the Model to Vary the Size of Cities within a Given Country

The result just developed above is designed to reflect the conditions that govern urbanization within a given country. In a country, absent any role for amenities, population mobility should equate real income across cities. In addition, there is a presumption that urban technology including transportation cost, the price of structure inputs, and the opportunity cost of converting land to urban use are identical. Finally, the fraction of land available for development is assumed constant, and there is no congestion that raises transportation cost per unit distance as city size increases. Clearly, the actual city size distribution in a country includes variation among areas in all these factors. The discussion here can either be thought of as abstracting from all these factors or as treating their variation as idiosyncratic and unrelated to the fundamental driver of size variation, differences in earnings of workers across cities in a given country.

Having Cities of Different Sizes. In order to generate cities of different sizes, the standard approach is to formulate an open city model in which wages vary due to exogenous labor productivity differences. It is then possible to establish a relation between earnings paid to workers in the city and total population of the city. Once the relation between earnings and population is established, the implications for all the other basic city characteristics can be established and the question of whether these changes are all proportional to differences in population (in which case population is a sufficient statistic to describe city differences within a country) or not can be determined.

Imagine that all cities begin with identical characteristics. The price of housing at k^* must be identical in all cities because the cost of housing production at the city limit is identical everywhere. Indeed, structure inputs are priced everywhere at p_s while land at k^* is always priced at R_A, leaving housing production cost, and thus housing price $r(k^*)$, identical across cities.

In order to grow a particular city, y must rise making that city more attractive. Raising earnings by Δy means that workers at the city edge who pay $r(k^*)$ for housing will be willing to pay an additional $\Delta y = t\Delta k$ for transportation to work. Conversely, the outer boundary of the city can expand by $\Delta k* = \Delta y/t$. Thus, if y increases in one city, k^* in that city will increase until the workers at the edge are indifferent between working at the CBD or outside the city. As k^* and thus total land area increase, population also increases.

In addition to expanding the city boundary, the rise in city earnings would raise the utility of workers living closer than k^* to exceed that of workers in other cities. Their willingness to pay for housing in the city would rise. In order to maintain the iso-utility condition among cities, rents must rise throughout the growing city as shown in Figure A6 with city 2 having higher earnings than city 1. The increase in rent Δr needed to maintain the iso-utility condition in response to this increase in wages, requires that housing expenditure must rise by Δrh to offset the increase in earnings Δy. It follows that $\Delta y = \Delta rh = \Delta rhr/r$. Given that $v = rh$, it follows that $\Delta y = v\Delta r/r$.

Letting η be the fraction of income spent on housing so that $v = \eta y$, this implies $y = \eta y\Delta r/r$ or $\Delta y/y = \eta\Delta r/r$ or that $dlny = \eta dlnr$ and $dlnr/dlny = 1/\eta$. Thus, the percentage increase in rent needed to offset a given percentage rise in income is equal to the inverse of the share of housing in income.

Totally differentiating equation 16 with respect to log income the effect of changing income on central density may be written as:

$$d\,ln\,D_o/d\,ln\,y = -\partial\,ln\,v/\partial\,ln\,y + (1/\eta)\partial\,ln\,r/\partial\,ln\,y = -1/\eta + 1/\beta\eta = (1-\beta)/\beta\eta(15) \qquad (18)$$

Central density thus varies inversely with β, which measures the importance of land as a factor of production for housing, and η, which is the fraction of income spent on housing.

City Population Size-Earnings Elasticity. Now, it is possible to determine the elasticity of city population size with respect to city earnings. The total differential of equation 17 with respect to income is:

$$d\,ln\,N/d\,ln\,y = (\partial\,ln\,N/\partial\,ln\,D_o)(\partial\,ln\,D_o/\partial\,ln\,y) + (\partial\,ln\,N/\partial\,ln\,\lambda)(\partial\,ln\,\lambda/\partial\,ln\,y) \quad (19)$$

From equation 19 it is evident that, as income increases, both central density and the density gradient will change, which then determines the consequent change in total population.

Using the results that $\partial lnN/\partial lnDo \approx 1$ and that $\partial lnN/\partial ln\lambda \approx -2$ from equation 17, substituting the result from equation 18 for $d\,ln\,D_o/d\,ln\,y$, and substituting $\lambda = t/(\beta v) = t/(\beta\eta y)$ into equation 19 yields the elasticity of city population with respect to city earnings:

$$d\,ln\,N/d\,ln\,y = (1-\beta)/(\beta\eta) + 2/\eta = 1/(\beta\eta) + 1/\eta \quad (20)$$

Equation 20 implies that the elasticity of household population with respect to income, or what is commonly called the urban wage premium, is a constant equal to a function of the share of land in housing construction and the share of housing in income. Thus, the urban wage premium varies inversely with the importance of housing in household consumption bundles and the share of land in housing construction. This result is consistent with but not based directly upon the Rosen-Roback model (Rosen, 1979; Roback, 1982), which states that, in order to attract labor, growing cities must overcome rising housing prices. The rise in housing price is larger when the share of land inputs, i.e. β, is larger, and has a bigger effect on cost of living when the fraction of income spent on housing, reflected in η, is larger. There is a transportation cost effect because commuting distance and cost are embodied in the model. However, because transportation cost per unit distance traveled is assumed constant within a country, perhaps because larger cities have more investment in transportation capacity, it does not enter the wage premium expression.[3]

It is possible to calculate the implied urban wage premium for the U.S. from equation 20 because there are well accepted values for the share of land in housing, β = 0.2, and the share of housing in consumption in income, η = 0.3.[4] Using these values, the calculated value of the premium, is $dlnN/dlny = (1/\beta\eta) + (1/\eta)$ = 20 or inverting this expression, a premium of 0.05, a 10% rise in population generated by 0.5% rise in earnings. There is an established literature on the urban wage premium for the U.S. which uses Mincer equations to compute the partial relation between city size and earnings change. The general consensus is that the premium is approximately 0.05 to 0.06 (Combes and Gobillon, 2015). The precision of these estimates is based on use of high quality micro data on wages or earnings that allows for differences in population composition, which are also associated with city size, to be removed. Estimating the premium for lower income countries is more challenging due to the inability to adjust for the tendency for human capital per worker to rise with city size. Recent estimates by Chauvin et al. (2016) place the urban wage premiums for the U.S., Brazil, China, and India at 0.054, 0.052, 0.088, and 0.077, respectively.[5] Given the difficulty of removing the effect of the tendency for human capital per

[3]Given $\lambda = t/(\beta v)$ it follows that $d\,ln\,\lambda/d\,ln\,y = \partial\,ln\,t/\partial\,lny - \partial\,ln\beta/\partial\,lny - \partial\,lnv/\partial\,lny$. It could be argued that $\partial lnt/\partial lny \neq 0$ as assumed here. The problem is that there are reasonable arguments for thinking that the effect of income on transportation cost could be negative or positive. Negative due to improved transportation technology and positive because higher earnings raise the value of time. Likewise, we could imagine that richer cities within a given country are better able to substitute structures for land when producing housing, so $\partial ln\beta/\partial lny < 0$.

[4]See, for example Glaeser and Gyourko (2018).

[5]These are OLS estimates. IV estimates for the U.S. and Brazil are similar but IV estimates for China and India

worker to rise with city size on estimates of the urban wage premium in China and India, and issues created by impediments to population mobility, these estimates are remarkably consistent.

A3. Using the Model to Compare Cities in Developed and Developing Countries

As noted above, the characteristics of large cities, whose populations are similar as we will see below, vary dramatically across countries. The purpose of this section is to take the same SUM model that explains the ability of population to reflect differences among cities *within* a country with high population mobility and determine if it can generate differences in cities of the same size that are observed *across* countries. The stylized example discussed here is the contrast between the typical city in a developed country, noted d, with cities in developing countries, noted g.

Decomposition of Total Population. It is useful to consider a simple decomposition of the determinants of total population because that is the principle characteristic that cities in developed and developing countries have in common. Total population is the product of land area, structure density, and "crowding" or population per unit interior space:

$$N = L(H/L)(N/H) \tag{21}$$

where N = urban household population, L = land area, H = interior space, so that (N/H) = population per square foot of interior space or "crowding".

More formally, in terms of equation 14 above, this is written as an integral rather than as the product of averages:

$$N = \int_0^{k^*} \theta D(k)dk = \int_0^{k^*} \theta(H/L)(k)/h(k)dk \tag{22}$$

where $\theta = 2\pi$ (fraction of land for housing) (if development is unrestricted and the city is a circle), $D(k)$ = population density function, $(H/L)(k)$ = density of interior space per unit land at k, $h(k)$ = interior space consumed or used per household at k and k^* is the city limit.

Differences Within vs. Across Countries. The stylized facts that we will establish later on show that, holding population constant, land area (hence k^*), building heights ($(H/L)(k)$) and interior space consumption per household ($h(k)$) *all* increase directly with GDP per capita so that all three are larger in developed than developing cities. Because population varies directly with land area and structure density and inversely with interior space consumption, Equation 21 guarantees that, holding population constant, cities in high income countries are very unlike those in low income countries. The result is that knowing population N is, by itself, not sufficient to characterize any of the three major spatial city characteristics in equation 22 and similarly for other characteristics of the city including total output. Given that population N varies directly with land area (hence k^*) and building heights ($H/L(k)$), and inversely with interior space consumption per household ($h(k)$), it must be that the latter is sufficiently lower in developing countries to compensate for the influence on population of the lower values of the former two variables.

Therefore, the claim, based on the discussion above, that *within* a country there is a regular relation among land area L, building heights (H/L), and interior density (N/H) along with other

exhibit wide variance.

aspects of urban spatial structure, need not hold *across* countries. The intuitive reason for this difference is that, within a country, the iso-utility condition keeps housing expenditure v constant and the sharing of technology for producing housing and transportation services limits variation in A, t and p_s. Therefore, the principal effect of rising population is to increase land area through rising k^* and the analysis of the implications of the SUM above indicated that the elasticity of population with respect to income depends on factors, specifically β and η, which should not vary appreciably across cities of a same country.

In contrast, there is no free mobility across countries of different income levels.[6] Comparing developed and developing cities, there are obvious differences in key parameters. First, real incomes are an order of magnitude higher in the cities of developed countries ($y_d \gg y_g$), and thus total expenditure on housing is higher ($v_d \gg v_g$). It follows that housing consumption per household is higher $h_d > h_g$. Superior building technology means that $A_d > A_g$. One important element of building technology is secure property rights and financial markets that facilitate construction at a high structure/land ratio. Likewise, developed countries are better able to substitute structure inputs for land ($\beta_d < \beta_g$). Lastly, it might be that $t_d < t_g$ due to advanced transportation technology in rich countries. However, one element of transportation cost is the opportunity cost of time and, given that wages are an order of magnitude higher in rich cities, this inequality is ambiguous. Similarly, differences in cost of structure inputs p_s are not obvious.

The implications of differences in the parameters for city structure are evident in equations 5, 10, 11, and 13 above. First, equation 5 shows that the slope of the housing price gradient becomes flatter as v rises and t decreases. From equations 10 and 11, housing structure density is increasing in A and the slope of the density gradient is decreasing in v. Finally, equation 13 shows that the slope of the population density function flattens as t decreases, β increases, and v increases. This last result is important because, comparing developed or developing cities with equal population, if the density function of the developed city is flatter, then the radius of the developed city is larger and the central density of the developed city must be lower. This relation is shown in Figure 1, which plots the population density functions of a developed city and a developing city of similar sizes. Note that the area under the $D(k)_g$ graph appears larger than the area under the $D(k)_d$ graph but the weight of land at each distance is proportional to the square of k.

Results on Structure and Population Densities. While our results suggest that there may be similarities in the elasticity of city size with respect to earnings, the model implies that, in other respects, the cities are very different. Consider structure and population densities. Equation 10 reproduced here implies that structure density is determined by construction technology, the share of housing in land, the price of structure inputs, and the price of housing space:

$$(H/L)(k) = ln\,(A\phi) + [(1-\beta)/\beta] ln\, r(k) - (1-\beta)/\beta ln\, p_s \tag{23}$$

Considering central structure density at $(H/L)(0)$. Equation 23 suggests that unless p_s is much lower or $r(0)$ is much higher in developing cities, the positive relation between A and

[6] Since our goal is to compare patterns of physical urban development between developed and developing countries, we purposely ignore cases such as the European Union where there is almost free mobility across countries.

development should produce much higher central structure densities in cities in developed countries. However, the slope of the structure density gradient shown in equation 11 to be $d\,ln\,(H/L)/dk = -t(1-\beta)/(\beta v)$ is clearly flatter in higher income countries as housing expenditure v, is higher by an order of magnitude. Figure 2 displays the relation between structure density functions of cities with the same population size in developed and developing countries.[7]

The SUM also predicts that, holding population constant, central population density differences between cities in developed and developing countries will be the opposite of structure density. Recall from equation 13 reproduced here that:

$$lnD(k) = ln\,(A\phi) + (1/\beta)ln\,r(k) - ln\,v - (1-\beta)/\beta ln\,p_s \tag{24}$$

While it is true that the positive association between building technology, A, and development tends to raise central density, $D(0)$, the fact that income and hence v is higher by an order of magnitude produces a negative association between central density and development, again unless structure prices are much lower or rents much higher in the developing city. While central population densities are predicted to be lower, the slope of the population density gradient is based on equation $d\,ln\,D(k)/dk = -t/(\beta v)$ which, like the structure density function, will flatten with the level of economic development due to the rise in v with income. The relation between population density functions in cities in developed and developing countries with identical population is depicted in Figure 1.

In conclusion, when comparing cities in developed and developing countries, the relation between total population and other city characteristics is problematic because the ratio of labor to land or labor to capital stock is far lower in developing cities. The density of population is much higher but structure densities are lower and land area is much smaller.

However, despite these substantial differences in levels for cities in developed and developing countries with the same population, this does not mean that the elasticities of land, structures, or output, with respect to population must be different. As noted in equation 20, it is possible for the elasticity of income with respect to population to be similar for cities in developed and developing countries. This means that, the ratio of output, land area, or structure capital to population for two cities in a developed country can be the same as the ratio in a developing country. Accordingly, research that uses population ratios to reflect ratios of other city characteristics could possibly pool observations from cities in developed and developing countries.

The same forces that produce differences in population density between cities in high and low income countries also apply to employment density. Higher income countries have greater structure density due to building technology differences but space per worker also increases with income as firms substitute physical capital in the form of space when labor costs rise. The result is the same combination of higher structure density near the city center but lower employment density as space per worker expands faster than space per unit land. Just as was the case with

[7]It might also appear that the structure density gradient is flatter in cities in developed countries because t falls with improved transportation technology. While this is true, transportation cost also includes the value of time which rises with income. Therefore, the relation between t and economic development may be ambiguous.

housing and population density, the amount and spatial pattern of non-housing investment is very different in high and low income cities although total employment may be similar.

B Data Appendix - Details on the Building Heights Data

The original data set of CTBUH (2018) (accessed between January 2017 and January 2018) has 27,652 *tall buildings*. Once we keep "buildings" and "tower-buildings" that are completed or about to be completed (architecturally or structurally topped out), we are left with 19,132 buildings.

Heights. According to the website of CTBUH (2018), they do not use a consistent definition of tall buildings across all cities. We thus study how heights vary across the data set. To do so, we need to obtain height for as many buildings as possible. For most buildings, we know height to tip of the building (no matter the function of the highest element) and/or height to the architectural top of the building (which may include spires but excludes antennae) and/or height to the highest occupied floor and/or height of the observatory of the building if there is one and/or the number of floors above ground. Height to the highest occupied floor may be the best measure but it is only available for 11.6% of buildings whereas architectural height is available for 84.7%, height to the tip 60.6%, height to the observatory 1.1%, and the number of floors 98.2%. We thus use architectural height as our main measure. Since it is missing for 15.3% of buildings, we impute it when possible with data on height to the tip (correlation between architectural height and this height = 0.99), then data on height to the highest occupied floor (correlation = 0.98), and finally data on height to the observatory (correlation = 0.96). We then regress our measure of height on the number of floors and find a coefficient of 3.8***, which indicates that a floor corresponds to 4 meters for most buildings in the world (95% conf. interval = [3.77; 3.87]). We can then impute heights for the remaining buildings. In the end, we obtain a consistent measure of heights (m) for 99.6% of buildings. We will nonetheless verify results hold if we only use architectural height or height not using information on number of floors.

As can be seen in Web Appendix Figure A1 which plots the Kernel distribution of building heights in the data set, the mode of the distribution is 80 m. Since cities are likely to have more buildings below 80 m than above 80 m, and since the distribution of buildings is relatively smooth after 80 m, this suggests that the data set captures buildings above 80 m only. Thus, the data set is unreliable for buildings below 80 m. We are then left with 14,729 buildings of more than 80 m.

Year of Construction. For most buildings, we know the year of completion and/or the year construction started and/or the year construction was proposed. We use the year of completion as our main measure, since it is available for 96.6% of buildings. For the remaining buildings, we impute the year of completion using information on the year construction started (correlation with the year of completion = 0.99), then the year construction was proposed (0.97). On average, a building is completed 5.5 years after construction is proposed and 3.3 years after construction is started. We then obtain the year of completion for 96.8% of buildings.

Function. The function(s) of the buildings is then available for 98.9% of buildings. Many buildings have multiple functions. Among buildings for which we know the function, 49.9% of them are used for residential purposes and 41.8% of them include offices. Other important functions include hotels (11.8%) and retail (3.9%). Other functions are more marginal.

C Robustness for the City Population Size-Income Elasticity

Column (1) of Web Appendix Table A2 replicates the baseline results of row 2 and columns (1)-(2). Low elasticities are also obtained for 300K+ agglomerations when using other data sets and their own population estimates for the year 2015, whether Demographia (2017) (col. (2)), European Commission (2018) (col. (3)) or Atlas of Urban Expansion (2016) (col. (4)). Finally, elasticities remain unchanged if we drop each country one by one (not shown) or remove the largest cities of each country, to study non-primate cities only (not shown).

National per capita incomes are used because consistent income or earnings data does not exist for enough individual cities. In addition, at the world level, income differences are driven by differences *between* countries rather than by differences across cities *within* countries. Indeed, with free mobility and abstracting from amenities and worker skill heterogeneity, all cities in a country should offer similar wages in local PPP terms (i.e., net of housing and other prices). National per capita GDP is indeed strongly correlated with: (i) city per capita GDP (source: Oxford Economics (2019); 0.91-0.92);[8] and (ii) night lights per capita (source: NGDC (2015); 0.78-0.81).[9] The between component, i.e. differences across countries, accounts for 91-93% and 67-68% of the variation in city per capita GDP and night lights per capita, respectively. We nonetheless verify that low elasticities are found when using city per capita GDP (col. (5) and (6)) or night lights per capita (col. (7) and (8)), whether for the full sample or the main sample.

D Robustness for the City Land Area-Income Elasticity

There could be classical measurement error in city land area. However, since city land area is the dependent variable, this would only affect precision. Yet, if the quality of the land area measurement is correlated with income, the elasticity is mis-estimated. For example, if land area is under-estimated in developing (developed) country cities, the elasticity will be over-estimated (under-estimated). We thus check how robust our results are when using alternative data sets of city land areas, since the different data sets use different sources and methodologies. Next, the availability of city land area estimates itself could be endogenous. Finally, classical measurement error in per capita income causes a downward bias, whereas non-classical measurement error in income could generate either an upward bias or a downward bias.

First, log land area in Demographia (2017) is strongly correlated with log land area in Atlas of Urban Expansion (2016) (0.91-0.91; N =140) and European Commission (2018) (0.78-

[8]The Oxford Economics (2019) database, prepared for the World Bank, includes per capita GDP (constant 2012 million $) for 775 agglomerations annually from 2000 to 2017. According to their note on methods, they used detailed national accounts data or estimates provided to them by the administrations of these cities. When the data was unavailable, they used per capita GDP for regions encompassing the agglomerations. Because no details are provided for each estimate, and estimates may have been spatially and/or temporally interpolated and/or extrapolated, we believe that the quality of this data is not high and do not use it as our main source of income data.

[9]Night lights per capita is the sum of night lights of the city (2011) divided by the population of the city (2010). Satellite images are provided by NGDC (2015), and are available at a fine spatial resolution, annually from 1996-2011. Note that we use the radiance calibrated version of this data, to avoid issues related to top-coding, especially for cities. In particular, radiance calibrated data imply that night light intensity can go beyond the typical upper-bound of 63. The Global Rural-Urban Mapping Project, CIESIN (2017), then provides geocoded polygons of urban extent boundaries circa 1995. Using GIS, we obtain the sum of satellite night lights in each polygon from 1996 to 2011. We match the agglomerations in our sample to their respective geocoded polygon.

0.87; N =1,007). Column (1) of Web Appx. Table A3 then replicates the baseline results. The elasticity of land area with respect to income remains close to 0.5 if we use Atlas of Urban Expansion (2016) (col. (2)). If we use residential built-up area or non-residential built-up area instead of total land area, since these measures are available in Atlas of Urban Expansion (2016), we find very similar elasticities (not shown). Note that built-up areas account for 70% of total land area, of which two thirds is "residential". The fact that elasticities differ little between residential and non-residential areas suggests that rich country cities use more land for *all* purposes. Now, if we use European Commission (2018) data, elasticities become smaller but only if we use total land area (col. (3)), not if we use built-up area estimates and focus on the main sample of 1,010 cities (col. (4)).[10] Elasticities also remain unchanged if we drop each country one by one, except the U.S.. In that case, the elasticities are lower, at 0.36***-0.43*** (not shown). U.S. cities disproportionately consume more land given their income level. If we drop primate cities, elasticities are, if anything, higher, at 0.66***-0.71*** (N = 863, not shown). Land area in Demographia (2017) is then only available for 1,010 out of 1,773 agglomerations of more than 300,000 inhabitants in 2015. We verify results hold if we use weights that makes the main sample of 1,010 cities representative of the income distribution of the full sample of 1,773 cities (col. (5)).[11]

Finally, if we use other per capita income measures, we find similar elasticities of land area with respect to city per capita GDP (source: Oxford Economics (2019); col. (6)), but lower elasticities for night lights per capita (0.30***-0.34***; source: NGDC (2015); N = 1,009; col. (7)).[12]

E Robustness for the City Building Heights-Income Elasticity

Col. (1) of Table A4 replicates the baseline results of col. (1)-(2) in row 4 of Table 2. Our data comes from CTBUH (2018) and captures *tall buildings* above 80 m (about 20 floors). However, there could be classical measurement error in city building stocks. Yet, because average building height is the dependent variable, this would only affect precision. But if the quality of the tall building data set is correlated with income, the elasticity is mis-estimated. For example, if the total number and height of buildings are under-estimated in developing (developed) country cities, the elasticity will be over-estimated (under-estimated). For our main hypothesis that developed country cities disproportionately grow by building up, non-classical measurement error is an issue if and only if we over-estimate the elasticity. The two important questions are then: (i) whether our data disproportionately miss *tall* buildings in poorer countries? and (ii) whether *non-tall* buildings (low-rise buildings, houses, shacks and tents) are taller in poorer countries?

To compare results, we first collected data from a second source, Emporis (2019), another global provider of building information.[13] Note that Emporis (2019) claims to capture all *high-rise buildings*, which they define as buildings above 35 meters (about 9 floors). They then classify

[10] We believe that this has to do with the algorithm that they use, which appears to disproportionately under-estimate urban open spaces in richer countries or over-estimate urban open spaces in poorer countries.

[11] More precisely, the weights are constructed to reflect the respective population shares of high-income, upper-middle income, lower-middle income and low income countries among the 1,773 agglomerations.

[12] Since the elasticity of night lights per capita with respect to national per capita GDP is 1.29*** once we control for log city populations (N = 1,009; not shown), the implied elasticities with respect to national per capita GDP are 0.40-0.45.

[13] Their website says they rely on their extensive member network to gather information on buildings.

as *skyscrapers* buildings above 100 meters. Finally, they use the number of floors of each 35m+ building to compute for each city a Skyline index. We do not have access to their raw data but their website reports useful information for the 100 top cities in the world.[14] For 90 of these cities also in our data, and using as weights the sum of heights in our data in order to focus on the cities with the most tall buildings, the correlation between the log of their number of skyscrapers and the log of our own number of buildings above 100 meters is 0.90. Next, the correlation between the log of their Skyline index and the log of their number of skyscrapers is 0.83. The correlation of their Skyline index (based on 35m+ buildings) with our own reconstructed index (based on 80m+ buildings and their formula) is 0.79. Thus, our measure is a good proxy for 35m+ buildings.

Is our measure also a good proxy for formal structures below 35m, whether low-rise (four plus one) buildings or houses? Based on Emporis, which reports the number of low-rise buildings for 7 North American cities, the (80m+) buildings in our data account for between half and two thirds of total heights including low-rises. In addition, for each building, we know the main material used. While it was steel around 1950, the use of concrete increased dramatically over time, reaching 90% in the 2000s (the mean share of concrete over the period 1950-2017 is 73%). We then obtained from the *Minerals Yearbooks* of USGS (2020) and for 144 countries and each decade from 1950 the total production of cement – the main ingredient of concrete – which we use as a proxy for cement consumption.[15] As expected, the correlation between decadal tall building construction and decadal cement use is very high, at 0.77 (N = 870). Adding country and year fixed effects, we obtain a correlation of 0.80 (but 0.99 if we use urban population as weights). Thus, we believe tall building construction is a very good proxy for the formal construction sector.

Do cities with taller buildings have fewer informal structures? If that is the case, under rather than over-estimation of the elasticity is likely. Slum shares are not available or not consistently constructed for enough world cities. However, we obtain from the *United Nations Global SDG Database* the proportion of urban population living in slums (%) at the country level circa 2016. For 116 countries with data on the population share of slums, urban population, building heights, urban land area, and per capita GDP (PPP), we then examine if adding the slum sector to the tall sector reduces the income elasticity of building height. First, we verify for this sample that the urban population-income elasticity is small. When using urban population sizes as weights, we obtain a non-significant elasticity of 0.20 (not shown). The area-, height- and occupant density-income elasticities are then 0.89***, 0.84*** and -1.53*** (not shown).[16] The area- and occupant density-income elasticities are higher in magnitude than for individual cities because the urban sector includes smaller cities for which land expansion is the main component of growth. Adding 0.89, 0.84 and -1.53, we also obtain exactly 0.20. The slum share then very strongly decreases with log per capita GDP (-18.1***; not shown) and average (tall) building height (-7.4***; not shown). Thus, we are unlikely to under-estimate structure heights in poorer countries.

[14] Accessed on 12-11-2019: https://www.emporis.com/statistics/skyline-ranking.

[15] Because cement is a low-value bulky item, the world trade of cement only accounts for 3% of world cement production (World Cement, 2013). Cement production is thus a very good proxy for cement consumption.

[16] Recall that we proxy occupant density by the number of city residents divided by the sum of tall building heights.

To more directly see how slums modify the height-income relation, we calculate for each country the total number of slum residents. We then assume that each slum resident occupies 3 sq m and that slum structures consist on average of two low-ceiling floors of 2 m each, i.e. are 4 m tall.[17] We can then compute the total surface area and volume used by the slum sector. Likewise, for 2,933 tall buildings in our data, we know the gross floor area, the number of floors, and the height. We then compute lot area and obtain its relation to height and height squared (while adding country fixed effects). For the 12,055 buildings built in 2015 or before, we predict lot area, and obtain the total area, and total volume, of the formal tall sector for each country in 2015. We then add up the volumes and areas of the two sectors, and divide volume and area to obtain average building height. Regressing this on per capita GDP, the height-income elasticity is now higher (1.12***; not shown), not lower. Assuming even slum structures of one storey only (2 meters) returns an even higher elasticity (1.20***; not shown), so our results do not depend on the assumptions made for slums. More generally, adding slums only reinforces our results.[18]

Next, we know land area in 2016 for 1,040 cities from Demographia (2017). However, only 405 of these have tall (80m+) buildings in our data. So far, we arbitrarily assumed that building height in the remaining 635 cities is 40 m, in order to avoid dropping cities with 0 tall buildings when using logged average building height. The elasticities remain similar if we drop cities without tall buildings (col. (2)) or give these cities one third (27 m) or two thirds (53 m) of a tall building (col. (3)-(4)). Elasticities are unchanged if heights are not imputed using the number of floors (col. (5)) or if we use raw architectural heights (not shown). Next, because the website of CTBUH (2018) is in English or Mandarin, we verify results hold if we drop countries where these are not official languages (col. (6)), in case their tall building stocks are more likely to be mismeasured. Lastly, taller buildings could be better reported than smaller buildings, because there are fewer of them and they stand out more. The results of the population-weighted regressions hold if we only use buildings above the 25th percentile height (100 m, col. (7)), median height (125 m, col. (8)) or mean height (135 m, col. (9)). Finally, results hold if we use weights that makes the main sample of 1,010 cities representative of the income distribution of the full sample of 1,773 cities (col. (10)).

Is it possible that measurement error in building stocks could be correlated with per capita income?[19] One could imagine that more buildings are reported in richer countries, because their real estate industry is more mature and CTBUH (2018) has more experts and contributors from these countries. In that case, we over-estimate the elasticity. Conversely, we could imagine that poorer countries have fewer tall buildings, making them easier to report.[20] The elasticity is then under-estimated, which is less of an issue for us. We verify that results hold if we drop countries in the bottom or top 10% in income per capita (col. (11)-(12)). Alternatively, we can drop cities in the top 10% in terms of number of buildings (col. (13)), in case cities with more tall buildings have more misreported buildings. Next, tall buildings could be reported more accurately in smaller

[17]Per capita space availability is 2.73 sq m in Mumbai's slums (The Indian Express, 2016). Mumbai's slum typically have one storey. However, many slum structures in the rest of the world have multiple storeys (UN-Habitat, 2003).

[18]This analysis ignores the formal non-tall sector for which international data is simply non-existent.

[19]Again, classical measurement error in the dependent variable, tall building heights, would only lower precision.

[20]For example, in our data, New York has 1,247 tall buildings but cities like Cairo or Rio have fewer than 50.

cities, since they stand out more there. However, large cities may have more people reporting buildings. Because city size does not vary with income, this should not be an issue. Yet, we verify results hold when we drop cities in the bottom or top 10% of city size (col. (14)-(15)).[21] Finally, elasticities are unchanged if we drop each country one by one (not shown). Lastly, elasticities increase if we use city per capita GDP (col. (16)), and decrease if lights data is used (col. (17)).[22]

F Does the SUM Hold for Both Rich and Poor Countries?

In this section, we examine if the SUM's main prediction for the analysis of urban development within countries holds. In particular, equation (6) in the main text suggests that the within-country city population size-income relationship should not necessarily differ across countries of different income levels, implying that urban systems do not behave differently around the world.

F1. Available Estimates in the Literature

The previous section suggests that the SUM holds in both rich countries and poor countries. While we showed that its implications for the analysis of urban development across countries hold, we now examine if its main prediction for the analysis of urban development within countries holds. Equation (20) contains the new proposition that the city population size-income elasticity, the inverse of the urban wage premium, is a constant that depends on the housing expenditure share and the share of land in housing. It is possible to calculate the elasticity for the U.S. because there are well accepted values for the share of housing in consumption in income, $\eta = 0.3$, and for the share of land in housing, $\beta = 0.2$ (Glaeser and Gyourko, 2018). Using these values, the calculated elasticity is 20. Inverting this expression, we obtain a premium of 0.05, so a 10% rise in population generates a 0.5% rise in earnings. The general consensus is that the premium is 0.05-0.06 for the U.S. (Combes and Gobillon, 2015). Recent estimates by Chauvin et al. (2016) place the urban wage premiums for the U.S., Brazil, China, and India at 0.05, 0.05, 0.09, and 0.08, respectively.[23] Given measurement concerns and issues created by impediments to population mobility, there estimates are remarkably consistent. Thus, available estimates in the literature appear consistent with eq. (20). But empirical testing of the proposition should be accomplished with a much larger sample of countries, even at the sacrifice of data quality. We turn to this exercise now.

We focus our analysis on 31,361 urban polygons/agglomerations of at least 1,000 inhabitants in 223 countries in 2000 according to CIESIN (2017). We use 1,000 inhabitants instead of 300,000 as in the previous section because we need to have enough cities for each country to estimate individual urban wage premiums (some countries have few cities). We will show that results hold when using higher population cut-offs. Next, given consistent income data does not exist for enough individual cities, we use for each city the sum of (radiance calibrated) night lights in 2000 (NGDC, 2015).[24] In the end, we obtain 28,259 world cities with population estimates and night

[21]Elasticities are mechanically lower when dropping the top cities since tall buildings are a feature of large cities. Likewise, elasticities are lower, at 0.25*** and 0.51***, if we drop the primate cities (N = 863, not shown).

[22]Since the elasticity of night lights per capita with respect to national per capita GDP is 1.29*** once we control for population (N = 1,009; not shown), the implied elasticities with respect to national per capita GDP are 0.14 and 0.39.

[23]These are OLS estimates. IV estimates for China and India exhibit wide variance.

[24]This night lights data is not top-coded. Unlike the stable lights product on which many researchers have relied,

lights in 2000. Lastly, we examine if population captures the sum of night lights across cities, first *within* countries, and then *across* countries.

F2. Urban Wage Premiums for Cities in Countries across the World

We first consider two countries with enough cities to achieve precise estimates, the U.S. and India (N = 5,034 and 2,733, respectively), and study how the log sum of night lights correlates with log population in 2000. As seen in Appx. Fig. A7, the slope is similar for both countries, at 1.04*** for the U.S. and 1.11*** for India. The intuition behind a positive wage premium in larger cities based on cost of living is not difficult. However, if larger cities have better housing and commuting technologies or amenities, they attract a larger population without having to offer higher wages, which leads to a flatter relation. While these factors can vary across countries, it seems they do not disproportionately vary across the whole urban spectrum between the U.S. and India. Although Mumbai, Delhi and Kolkata offer higher amenity and access to better technologies than other cities in India the same way New York, Los Angeles and Chicago do in the U.S., the relation between log population and log earnings remains linear even for the largest cities.

The level of night lights for a given level of population is higher in the U.S. than in India. New York and Mumbai have similar population levels, but New York is much brighter. Likewise, if we pool the two samples and add country fixed effects, we obtain a slope of 1.06***, with a R-squared of 0.80 (N = 7,767, not shown). The coefficient of the U.S. effect is 3.74***. U.S. cities thus have 3.74 log points more of night lights than Indian cities. This is a noticeable difference: The log sum of night lights is equal to 7.9 for the average U.S. city, and 16.0 for New York.

The same analysis can be conducted for our full sample of world cities. More precisely, we focus on 171 countries with at least 5 cities of 1,000 inhabitants or more (N = 28,259). Results are reported in Appendix Table A6.

Urban Wage Premiums for Countries of the World. Adding country fixed effects, the coefficient is 1.11*** (Col. (1) of Row 1; R2 = 0.83). The urban wage premium for the world is then 11%.

There are several challenges to estimating the urban wage premium when micro data on workers is not available. First is the possibility for human capital per worker to rise with city size. When differences in human capital are an omitted variable, any positive correlation with city size biases the coefficient relating city size to earnings upward. In the case of our estimates, which are based on night lights (or GDP below) per capita, there is a further issue raised by the possibility that the share of labor in output varies with city size. In this case, if the share of labor varies inversely with city size, the estimate premium is also biased upward. Therefore, we anticipate that the estimated elasticity of city size with respect to population may be biased upward. However, we do not know how the size of any bias would vary with the level of income of the country.

We obtain somewhat similar results when using other data sets. For these, we also restrict our analysis to countries with at least 5 observations, still a small number to estimate an urban wage premium. In column (2), we use data for 1,590 cities above 300,000 inhabitants (for the year

this data indeed records levels of luminosity beyond the normal digital number upper bound of 63.

2015) in United Nations (2018). In col. (3)-(4), we use data for 593 cities (2000 and 2015) in Oxford Economics (2019). This data set is useful because city per capita GDP is available. In col. (5), we use data for 11,261 cities above 50,000 inhabitants (2015) in European Commission (2018). For this data set, we use the night lights estimates they provide. Finally, in col. (6)-(7), we use data for 250 cities (2003 and 2012) in OECD (2018). For this data set, city per capita GDP is also provided but Mexico is the only developing country included. Overall, and with only one exception (col. (2) of row 1), the urban wage premium for the world is close to, or above, 10%. Finally, the slope coefficients should be better measured in countries with more city-observations. If we use as weights the number of observations in the country, we also obtain similar coefficients and R2 (col. (8)).

Urban Wage Premium for Each Country. The slope of the log night lights-log population relationship is estimated for each individual country. As can be seen in Appx. Fig. 8(a), the estimated slope coefficient is close to 1.10 for most countries.[25] As predicted based on the SUM, the slope coefficient is not correlated with national per capita GDP. When regressing the former on the latter, we find a non-significant effect of -0.00 (R2 = 0.00, col. (1) of row 2). The relation varies little across income levels, which is an impressive feature of urban systems. Note that these results are not driven by countries having only a few cities. If we focus on countries with at least 10, 25 or 50 cities, the coefficient remains small and non-significant, at -0.02 (N = 144; R2 = 0.01), -0.02 (N = 106; R2 = 0.02) and 0.01 (N = 68; R2 = 0.00), respectively (not shown). Conversely, if we restrict the sample to cities above 300,000 inhabitants, the coefficient is unchanged, at -0.03 (N = 44; R2 = 0.01; not shown).

Next, if we use the alternative data sets or the number of city-year observations in each country as weights, we still find that slope coefficients are not, or little, correlated with national per capita income (see col. (2)-(8) of row 2). In addition, the R-squared are usually small.

Country Effects. If we plot the country effects (relative to the U.S.) against national per capita GDP (PPP, constant 2011 international dollars) in 2000, we find a strong and significant negative relation (Appx. Fig. 8(b)). The coefficient of log per capita GDP is 1.00*** (R2 = 0.68: N = 170; see col. (1) of row 3). If we restrict the sample to countries with at least 10, 25 or 50 cities, the coefficient remains positive and significant, at 1.01*** (N = 144; R2 = 0.68), 1.02*** (N = 106; R2 = 0.74) and 1.08*** (N = 68; R2 =0.72), respectively (not shown).

Next, if we use the alternative data sets or the number of city-year observations in each country as weights, level differences between countries remain strongly correlated with national per capita income, as shown by the large point estimates and the high R-squared (see col. (2)-(8) of row 3).

Summary. Urban systems appear to behave relatively similarly across countries, thus increasing our confidence that the SUM can be used to make sense of the empirical patterns we observed.[26] The other important fact is that rich country cities are much wealthier than poor country cities for a given city size, and these level differences are almost entirely explained by the income levels

[25]It is significantly different from 1.11 for only about one third of the countries (not shown).

[26]Coming back to equation 20, if the share of land in housing and the share of housing in income are both one third, the urban wage premium is about 8%. Unfortunately, we do not know how these parameters vary between rich and poor countries, but if they vary, they must do so in a way that is neutral for the urban wage premium.

of the countries. Thus, across countries, it is not surprising that city population size does not vary with city per capita income. Likewise, because this holds for smaller and larger cities, urban systems population size does not vary with national incomes.

REFERENCES

Atlas of Urban Expansion, *Atlas of Urban Expansion.* 2016.

Brueckner, Jan K., "A Note on Sufficient Conditions for Negative Exponential Population Densities," *Journal of Regional Science*, 1982, *22* (3), 353–359.

__, "Chapter 20 The structure of urban equilibria: A unified treatment of the muth-mills model," in "Urban Economics," Vol. 2 of *Handbook of Regional and Urban Economics*, Elsevier, 1987, pp. 821 – 845.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio, "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States," *Journal of Urban Economics*, 2016, pp. –.

CIESIN, *Global Rural-Urban Mapping Project. Verion 1 (GRUMPv1): Urban Extent Polygons, Revision 01. Center for International Earth Science Information Network* 2017.

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in "in," Vol. 5, Elsevier, 2015, chapter Chapter 5, pp. 247–348.

CTBUH, *The Skyscraper Center: The Global Tall Building Database of the CTBUH.* 2018.

Demographia, *World Urban Areas (500,000+): Population, Density.* 2017.

Emporis, 2019.

European Commission, *Global Human Settlement* 2018.

Glaeser, Edward and Joseph Gyourko, "The Economic Implications of Housing Supply," *Journal of Economic Perspectives*, Winter 2018, *32* (1), 3–30.

Kim, Kyung-Hwan and John F. McDonald, "Sufficient Conditions for Negative Exponential Densities: A Further Analysis," *Journal of Regional Science*, 1987, *27* (2), 295–298.

Mills, Edwin S., *Studies in the Structure of the Urban Economy.* 1972.

Muth, R., *Cities and housing: the spatial pattern of urban residential land use.* 1969.

NGDC, *Global Radiance Calibrated Nighttime Lights* 2015.

OECD, *OECD Metropolitan Areas Database* 2018.

Oxford Economics, *Global Cities 2030s, Prepared for the World Bank* 2019.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, December 1982, *90* (6), 1257–78.

Rosen, Sherwin, "Wages-Based Indexes of Urban Quality of Life," in Peter Miezkowski and Mahlon Straszheim, eds., *Current Issues in Urban Economics*, Johns Hopkins, 1979.

The Indian Express, *Average living space in Mumbai: Each resident has just 8 sq m to call own* 2016.

UN-Habitat, *Global Urban Indicators Database Version 1* 1993.

__, *The Challenge of Slums* 2003.

United Nations, *World Urbanization Prospects: The 2018 Revision* 2018.

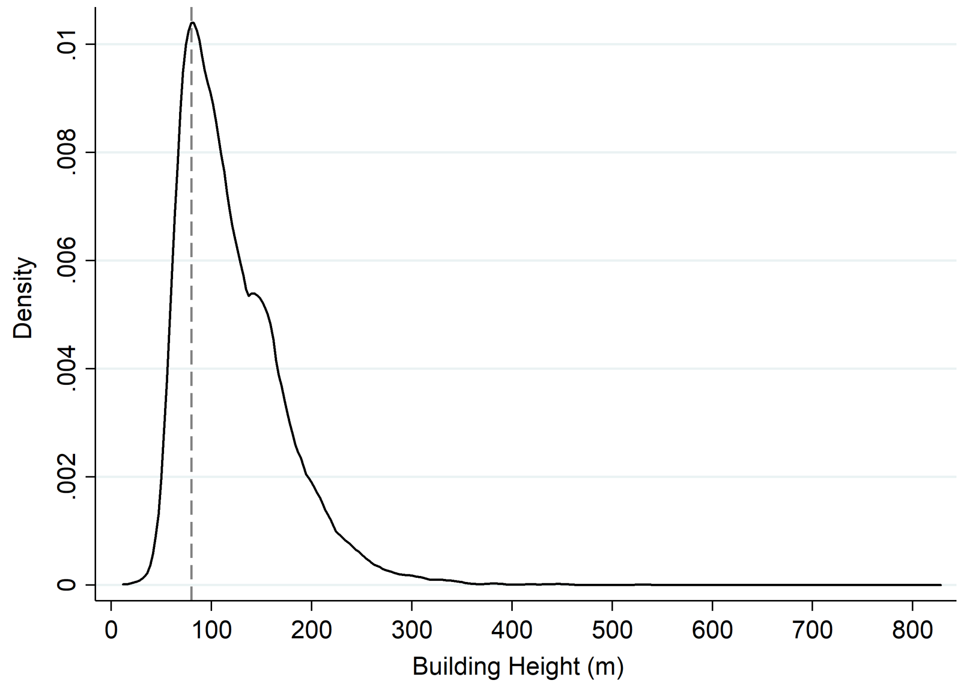
USGS, *Minerals Yearbook* 2020.

World Bank, *World Development Indicators* 2018.

World Cement, *Global Trading Patterns in Cement* 2013.

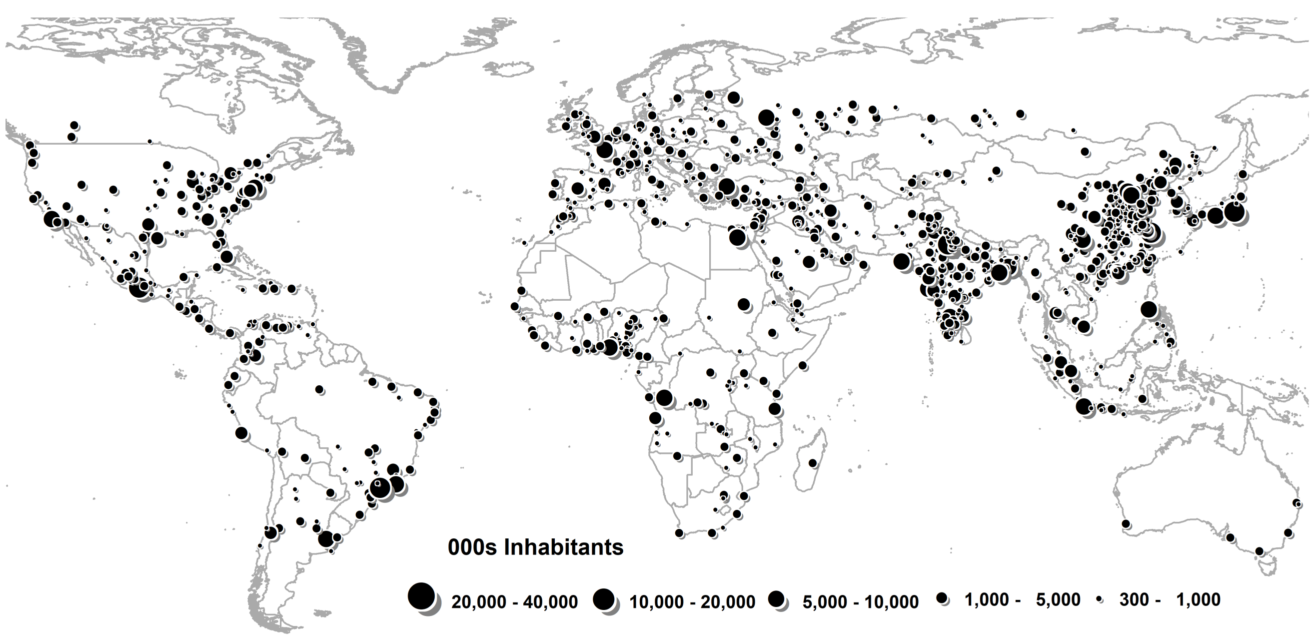
Zhao, Weihua, "The unitary elasticity property in a monocentric city with negative exponential population density," *Regional Science and Urban Economics*, 2017, *62*, 1 – 11.

Figure A1: **KERNEL DISTRIBUTION OF BUILDING HEIGHTS**



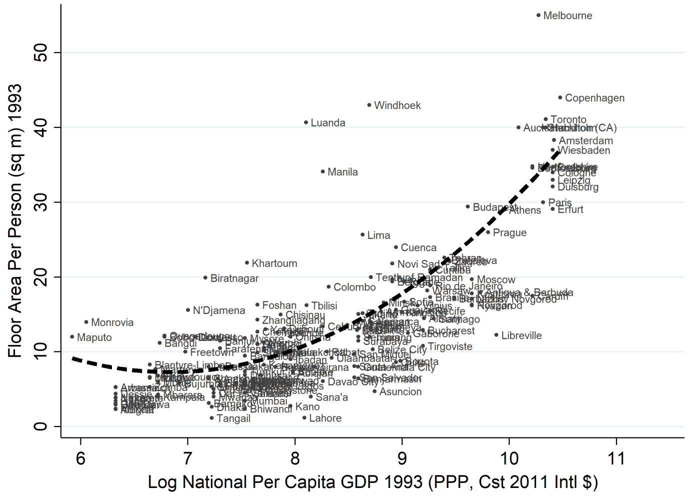
Notes: This figure shows for 19,054 tall buildings in the CTBUH (2018) data set the Kernel distribution of heights (m). As can be seen, the mode of the distribution is 80 m. Data on building heights comes from CTBUH (2018).

Figure A2: **1,010 URBAN AGGLOMERATIONS OF THE MAIN SAMPLE**



Notes: This figure shows the 1,010 urban agglomerations of the main sample. The 1,010 agglomerations are selected as follows: (i) We first select 1,773 urban agglomerations of more than 300,000 inhabitants in 2015 (source: United Nations (2018)). These are urban agglomerations, so they include both the central city and peripheral areas, and should thus be thought as commuting zones; (ii) Among these 1,773 urban agglomerations, we know area for only 1,010 of them (source: Demographia (2017), which focuses on urban agglomerations of more than 500,000 inhabitants circa 2017).

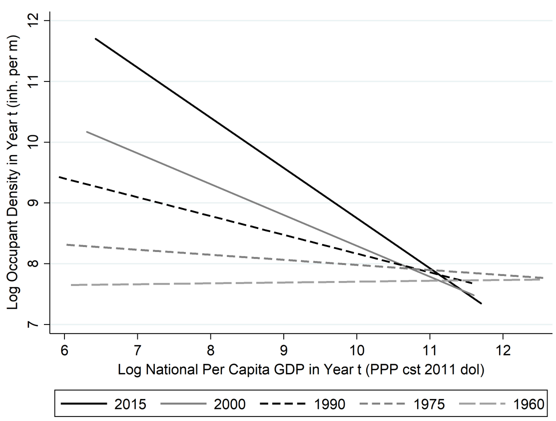
Figure A3: **ECONOMIC DEVELOPMENT AND CITY FLOOR AREA PER PERSON, 1993**



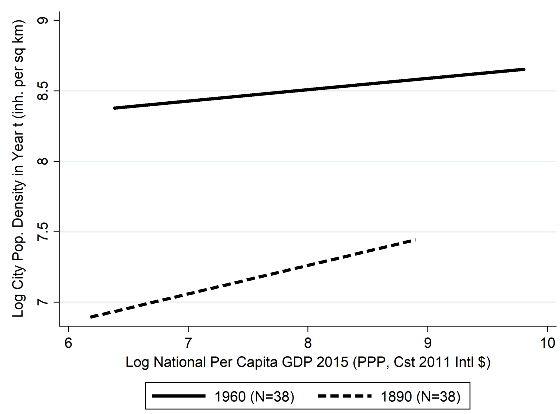
Notes: This figure shows for 181 urban agglomerations the relationship between floor area per person (sq m) in 1993 and log national per cap. GDP (PPP and cst 2011 intl $): Y = -39.82*** + 6.54*** X (N = 181; R2 = 0.56). Data on floor area per person comes from UN-Habitat (1993). Data on per capita GDP comes from World Bank (2018).

Figure A4: **EVOLUTION OF SELECTED RELATIONS, 1890-2015**

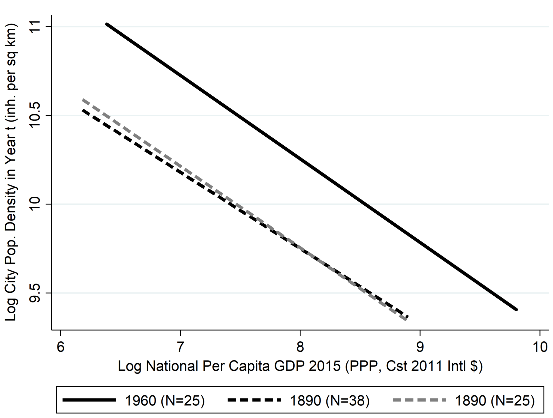
(a) Occ. Dens. (Resid.)-Income Relation 1960-2015



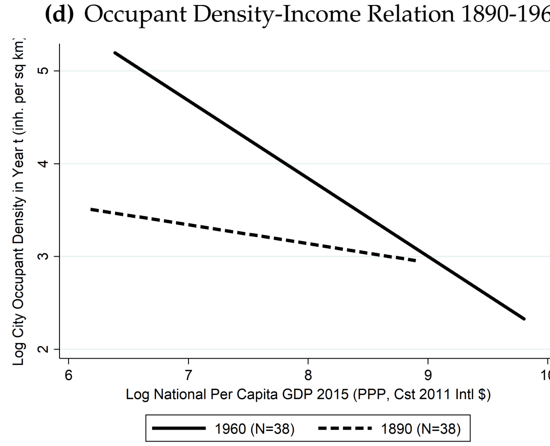
(b) Population-Income Relation 1890-1960



(c) Population Density-Income Relation 1890-1960



(d) Occupant Density-Income Relation 1890-1960



Notes: Subfig. 4(a) shows the relation between occupant density calculated using residential buildings only and national per cap. GDP in 1960-2015. Subfig. 4(b) shows the relation between population size and national per cap. GDP in 1890-1960 (N = 38). Subfig. 4(c) shows the relation between pop. density (available for 38 large world cities in 1890 and 25 large world cities in 1960) and national per cap. GDP in 1890-1960. Subfig. 4(d) shows the relation between occupant density and national per cap. GDP in 1890-1960 (N = 38). City pop. sizes in the same year are used as weights.

Figure A5: **RELATION AMONG GRADIENTS FOR HOUSE PRICE, LAND RENT, STRUCTURE DENSITY AND POPULATION DENSITY**

Figure A6: **RAISING CITY RADIUS AND CITY RENTS BY RAISING CITY EARNINGS**

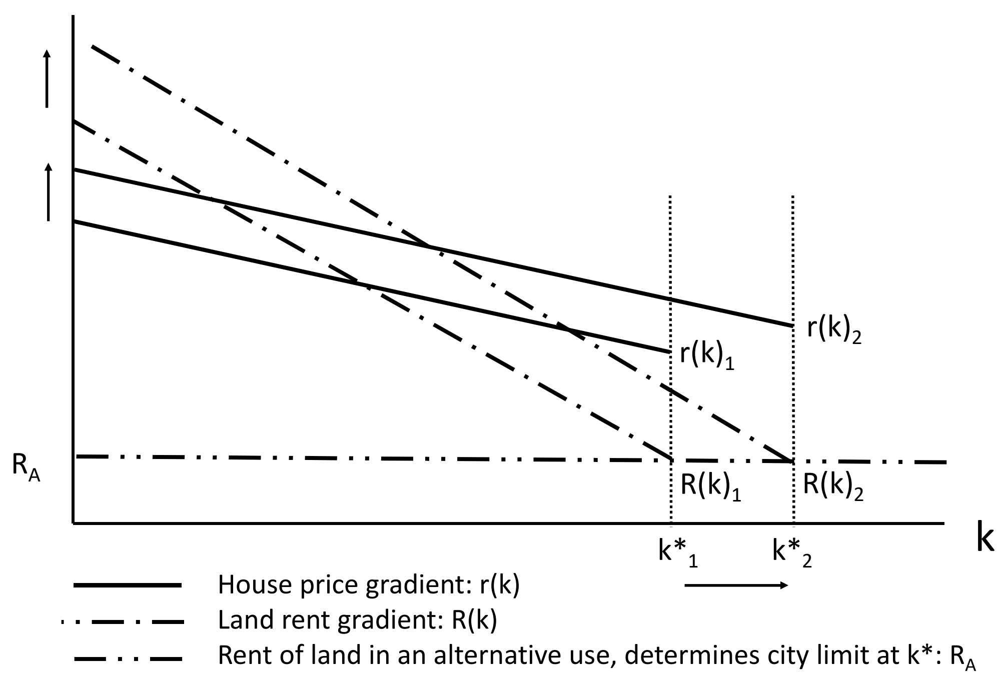
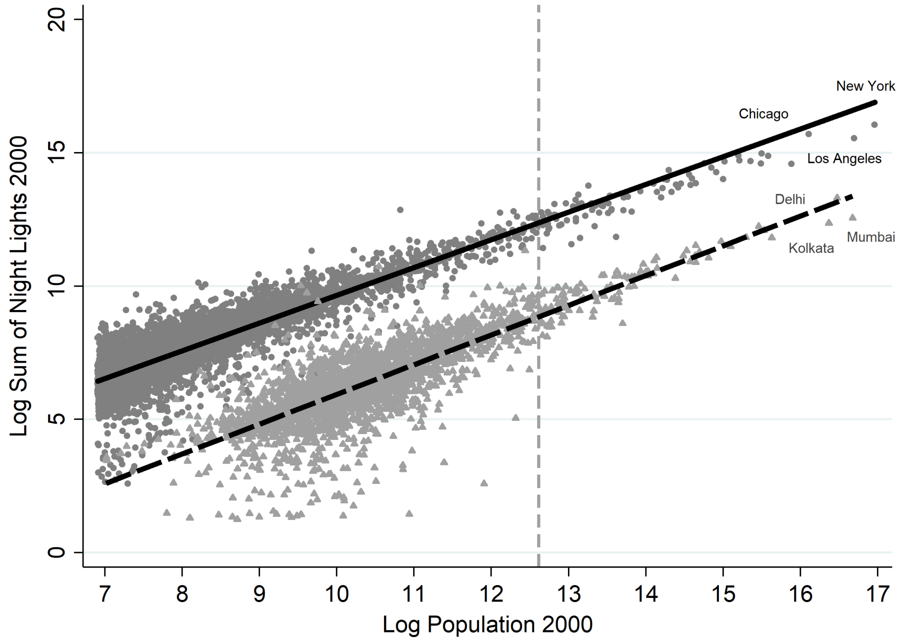


Figure A7: **CITY POPULATION AND CITY NIGHT LIGHTS, U.S. AND INDIA, CIRCA 2000**

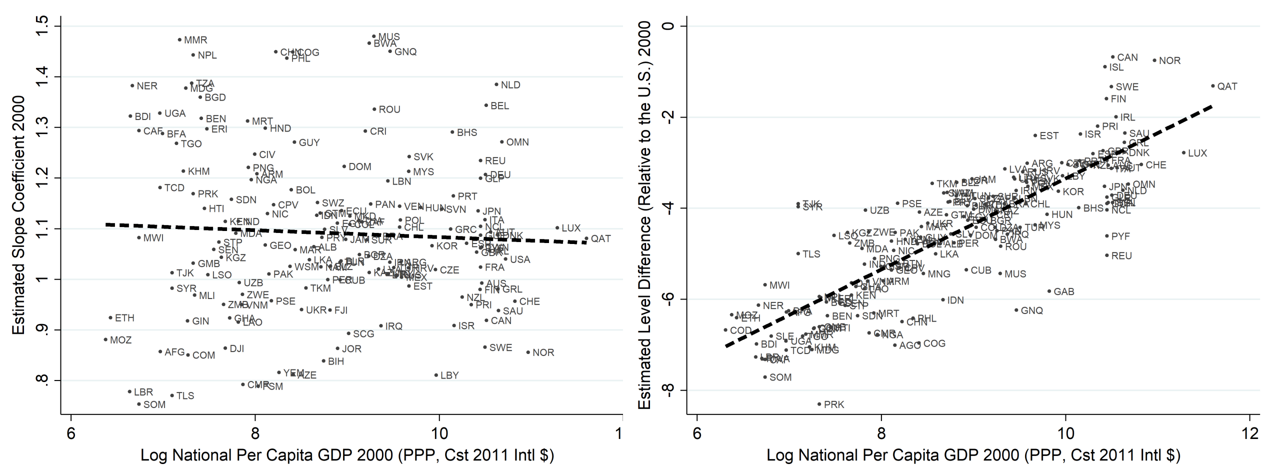


Notes: This figure shows for 5,034 urban agglomerations of more than 1,000 inhabitants in the United States and 2,733 urban agglomerations of more than 1,000 inhabitants in India the relationship between log population in 2000 and the log sum of night lights in 2000 (for the U.S.: Y = -0.74*** + 1.04*** X; R2 = 0.80; N = 5,034; for India: Y = -5.21*** + 1.11*** X; R2 = 0.65; N = 2,733). The dashed vertical line separates urban agglomerations below vs. above 300,000 inhabitants. Data on city population sizes comes from CIESIN (2017). Data on city night lights comes from NGDC (2015).

Figure A8: **SLOPE COEFFICIENTS, LEVEL DIFFERENCES & NATIONAL INCOMES, 2000 (FOR CITIES l IN COUNTRY c, LOG NIGHT LIGHTS$_{l,c}$ = a + b* LOG POP.$_{l,c}$ + λ_c + $u_{l,c}$)**

(a) Estimated Slope Coefficients (b's) & Incomes

(b) Estimated Level Differences (λ's) & Incomes



Notes: This figure shows for 170-171 countries with at least 5 urban agglomerations of 1,000 inh. or more the relationship between the estimated slope coefficient – of the relationship between the log sum of night lights and log city population size in 2000 – and estimated level difference (relative to the United States, the omitted country fixed effect) and log mean per capita GDP (PPP and constant 2011 international $) for all available years in 1998-2002 (left: Y = 1.13*** - 0.00 X; R2 = 0.00; N = 171; right: Y = -6.67*** + 1.00*** X; R2 = 0.68; N = 170). Data on city populations comes from CIESIN (2017). Data on city night lights comes from NGDC (2015). Data on national per capita GDP comes from World Bank (2018).

Table A1: **RELATIVE SIZE OF CITIES IN DEVELOPING COUNTRIES, CIRCA 2015**

Dimension:	Population	Land Area	Interior Space	National PCGDP	Night Lights
Panel A:	% in Developing Country Cities, Including Upper-Middle Income Countries				
1. Unbalanced Sample	74.2	50.3	35.9	42.1	23.7
Observations	1,773	1,010	1,773	1,773	1,765
2. Balanced One (Data for the 5 Dim.)	74.4	50.3	35.6	42.1	27.2
Observations	1,009	1,009	1,009	1,009	1,009
Panel B:	% in Developing Country Cities, Excluding Upper-Middle Income Countries				
1. Unbalanced Sample	30.8	16.6	7.7	9.1	5
Observations	1,773	1,010	1,773	1,773	1,765
2. Balanced One (Data for the 5 Dim.)	31.7	16.6	7.5	9.4	5.4
Observations	1,009	1,009	1,009	1,009	1,009

Notes: Panel A: This table shows the respective percentage contributions of urban agglomerations in developing countries (based on the classification of the World Bank in 2016, so low-income countries, lower-middle income countries and upper-middle income countries) to the total population (source: United Nations (2018)), total land area (sq km; source: Demographia (2017)), total interior space (proxied by the sum of building heights (m); source: CTBUH (2018)), sum of national GDP (PPP and constant 2011 international $; source: World Bank (2018)), and sum of night lights (source: NGDC (2015)) in the sample of urban agglomerations for which data is available. Balanced one: Sample for which all 5 variables are available. *Panel B:* We show the same shares if we reclassify upper-middle income countries as developed countries, thus only considering low-income countries and lower-middle income countries as developing countries.

Table A2: **POPULATION-INCOME ELASTICITY, ROBUSTNESS CHECKS**

	(1)	(2)	(3)	(4)	(5)	(6)
Check:	2015	Demographia	GHS	AUE	City PCGDP	City Lights PC
1. No Pop. Weights	0.02	0.00	-0.08	0.07	0.09	-0.01
	[0.03]	[0.03]	[0.05]	[0.09]	[0.06]	[0.02]
2. Pop. Weights	0.05	0.03	0.03	0.10	0.15	-0.04
	[0.09]	[0.07]	[0.16]	[0.08]	[0.10]	[0.06]
Observations	1010	1040	1977	200	747	1852
	(7)	(8)	(9)	(10)	(11)	(12)
Check:	City PCGDP	City Lights PC	2000	1990	1975	1960
1. No Pop. Weights	0.09	0.00	0.13***	0.25***	0.33***	0.43***
	[0.05]	[0.02]	[0.03]	[0.06]	[0.07]	[0.08]
2. Pop. Weights	0.15	-0.02	0.17*	0.27**	0.32***	0.33***
	[0.09]	[0.06]	[0.10]	[0.11]	[0.09]	[0.06]
Observations	694	1009	1006	1005	1005	1004

Notes: This table shows that the income elasticity of city population size remains similar to the baseline results if we implement various robustness checks described in the main text. The specifications in row 1 and row 2 are the same as in col. (1) and (2) of Table 2, respectively. In row 2, city population sizes in the same year are used as weights. Standard errors are clustered at the country level.

Table A3: **LAND AREA-INCOME ELASTICITY, ROBUSTNESS CHECKS**

Check:	(1) Baseline	(2) AUE	(3) GHS	(4) GHS Built (B)	(5) Repr. World	(6) City PCGDP
1. No Pop. Weights	0.52***	0.55***	0.27***	0.46***	0.52***	0.49***
	[0.13]	[0.13]	[0.07]	[0.10]	[0.13]	[0.10]
2. Pop. Weights	0.58***	0.63***	0.32***	0.53***	0.57***	0.59***
	[0.12]	[0.13]	[0.10]	[0.11]	[0.12]	[0.09]
Observations	1,010	162	1,749	1,003	1,010	694
Check:	(7) City Lights PC	(8) AUE 2000	(9) AUE 1990	(10) GHSB 2000	(11) GHSB 1990	(12) GHSB 1975
1. No Pop. Weights	0.30***	0.62***	0.72***	0.43***	0.44***	0.50***
	[0.07]	[0.11]	[0.09]	[0.08]	[0.09]	[0.09]
2. Pop. Weights	0.34***	0.58***	0.68***	0.48***	0.51***	0.55***
	[0.07]	[0.13]	[0.12]	[0.10]	[0.11]	[0.12]
Observations	1,009	161	161	1,003	1,001	1,002

Notes: This table shows that the income elasticity of land area remains similar to the baseline results if we implement various robustness checks described in the main text. In row 2, city pop. in the same year are used as weights. SEs clustered at the country level.

Table A4: **BUILDING HEIGHTS-INCOME ELASTICITY, ROBUSTNESS CHECKS**

Check:	(1) Baseline	(2) No Build.	(3) Build 1/3	(4) Build 2/3	(5) No Floors	(6) Engl. Mand.
1. No Pop. Weights	0.34***	0.34**	0.43***	0.28***	0.34***	0.50***
	[0.06]	[0.16]	[0.07]	[0.06]	[0.06]	[0.10]
2. Pop. Weights	0.71***	0.57***	0.78***	0.66***	0.73***	0.87***
	[0.10]	[0.17]	[0.10]	[0.10]	[0.10]	[0.20]
Observations	1,010	405	1,010	1,010	1,010	534
Check:	(7) $\geq$ 25th Pctile	(8) $\geq$ Median	(9) $\geq$ Mean	(10) Repres. World	(11) No Bottom Inc.	(12) No Top Inc.
1. No Pop. Weights	0.24***	0.09	0.04	0.35***	0.49***	0.29***
	[0.06]	[0.07]	[0.08]	[0.06]	[0.10]	[0.08]
2. Pop. Weights	0.66***	0.57***	0.52***	0.71***	0.82***	0.73***
	[0.11]	[0.13]	[0.14]	[0.10]	[0.15]	[0.10]
Observations	1,010	1,010	1,010	1,010	922	908
Check:	(13) No Top Stock	(14) No Bottom Pop.	(15) No Top Pop.	(16) City PCGDP	(17) City Lights PC	(18) AUE 2015
1. No Pop. Weights	0.24***	0.39***	0.27***	0.52***	0.11***	0.71***
	[0.06]	[0.06]	[0.07]	[0.06]	[0.04]	[0.14]
2. Pop. Weights	0.49***	0.73***	0.46***	0.75***	0.25***	0.86***
	[0.10]	[0.11]	[0.08]	[0.09]	[0.07]	[0.13]
Observations	969	908	909	694	1,009	140
Check:	(19) AUE2000	(20) AUE1990	(21) GHSB 2015	(22) GHSB 2000	(23) GHSB 1990	(24) GHSB 1975
1. No Pop. Weights	0.39**	0.09	0.40***	0.22**	0.06	-0.19***
	[0.15]	[0.22]	[0.07]	[0.09]	[0.10]	[0.07]
2. Pop. Weights	0.46**	0.23	0.75***	0.59***	0.44***	0.24
	[0.18]	[0.18]	[0.09]	[0.07]	[0.14]	[0.17]
Observations	139	139	1,007	1,003	1,001	1,002

Notes: This table shows that the income elasticity of building heights is similar to the baseline results if we implement various robustness checks described in the main text. In row 2, city pop. in the same year are used as weights. SEs clustered at the country level.

Table A5: ESTIMATED ELASTICITIES OVER TIME, RESIDENTIAL BUILDINGS

Panel A:	Elasticity of City Building Heights wrt National Per Capita GDP				
	2015	GHS Built 2015	GHS Built 2000	GHS Built 1990	GHS Built 1975
1. No Pop. Weights	0.32***	0.12	0.07	-0.03	-0.19***
	[0.07]	[0.07]	[0.07]	[0.06]	[0.05]
2. Pop. Weights	0.68***	0.47***	0.29***	0.14*	-0.09
	[0.10]	[0.11]	[0.09]	[0.08]	[0.06]
Observations	1,010	1,007	1,003	1,001	1,002
Panel B:	Elasticity of City Interior Space wrt National Per Capita GDP				
	2015	2000	1990	1975	1960
1. No Pop. Weights	0.58***	0.50***	0.41***	0.31***	0.34***
	[0.08]	[0.06]	[0.08]	[0.08]	[0.09]
2. Pop. Weights	1.01***	0.77***	0.65***	0.45***	0.36*
	[0.10]	[0.11]	[0.14]	[0.15]	[0.19]
Observations	1,010	1,006	1,005	1,005	1,005
Panel C:	Elasticity of City Occupant Density wrt National Per Capita GDP				
	2015	2000	1990	1975	1960
1. No Pop. Weights	-0.55***	-0.37***	-0.17	0.02	0.09
	[0.07]	[0.06]	[0.12]	[0.12]	[0.14]
2. Pop. Weights	-0.96***	-0.60***	-0.38**	-0.13	-0.04
	[0.08]	[0.11]	[0.15]	[0.15]	[0.15]
Observations	1,010	1,006	1,005	1,005	1,004

Notes: This table shows the evolution of the elasticities of city average building heights, city total interior space and city occupant density – all constructed using residential buildings only – with respect to national per capita GDP (see text for details). The specifications in row 1 and row 2 are the same as in columns (1) and (2) of Table 2, respectively. In row 2, city population sizes in the same year are used as regression weights. Standard errors are clustered at the country level.

Table A6: RELATION BETWEEN URBAN SIZE PREMIUM AND INCOME, ROBUSTNESS

Source Year t Measure of City Income	Base 2000 Lights (1)	U.N. 2015 Lights (2)	Oxford 2000 PCGDP (3)	Oxford 2015 PCGDP (4)	GHS 2015 Lights (5)	OECD 2003 PCGDP (6)	OECD 2012 PCGDP (7)	Base 2000 Lights (8)
1. World Elasticity, Country FE	1.11***	0.97***	1.08***	1.12***	1.25***	1.09***	1.09***	1.10***
	[0.00]	[0.03]	[0.03]	[0.02]	[0.01]	[0.02]	[0.02]	[0.02]
Number of Cities	28,259	1,590	593	593	11,261	250	250	28,259
R-Squared	0.83	0.72	0.94	0.93	0.80	0.96	0.95	0.78
2. Effect of LPCGDP on Slope Coef.	-0.00	-0.13**	-0.02	-0.09**	-0.16***	-0.07	-0.11	-0.07*
	[0.01]	[0.06]	[0.04]	[0.04]	[0.02]	[0.06]	[0.06]	[0.04]
Number of Countries	171	54	32	32	134	13	13	171
R-Squared	0.00	0.12	0.01	0.15	0.33	0.07	0.11	0.22
3. Effect of LPCGDP on Level Diff.	1.00***	1.34***	1.34***	1.38***	1.40***	1.03***	1.05***	1.20***
	[0.06]	[0.11]	[0.05]	[0.07]	[0.07]	[0.12]	[0.12]	[0.19]
Number of Countries	169	54	32	32	134	13	13	169
R-Squared	0.68	0.77	0.95	0.92	0.75	0.95	0.94	0.64
Min. City Pop. Size (000s)	1	300	222	392	50	422	439	1

Notes: Row 1 shows for various data sets the elasticity of the city sum of night lights or city GDP with respect to city population size (country FE included). Row 2 shows the effect of log national per capita GDP (PPP, cst 2011 intl $; source: World Bank (2018)) on the estimated slope coefficient for each country. Row 3 shows the effect of national per capita GDP on the estimated level difference (relative to the U.S., the omitted country FE). Col. (1): We use the baseline sample of cities available in CIESIN (2017). Col. (2): U.N. corresponds to the main sample of agglomerations of at least 300,000 inh. in United Nations (2018). Col. (3)-(4): Oxford corresponds to the sample of cities available in Oxford Economics (2019). Col. (5): GHS corresponds to the sample of cities available in European Commission (2018). Col. (6)-(7): OECD corresponds to the sample of cities available in OECD (2018). Night lights in 2012 are used as a proxy for 2015 and come from NGDC (2015) except in col. (5) where we use night lights estimates (2015) available in European Commission (2018). Col. (8): We use as regression weights the total number of city-obs. (≥ 5) in each country in the first-step regressions.