

1/June/2020

Building the city: from slums to a modern metropolis

J. Vernon Henderson, Tanner Regan, and Anthony J. Venables

Abstract: We model the building of a city, estimate parameters of the model, and calculate welfare losses from institutional frictions encountered in changing land-use. We distinguish formal and slum construction technologies; in contrast to slums, formal structures can be built tall, are durable, and non-malleable. As the city grows areas are initially developed informally, then formally, and then redeveloped periodically. Slums are modelled as a technology choice; however, institutional frictions in land markets may hinder their conversion to formal usage that requires secure property rights. Using unique data on Nairobi for 2003 and 2015, we develop a novel set of facts that support assumptions of the model, estimate all parameters of the model, and calculate welfare losses of conversion frictions. We track the dynamic evolution of the city and compare it with model predictions. In the core city formal sector, about a third of buildings were torn down over 12 years and replaced by buildings on average three times higher. For slums in older areas near the centre, even after buying out slumlords, overcoming institutional frictions would yield gains amounting to about \$18,000 per slum household, 30 times typical annual slum rent payments.

Keywords: city, urban growth, slums, urban structure, urban form, housing investment, capital durability.

JEL classification: O14, O18, R1, R3, H

Acknowledgements: We gratefully acknowledge the support of an Africa Research Program on Spatial Development of Cities at LSE and Oxford funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. We acknowledge the excellent research assistance of Patricia Freitag, Piero Montebruno and Ilia Samsonov. Thanks to referees, the editor and to seminar participants at LSE, CURE, Berkeley, Pennsylvania, Lausanne, Oxford, Helsinki, Luxembourg, Bristol, USC, Melbourne, Monash, NBER and RIETI Tokyo for helpful comments.

Authors' affiliations & emails: LSE (J.V.Henderson@lse.ac.uk); LSE (t.regan1@lse.ac.uk); Oxford and Monash Universities (tony.venables@economics.ox.ac.uk)

1. Introduction

The urban population of Sub-Saharan Africa is predicted to increase from 450 million to 1250 million over the next 30 years (UN, 2018). The populations of many large cities are increasing by about 50% every ten years, implying a huge demand for increased building volumes. World Bank (2006) suggests that about two thirds of a country's non-governmental capital stock are buildings, and urban construction and maintenance are a large and rising share of many countries' total investment. Yet we know little about this investment process that reshapes cities and redeploys nations' capital stocks. African cities are subject to rapid redevelopment and expansion in the formal building sector; while at the same time, there is formalization of some slums, densification of others, and spread of new slums at city edges. In parts of cities, formal sector high-rise buildings on high price land are adjacent to low-level slums with limited returns to land, suggesting inefficient land-use and distorted investment decisions.

This paper models the process of a city's growth and applies the model to a unique data set we have developed on Nairobi, the capital of Kenya. The paper does four novel things. First, we develop a model of the built environment of a growing city. Given the difficulty of having both continuous time and space in the same model, especially with durable capital, virtually all models of cities are static. Fujita (1982) has an early attempt at a dynamic model and Braid (2001) put forth the first complete example of a dynamic monocentric city model with durable capital. To study the development of rapidly growing cities in middle- and low-income countries a full dynamic framework is essential. We extend and enrich the base model designed for developed countries by adding a slum sector and a role for geographical and institutional frictions, and we develop clear propositions about city growth.

Dynamics involve spread of a city's land area, intensified land use within the city, and demolition and reconstruction with increased density and building height. Our model captures these features, with a key distinction between formal and informal (slum) sectors. Formal buildings involve sunk capital costs, can be built tall, cannot be modified once constructed, and involve investment decisions based on expected future rents. As the city and housing prices grow formal sector buildings are periodically demolished and redeveloped to a greater height. In contrast, in the informal sector, buildings generally are much more malleable and built with non-weight-bearing materials. They are built low, with high land intensity. The building volume delivered by slums increases through time not by building taller, but by increases in crowding and already high cover-to-area ratios. New slums appear near the expanding city edge as the highest and best use of that relatively cheap edge city land. Within the city, the model predicts that as housing prices rise, there should be on-going conversion of older slums to formal sector use.

Second, we assemble a rich data set on the built fabric of Nairobi and present aspects of the 2015 cross-section and its evolution since 2003. While there is work on the USA using census data of building ages (Brueckner and Rosenthal, 2009), we know of no work which utilizes citywide data on individual buildings, with demolition, redevelopment, and infill to detail the changes in the urban landscape. For the Nairobi data set, we have tracings of all buildings in the city from aerial photo images for 2003 and 2015, which give a precise delineation of the built environment. We develop an algorithm to overlay the two cross-sections of polygons and determine which building footprints are unchanged, which

buildings were demolished and/or redeveloped and where and to what extent infill occurs. Building heights are derived from LiDAR data allowing us to measure building volumes. We also have formal sector vacant land prices for 2015, housing rent data for the formal and informal sectors in 2012, and data on land quality throughout the city, covering slope, ruggedness, water cover and the like.

With these data we construct a rich picture of the built environment of Nairobi, both in the spatial cross-section and its evolution through time. The UN (2018) reports a population for the agglomeration of just over 4 million, a population that grew at 4.4% p.a. from the 1999 to 2009 census. We find that the built volume of the 2015 city area increased at 4.0% p.a., expanding by 60% between 2003 and 2015. Within just the 2003 spatial area of the city, overall volume growth was an astounding 47%, accounting for two-thirds of total volume growth within the 2015 city area. Much of this growth involves formal sector redevelopment. For example, at 3-4 km from the centre, redevelopment increased volume by 30%, achieved by tearing down one third of buildings and reconstructing more intensively. We find that newly redeveloped buildings are about triple the height of those torn down. Slower growing developed country cities generally have nowhere near this degree of change.¹

Nairobi's slums deliver similar built volumes per unit land as the formal sector, and house just under 30% of the population in both 1999 and 2009, similar to UN-Habitat's (2016) estimates for many developing country cities. Slum volume growth of 50% between 2003 and 2015 was slower than formal sector volume growth of 61%. As noted above, new slums appear on the city edge, often on land with private property rights, highlighting the on-going role for slums to provide low cost housing in developing countries. However apart from growth at the edge, a key feature is the persistence of several slum areas on high value land near the city centre, most notably Kibera, often referred to as Africa's largest slum.

Third, we estimate all the parameters of the dynamic model based on the characteristics of the 2015 cross-section of the built environment and on land sale and house-rent data, together with the discount rate that we take from World Bank data on Kenya. These parameters describe the intensity of land-use in the informal and formal sectors and gradients telling us how prices, building volumes and heights change with distance from the city centre in both sectors. We then use these parameters and the model to make predictions about the evolution of the built environment and population from 2003 to 2015 and compare these with the evolution we see in the data.

Finally, we calculate the welfare cost of inefficient land-use from delayed formalisation of slums near the city centre. This highlights the role of policy for fast growing cities with major land market failures that deter investment. We argue that institutional frictions in Nairobi delay timely redevelopment of inner-city slums and quantify the losses from such delays. For older slums at 3-5km from the centre, we estimate that the city could pay off illegal slumlords operating in these slums for the value of their land (reflecting their profits) in perpetual slum use. Even after that, formalisation would still bring a gain of about \$18,000 per slum household, in a context where slum households spend on average about

¹ To give perspective on the demolition rate of one third, in the USA, the American Housing Survey data gives demolition and removal by disaster (fire, hurricane and the like). Annual rates of demolition and removal by disaster range from 0.5% to about 1.2%. For 12 years this would involve 6-15% of building removal. Nairobi is thus 2-5 times that.

\$500-700 p.a. on housing. Our calculations do not capture the wider social costs and benefits of formalisation which we discuss, but they do quantify the direct land-use efficiency loss of delayed formalisation.

1.1 Context and literature

Our modelling judgements are based on several key features of Nairobi, features that we think are applicable to many other developing cities. First, we treat Nairobi as a monocentric city. Monocentricity is the workhorse assumption of urban economics, including contemporary papers such as Combes, Duranton and Gobillon (2018) on modern French cities. As we will see, Nairobi appears to be primarily monocentric with building height, volume, and land prices all declining sharply with distance from the centre. This monocentricity appears to be a feature of many major African cities. Lall, Henderson, and Venables (2017, p 21-23) show that African cities like Nairobi, Addis Ababa, Dar es Salaam, and Kigali are much more monocentric than Paris, London, Barcelona, or Atlanta. For our own sample of world cities similar to Nairobi in size, in Appendix Section A2.4, we show that developing country cities and Nairobi in particular display greater monocentricity than most developed country cities, with density dropping off three times more rapidly from the centre.

The second feature is that there are relatively well functioning markets for the allocation of housing within both the formal and informal sectors, but a problem in the conversion of informal to formal sector land-use. While the two sectors offer very different qualities of housing, both are overwhelmingly rental markets; 87% of Nairobi residents overall and 89% of slum dwellers rent, proportions similar to those in many African cities.² The rental markets allocate land between occupants, but the individuals who control slum land and to whom rents are paid often have insecure land rights. This does not impede the functioning of rental markets, but insecure rights do deter those who control the land from sinking the capital required to construct formal sector buildings. Without security of ownership such investments are too risky to be undertaken. This is a problem particularly on government owned – but not managed – lands nearer the city centre in Nairobi. Politically well-connected slumlords ‘illegally’ operate vibrant rental markets, but they cannot switch these lands to formal sector use because that would open their control of the land to challenge. Their ability to operate the slum rental market involves corruption, and a complex political history, as we describe in Section 6. Marx, Stoker, and Suri (2013) highlight the importance of “investment inertia”, where small groups of powerful individuals control large slum housing markets. Formalisation and similar initiatives would challenge entrenched land control, so the status quo of informal status persists. These situations seem to be common in older slum lands near city centres in many developing country cities in Asia and Africa.³

² From Un-Habitat (2011) on Ghana about 75% of residents in Accra and Kumasi rent. From authors’ own calculations from IPUMS about 40% in Dar es Salaam rent. At a country level among urban dwellers as defined in IPUMS for Ethiopia (2007), Nigeria (2010), and Uganda (2002), respectively 51%, 60%, and 58% rent, where we expect higher amounts in primate cities. So while the urban Kenya rate in IPUMS is 56%, in Nairobi as noted in the text it is 87%. Some other countries including Mozambique and South Sudan have higher ownership rates.

³ There is no central source of information on slum locations and sizes. We explored sources including the UN-Habitat Slum Almanac (2016), Davis (2006) and the repository of CIESEN at Columbia University. We compiled a long list of slums near city centres. Well known examples include Dhavari in Mumbai with population estimates of 7-800,000; Manshiyat Nasser (“Garbage City”) in Cairo with its edge within 3km of the centre and population

These cities exhibit the oft-photographed hotchpotch of land uses, with tall formal sector buildings bordering pockets of single storey corrugated iron sheet slum housing in the urban core as institutional frictions prevent conversion of these slums to formal sector development. It is this that motivates our modelling of distinct formal and slum sectors in this context, following the literature that generally makes the same sharp distinction, even if the line between informal and formal sector housing may be fuzzy at times.

Similar institutional barriers to formalisation arise in slum areas around the edge of Nairobi. Here a higher proportion of land is privately owned. However, some owners may have only possessory rights which need to be converted to formal modern title prior to sinking capital in formal sector development. This again involves formalisation costs, in this case in the form of fees, legal costs, delay and possibly bribes. Frictions arise for natural as well as institutional reasons, including poor geographic conditions such as steep slope, ruggedness, marshes and exposure to flooding. We examine these empirically to establish when and where each set of frictions matters. Our results suggest that institutional rather than natural frictions play the dominant role for older slums near the centre.

On other literature, we note that recent papers have varying perspectives in modelling the informal-formal sector distinction. For the Brazilian context, in Cavalcanti, da Mata, and Santos (2019) slum owners make a choice between formalising, which involves paying for secure rights and incurring taxes and regulation, versus staying in the unregulated informal sector with insecure land rights. We do not focus on taxes and regulation, because in Nairobi tax systems generally apply to both sectors and there is very low compliance in either sector. Similarly, land use regulations appear not to bind.⁴ Two recent papers investigate whether slum upgrading programs inhibit timely formalisation. For Jakarta, Harari and Wong (2019) argue that slum improvement programmes serve to delay redevelopment of slums. Similarly, Michaels et al. (2019) argue that, compared to untreated neighbouring slums, improved slums either look no different than untreated ones 2-3 decades later, or have poorer characteristics, in Dar es Salaam. Finally, Getcher and Tsivanidis (2018) argue that in Mumbai very large formal sector developments can spur redevelopment of nearby slum areas. Our work complements these papers looking at the dynamic aspects involved in investing in land and using the dual structure of urban land markets that we think is relevant to many developing country situations.⁵

up to 800,000; and Ga Mashie and Old Fadama in Accra housing about 400,000. City centre slums in Jakarta and Dar es Salaam city slums are analysed in Harari and Wong (2019) and Michaels et al. (2019).

⁴ In Nairobi land use regulations seem to be largely ignored (Mwaura, 2006) and property taxation of formal sector lands has very limited implementation and compliance (Kelly, 2003). Narubi (2015) shows for property taxes compliance in Nairobi in 2014/15 was under 7%, and the collection ratio was about 16%, suggesting only some bigger commercial properties pay these taxes. Furthermore, from the 2010 cadastre for Nairobi county, most slum lands are assessed these same taxes. On regulations, the minimum lot size is generally 500 m², and never less. In Fig. A2.2 of the Appendix, from a full cadastre of the formal sector, we show that modal lot size is at 180 m², 54% of all buildings are in violation of the regulation, and that there is a long right tail for commercial buildings. Most particularly, the curve for that distribution has no mass, step, blip, or kink around 500 m².

⁵ Additionally, Selod and Tobin (2018) model owners as choosing between different gradations of property rights at different costs, an issue relevant for some contexts. Ferreira, Monge-Naranjo and Torres de Mello Pereira (2016) on Brazil focus on modelling how slums may be stepping-stones in skill formation and growth of cities.

The basic model and core theoretical results are set out in Section 2 of the paper. Section 3 presents data and estimates key empirical relationships. Section 4 backs out the parameters of the model, based on empirical evidence, and validates key assumptions of the model. Section 5 analyses the evolution of the built environment and compare model predictions with the data. Section 6 develops the welfare analysis to derive costs of misallocation of land to older slums, and Section 7 concludes.

2. Theory

In this section we develop the model of a growing city, focusing on investment decisions and consequent patterns of land-use and built structures. Sections 2.1 and 2.2 describe housing demand and supply, introducing preferences and technologies for informal (slum) and formal sector buildings. We derive the investment decisions and consequent land-rent that would be earned if land at a particular place and date were in informal or formal sector use (or left for agricultural use outside the city). In Section 2.3 we turn to landholders' dynamic optimisation problems of choice of land-use, i.e. whether, at each place and date, land should be occupied by slums or formal sector buildings, and the timing of formal sector redevelopment. We look first at a particular point in the city and examine its evolution through time as it transitions from agricultural use to informal development, then formalises and goes through successive waves of formal sector demolition and reconstruction. Section 2.4 draws out the way in which his path varies across points in the city, giving a description of both the cross-section of the city and its evolution through time. Section 2.5 adds frictions to the transition process, in particular spatial heterogeneity in formalisation and construction costs. Together, these analytical sections provide the relationships upon which empirics, estimation of key parameters and welfare analysis are based.

2.1 Bid-rents and housing demand

The value of a unit of housing varies across the city and through time, and we start by deriving the bid-rent for housing at places x and date t . A representative household has preferences between consumption of housing and other goods (the numeraire) described by $U(s(x, t)a, w(x, t) - q(x, t)s(x, t))$. In this expression $s(x, t)$ is the quantity of housing space consumed by an individual living at place x at date t , $q(x, t)$ is its observed rental price, and a is an amenity parameter which will be discussed further below. Household income available for housing and purchase of other goods is $w(x, t)$. We take the function $w(x, t)$ to be exogenous and assume that it is increasing in t and weakly decreasing in x . The former is based on the city using labour to produce output with constant returns to scale and experiencing productivity growth relative to the rest of the economy.⁶ The latter can be a simple labelling of places. However we generally interpret x as distance from the city centre, in which case $w(x, t)$ is interpreted as wages at the centre minus commuting costs to distance x .⁷ We impose a particular functional form on $w(x, t)$ from Section 2.4 onwards.

⁶ Agglomeration economies are not included in this model. The price of output is exogenous to the city, and any variations in the output price have the same effect as changes in physical productivity.

⁷ See Appendix Section A1.1 for derivation of these relationships. This is as in the standard urban model, where all households commute to the CBD. Newer approaches offer a richer model of commuting and the dependence of net wages on location (e.g. Ahlfeld et al. 2016). We assume also that households have no other source of

The city is ‘open’, meaning there is free migration between the city and outside, where utility takes exogenous value u_0 . Equilibrium rent leaves all households indifferent between living at each occupied place x in the city or receiving outside utility u_0 . Housing rent $q(x, t)$ therefore satisfies $U^*(q(x, t)/a, w(x, t)) = u_0$, where $U^*(.)$ is the indirect utility function (each household’s location and $s(x, t)$ optimally chosen). Since the amenity factor enters the utility function multiplicatively with $s(x, t)$, it enters indirect utility in the form $q(x, t)/a$. Henceforth, we work with rent per unit amenity, defined as $p(x, t) = q(x, t)/a$, and referred to as space-rent. The net wage $w(x, t)$ therefore determines space-rent through the relationship

$$U^*(p(x, t), w(x, t)) = u_0, \text{ for example, } p(x, t) = (w(t, x)/u_0)^{1/\beta}, \quad (1)$$

where the example has Cobb-Douglas preferences with housing share β and hence chosen quantity $s(x, t) = \beta w(x, t)/ap(x, t)$ (see Appendix Section A1.1). Outside utility is exogenous and will henceforth be suppressed in notation.⁸ We assume that housing-space is a normal good, implying that space-rent $p(x, t)$ is increasing in $w(t, x)$, and hence increasing with t and weakly decreasing with x . The dynamics of the model are therefore driven by urban productivity growth that increases $w(t, x)$, i.e. shifts this horizontal labour demand curve up through time. This increases space-rent, and the consequent growth of city population is determined by the supply response of urban housing.

We note that, while we have different types of building technology, we do not model different types of households. Were we to do so, land-use would be determined by the upper envelope of the land-rent functions generated by different households.⁹ In our empirical work, land-rents in some parts of the city are systematically lower for land in informal use than for equivalent land in formal use. This is indicative of inefficient land-use, regardless of whether the land-rent function is that of a single household type or the upper envelope of many types. We will return to discussing the relevance of having different types of households in Section 6 when we conduct our welfare analysis.

2.2 Building technology and housing supply

There are two distinct building technologies, formal and informal, which supply building volume per unit land in different ways. The formal sector (F) can build tall, and the informal sector (I) can ‘crowd’, increasing cover, i.e. the proportion of land covered by building footprints. The volume of building that a technology can supply on a unit of land at a particular place, x , at time t , is the product of height and cover, $v_i(x, t) = h_i(x, t)c_i(x, t)$, $i = I, F$. Formal sector buildings are durable structures, which incurs sunk costs of construction and increasing marginal cost of building tall. The informal sector is unable to build tall because the materials used are not sufficiently load-bearing, this also giving a ‘malleability’, enabling structures to be rapidly modified and implying that construction costs are not sunk. Informal floor-space can be added at constant marginal cost, but crowding reduces amenity and hence housing rent.

This distinction captures the difference in sunk costs associated with each housing type, and its

income, implying that land-rents are received by non-resident landowners or slumlords.

⁸ u_0 is potentially time varying, but in later analysis it is only the exogenous ratio $w(x, t)/u_0$ that matters.

⁹ See for example Duranton and Puga (2015).

technological basis accords well with the difference between slums and formal sector areas in Nairobi. The 2009 Nairobi Census indicates that, in slums, the majority (about 52%) of households live in buildings where walls are corrugated iron sheets which can be easily reconfigured like Meccano parts; much other slum housing involves mud construction (about 19%) and other material with short duration. These materials are not sufficiently load-bearing to allow much height, and our data show that 85% of slum buildings are under 5 meters tall. In contrast, 84% of formal sector households live in a building that is made of stone or some type of brick or block, many with substantial height. The main missing category in both sectors is wood construction, where wood buildings in slums are mostly very different quality structures than those in the formal sector. While, in reality, the distinction between slums and formal sector developments can be ambiguous, the binary distinction is analytically tractable. We also think the sharp distinction captures the key features of Nairobi and is supported by the literature, including Taubenbock, Kraff and Wurm (2018) who analyse the morphology of slums for a large selection of cities worldwide.

2.2.1 Informal sector

Informal sector construction materials are malleable and construction costs are a flow, occurring continuously through the life of the structure. This can be thought of as either the rental on ‘Meccano parts’ used in construction or as the cost of material whose life is one instant. This sector is unable to build tall, so height is fixed at $h_I = 1$. It can however change at any instant the proportion of each unit of land that is covered with buildings, $v_I(x, t) = c_I(x, t)$. Construction costs per unit volume in this sector are constant κ_I , so construction costs per unit land are $\kappa_I v_I(x, t)$. However, crowding more building onto land in slums reduces housing amenities, which we capture by assuming that the amenity is a diminishing and convex function of cover per unit land, so $a(v_I(x, t))$, $a' < 0$, $a'' > 0$. If land at x, t is in informal use, then land-rent is space-rent times amenity minus construction cost, all times volume per unit land,

$$r_I(x, t) = [p(x, t)a(v_I(x, t)) - \kappa_I]v_I(x, t). \quad (2)$$

The volume of housing supplied at any instant is chosen to maximise land-rent, taking space-rent $p(x, t)$ as exogenous and internalising the effect of crowding on amenity.¹⁰ The first order condition equates marginal revenue to marginal cost,

$$\partial r_I(x, t)/\partial v_I(x, t) = p(x, t)a(v_I(x, t))[1 + v_I(x, t)a'(v_I(x, t))/a(v_I(x, t))] - \kappa_I = 0. \quad (3)$$

Generally, we assume that informal amenity is iso-elastic in cover, $a(v_I(x, t)) = a_I v_I(x, t)^{(1-\alpha)/\alpha}$, $\alpha > 1$. Then the chosen volume of housing and maximised value of land-rent are

$$v_I(x, t) = [a_I p(x, t)/\kappa_I \alpha]^{\alpha/(\alpha-1)}, \quad (4)$$

$$r_I(x, t) = (\alpha - 1)\kappa_I v_I(x, t) = \kappa_I(\alpha - 1)[a_I p(x, t)/\kappa_I \alpha]^{\alpha/(\alpha-1)}. \quad (5)$$

¹⁰ A simplification is that we do not explicitly model this as an externality in which each developer’s choice of crowding affects neighbours.

The loss of amenity due to crowding, α , is a key parameter. It will be recovered in the empirics based on how informal sector volume in (4) varies with distance from the city centre. The parameter also determines the division of revenue between land-rent and construction costs. Eqn. (5) indicates that land-rent is multiple $(\alpha - 1)$ of construction costs, so total space-rent earned per unit land is divided with proportion $(1 - 1/\alpha)$ going as land rent, and $1/\alpha$ to construction. Iso-elasticity implies further that informal sector space-rent adjusted for amenity is constant throughout the city at value

$$p(x, t)a_I v_I(x, t)^{(1-\alpha)/\alpha} = \kappa_I \alpha. \quad (6)$$

Essentially, and as will be confirmed later in the space-rent data, increased crowding near the city centre offsets the advantage of improved access to the centre.

2.2.2 Formal sector

The formal sector differs in that buildings are durable, construction is a sunk cost, and volume is achieved by varying height, not cover. Height is chosen at date of construction, denoted τ_i , so $h_F(x, \tau_i)$ is fixed for the life of the structure, i.e. until demolition at date τ_{i+1} , where subscript $i = 1, 2, \dots$ denotes successive redevelopments (or ‘generations’) of formal structures. For simplicity, but also based on the cross-section data, we assume that formal sector land cover is uniform at $c_F = 1$, so the volume of building at x developed (or redeveloped) at τ_i is constant at $v_F(x, t) = v_F(x, \tau_i) = h_F(x, \tau_i)$ for $t \in [\tau_i, \tau_{i+1}]$. Construction costs per unit land are one-off and sunk, and are an increasing and convex function of height (= building volume) on that land, $k(v_F(x, \tau_i))$, $k' > 0$, $k'' > 0$. Demolition incurs no costs.

We assume there is no amenity loss or gain from building tall, so we set $a_F = 1$. The space-rent of a unit of building volume is $p(x, t)$, exogenous to the developer. The present value of land-rent that accrues over the life of a structure, $t \in [\tau_i, \tau_{i+1}]$, discounted to construction date τ_i at interest rate ρ is denoted $R_F(x, \tau_i)$. With sunk costs $k(v_F(x, \tau_i))$ and volume fixed at the date of construction this is

$$R_F(x, \tau_i) = v_F(x, \tau_i) \int_{\tau_i}^{\tau_{i+1}} p(x, t) e^{-\rho(t-\tau_i)} dt - k(v_F(x, \tau_i)). \quad (7)$$

We define the ratio of the present value of space-rent per unit volume over its life relative to space-rent at date of construction as

$$\Phi(x, i) \equiv \int_{\tau_i}^{\tau_{i+1}} [p(x, t)/p(x, \tau_i)] e^{-\rho(t-\tau_i)} dt, \quad (8)$$

so $R_F(x, \tau_i) = p(x, \tau_i) \Phi(x, i) v_F(x, \tau_i) - k(v_F(x, \tau_i))$. The integral $\Phi(x, i)$ is akin to the ‘value-to-rent ratio’ on a newly constructed property in the terminology of the real-estate literature (noting the time horizon in (8) is cut at τ_{i+1}).

If land at place x undergoes formal sector development (or re-development) at date τ_i , the volume chosen is intuitive and given by

$$\partial R_F(x, \tau_i) / \partial v_F(x, \tau_i) = p(x, \tau_i) \Phi(x, i) - k'(v_F(x, \tau_i)) = 0. \quad (9)$$

If the cost function is iso-elastic, $k(v_F) = \kappa_F v_F^\gamma$, $\gamma > 1$, then chosen volume and the maximised present value of land-rent are ¹¹

$$v_F(x, \tau_i) = \left[\frac{p(x, \tau_i) \Phi(x, i)}{\kappa_F \gamma} \right]^{1/(\gamma-1)}, \quad (10)$$

$$R_F(x, \tau_i) = \kappa_F (\gamma - 1) \left[\frac{p(x, \tau_i) \Phi(x, i)}{\kappa_F \gamma} \right]^{\gamma/(\gamma-1)}. \quad (11)$$

The diseconomy in building taller, γ , is another key parameter and will be estimated based on how the height of newly constructed buildings and hence volume in (10) varies with location as space-rent varies.

To define land-rent share, it is useful to have a continuous flow measure of land-rent, given by amortizing the one-off construction cost continuously over the life of the structure. If amortisation is a constant proportion μ of revenue over the life of the building, i.e. $\mu p(x, t) v_F(x, \tau_i)$, then construction costs are covered by setting μ to satisfy $\mu p(x, \tau_i) \Phi(x, i) v_F(x, \tau_i) = k(v_F(x, \tau_i))$. With iso-elastic cost function the amortization rate is then $\mu = 1/\gamma$.¹² Flow land-rent net of amortisation is therefore fraction $(1-1/\gamma)$ of gross revenue earned by land and structure together. This land-rent, net of amortization, is place, time, and date of development specific, and we denote it $r_F(x, t, \tau_i)$; it is given by

$$r_F(x, t, \tau_i) = (1 - 1/\gamma) p(x, t) v_F(x, \tau_i) = (1 - 1/\gamma) p(x, t) \left[\frac{p(x, \tau_i) \Phi(x, i)}{\kappa_F \gamma} \right]^{1/(\gamma-1)}. \quad (12)$$

Land-rents are therefore fraction $(1-1/\gamma)$ of space-rent in the formal sector, and fraction $(1-1/\alpha)$ in the informal sector (eqn. 5), relationships that we will use later in the paper. Notice also that, comparing (11) and (12) at date $t = \tau_i$,

$$r_F(x, \tau_i, \tau_i) = R_F(x, \tau_i) / \Phi(x, i). \quad (12a)$$

2.3 Land development and construction phases

Equations (5) and (11) give the maximised returns to respectively informal or formal sector use of land at place x and date t . We now look at the choice of where and when land is in one or other of these uses. We pose this as the decisions of a landholder of when to develop informal structures and when to develop or redevelop formal structures on land at place x . At some date (say time 0) the present value of land-rent at location x that has not yet been developed is

$$PV(x) = \int_0^{\tau_0} r_0 e^{-\rho t} dt + \int_{\tau_0}^{\tau_1} r_1(x, t) e^{-\rho t} dt + [R_F(x, \tau_1) - D(x)] e^{-\rho \tau_1} + \sum_{i=2}^{\infty} R_F(x, \tau_i) e^{-\rho \tau_i}. \quad (13)$$

¹¹ The iso-elastic form implies an elasticity of substitution between land and capital (i.e. construction cost) of unity. This is at the centre of the range suggested in by Ahlfeldt and McMillen (2014).

¹² Using $k(v_F) = \kappa_F v_F^\gamma$ the condition $\mu p v_F \Phi = k(v_F)$ becomes $\mu p \Phi = \kappa_F v_F^{\gamma-1}$, and using eqn. 11 to substitute for v_F gives $\mu = 1/\gamma$.

The first term is the present value of rent earned while the land is undeveloped land (flow rent r_0 which we take to be constant), discounted at rate ρ and accruing up to the date of first development, denoted τ_0 . The second term is the present value of rent while in informal urban use during interval $[\tau_0, \tau_1]$. The first formal sector development, occurring at date τ_1 , yields rent and incurs a potential one-time fixed cost $D(x)$ capturing frictions to formal development, such as unclear land rights or costs of land preparation; we discuss these further in Section 2.5. The final term in (13) gives the discounted value of land-rents earned over the lives of consecutive formal sector buildings, constructed at dates $\tau_2, \tau_3 \dots$. Equation (13) assumes this sequence of development, and we will establish conditions under which this sequence is followed. Our key results emerge from solving this dynamic optimisation problem and we proceed in stages, so as to explain results.

Dates of development and redevelopment are chosen to maximise $PV(x)$. For the first development the optimal τ_0 simply equates flow land-rents on undeveloped and informal land,

$$\frac{\partial PV(x)}{\partial \tau_0} = e^{-\rho t} [r_0 - r_I(x, \tau_0)] = 0, \quad (14)$$

which, with iso-elasticity and using (5) is,

$$r_0 = \kappa_I(\alpha - 1)[a_I p(x, \tau_0)/\kappa_I \alpha]^{\alpha/(\alpha-1)}. \quad (14a)$$

Since space-rent $p(x, t)$ is increasing through time, for each place x there is a unique date τ_0 at which informal development commences.

The first formal development takes place at date τ_1 satisfying (see Appendix Section A1.2)

$$\frac{\partial PV(x)}{\partial \tau_1} = e^{-\rho \tau_1} [r_I(x, \tau_1) - p(x, \tau_1)v_F(x, \tau_1) + \rho\{k(v_F(x, \tau_1)) + D(x)\}] = 0. \quad (15)$$

Using the expression for land-rent, (12), and formal volume, (10), together with iso-elastic construction technology this can be written as

$$r_F(x, \tau_1, \tau_1)(\gamma - \rho \Phi(x, i))/(\gamma - 1) - \rho D(x) = r_I(x, \tau_1). \quad (15a)$$

This gives a unique switching point from informal to formal if $\alpha > \gamma$, (from eqns. 5 and 12), which we assume and which is a condition that our estimates of α and γ clearly satisfy.

The first redevelopment of formal land is at date τ_2 satisfying (see Appendix Section A1.2)

$$\frac{\partial PV(x)}{\partial \tau_2} = e^{-\rho \tau_2} [p(x, \tau_2)v_F(x, \tau_1) - p(x, \tau_2)v_F(x, \tau_2) + \rho k(v_F(x, \tau_2))] = 0.$$

Generalising this for all redevelopments gives:

$$p(x, \tau_{i+1})[v_F(x, \tau_{i+1}) - v_F(x, \tau_i)] = \rho k(v_F(x, \tau_{i+1})), \quad \text{for } i \geq 1. \quad (16)$$

Again using (12), (10), and iso-elastic construction technology this can be written as

$$r_F(x, \tau_{i+1}, \tau_{i+1})(\gamma - \rho \phi(x, i + 1))/\gamma = r_F(x, \tau_{i+1}, \tau_i). \quad (16a)$$

Intuition on these redevelopment dates can be seen from inspection of (16).¹³ This says that demolition and reconstruction occur at the date at which the instantaneous revenue gain from the change in volume equals the interest cost of the construction expenditure incurred. Eqns. (14) – (16) implicitly define the dates at which sites are (re-)developed. These equations, together with the definition of the value-to-rent ratio, $\Phi(x, i)$ in eqn. (8), form the basis of the analysis of the next sub-section.

2.4 Analysis

What do we learn from the characterisation of development stages given above? Assuming space-rents grow at constant exponential rate $\hat{p} > 0$ yields analytical results, where this growth is driven by the growth in urban wages relative to the outside option, derived from eqn. (1) (see Appendix Section A1.1). We look first at the time series development of a particular place x , and then at the urban cross-section.

2.4.1 Urban dynamics at any location

The process of redevelopment of land that has been formalised is summarised as follows:

Proposition 1: If formal sector construction costs are iso-elastic in height (with elasticity γ), space-rents are growing at constant exponential rate \hat{p} , $\rho(\gamma - 1)/\gamma > \hat{p} > 0$, and agents have perfect foresight then:

- (i) The value-to-rent ratio takes constant value Φ and the time interval between successive formal redevelopments is constant $\Delta\tau$,

$$\Phi = \int_0^{\Delta\tau} e^{(\hat{p}-\rho)t} dt = \frac{1 - e^{(\hat{p}-\rho)\Delta\tau}}{\rho - \hat{p}}, \quad \Delta\tau = \frac{\gamma - 1}{\hat{p}} \ln \left[\frac{\gamma}{\gamma - \rho\Phi} \right]. \quad (17)$$

- (ii) Successive rounds of formal sector building have greater volume (height) by a constant proportional factor

$$\frac{v_F(x, \tau_{i+1})}{v_F(x, \tau_i)} = e^{\hat{p}\Delta\tau/(\gamma-1)} = \frac{\gamma}{\gamma - \rho\Phi} > 1. \quad (18)$$

- (iii) Land value (capital value of a unit of land) in the formal sector at location x newly redeveloped at time τ_i is given by

$$PV_F(x, \tau_i) = r_F(x, \tau_i, \tau_i)\Phi / \left[1 - e^{-\left(\rho - \frac{\hat{p}\gamma}{\gamma-1}\right)\Delta\tau} \right], \quad (19)$$

equivalent to constant growth rate per unit time of $\hat{p}\gamma/(\gamma - 1)$.

¹³ The switch dates in equations (15) and (16) are *not* when flow land-rents are equalised. This is because land-rents (net of amortization) jump as volume increases at date of construction. The terms in γ and $\rho\Phi$ make the appropriate adjustments. Notice that in eqn. (15) the term $(\gamma - \rho\Phi(x, i))/(\gamma - 1) = 1$ in a stationary model with $\hat{p} = 0$ and the first generation development lasting in perpetuity.

The first part of this proposition comes from integrating eqn. (8), using it in (16a), and noting that there is a unique solution solving the two parts of (17) with Φ and $\Delta\tau$ constant over time and space. The second part follows from (16a), noting that, from (12) the ratio of land-rents at date τ_{i+1} on vintages $i + 1$ and i is simply the ratio of volumes. Part (iii) gives the capital value of a unit of land in the formal sector at location x at newly redeveloped at time τ_i derived by using $R_F(x, \tau_i) = r_F(x, \tau_i, \tau_i)\Phi$ in (12), and summing over infinitely repeated cycles of redevelopment; the sum is finite if $\rho > \hat{p}\gamma/(\gamma - 1)$. Flow land-rent $r_F(x, \tau_i, \tau_i)$ increases at rate \hat{p} during the life of a building, and jumps up by factor $e^{\hat{p}\Delta\tau/(\gamma-1)} = \gamma/(\gamma - \rho\Phi) > 1$ at date of redevelopment, due to the greater building volume. The average growth of the land-value is greater than that of space-rent in anticipation of these jumps in volume. Relationship (19) between current land-rent and land-value (reflecting future rent increases) will be used in Section 3 to back out an estimate of \hat{p} , the rate of increase in housing rents. Note also that the proposition tells us that, if the rate of growth of space-rent is the same in all formalised locations, then so too are Φ , $\Delta\tau$, and the growth rates of building volume and land-value.

Proposition 1 deals with redevelopment of land that has previously been formalised. The timing of informal and first formal development are given by eqns. (14) and (15), and hinge on comparisons between r_0 , $r_I(x, t)$, and $r_F(x, t, t)$. With Φ now constant from proposition 1 we have:

Proposition 2: If formal sector construction costs are iso-elastic in height (with elasticity γ), informal amenity is iso-elastic in volume (with elasticity α), space-rents are growing at constant exponential rate \hat{p} , $\rho(\gamma - 1)/\gamma > \hat{p} > 0$, agents have perfect foresight and $D(x) \geq 0$ then:

- (i) Informal development exists for an interval of time $[\tau_0, \tau_1]$, positive if and only if τ^* solving $r_0 = r_F(x, \tau^*, \tau^*) (\gamma - \rho\Phi) / (\gamma - 1) - \rho D(x)$ is greater than τ_0 .
- (ii) If informal development occurs, then it precedes formal development.

The equation in the first part of the proposition implicitly defines τ^* as the date at which it would be profitable to switch undeveloped rural land directly to formal sector development. If $\tau^* > \tau_0$, then the switch date from rural to informal (eqn. 14) occurs prior to this date, so there exists a period of informal land-use prior to formalisation. Appendix Section A1.3 derives the inequality under which this holds, $((r_0 + \rho D)/(\gamma - \rho\Phi))^{(\gamma-1)/\gamma} (\kappa_F/\Phi)^{1/\gamma} \gamma > (r_0/(\alpha - 1))^{(\alpha-1)/\alpha} \kappa_I^{1/\alpha} \alpha/a_I$. As expected, an interval of informality is more likely the lower is κ_I , higher is a_I , the higher is κ_F/Φ , and the lower is r_0 .¹⁴ The second part of Proposition 2 is proved by noting that r_I is increasing at rate $\hat{p}\alpha/(\alpha - 1)$ (eqn. 6) and r_F is increasing at rate $\hat{p}\gamma/(\gamma - 1)$ (eqn. 12). Given $\alpha > \gamma$, there is therefore a unique crossing, switching from informal to formal. The assumption $\alpha > \gamma$ means that there are sharper diseconomies to informal sector crowding than to formal building height, and that the share of land-rent in revenue is higher in informal development than in formal.

¹⁴ r_0 determines the date of edge slum development, but has no bearing on subsequent time intervals between this and formalisation or subsequent redevelopments.

Fig. 1 illustrates these results in a stylized benchmark city without frictions, using model parameters estimated in Section 3.¹⁵ Building volume is given on the vertical axis (log units), and on the horizontal plane axes are time t and location x . Location is distance from the centre, and we discuss the cross-section (variation across x at a given t) in the next sub-section. For the moment, look just at the development of a particular location through time, i.e. fix x and look along a line sloping up and to the right parallel to the t axis. Initially (at low t) this land is rural. Building volume becomes positive at date τ_0 (specific to location x) when informal development takes place. The volume of informal development increases steadily, as increasing p causes Meccano pieces to be rearranged and building cover to increase. Formal development takes place at τ_1 and, as illustrated, leads to a small increase in volume, indicated by the second step. Subsequent redevelopments occur at fixed time intervals $\Delta\tau$ and bring the same proportionate increase in volume, achieved by building taller. The timing and volume of each of these formal investments is based on perfect foresight about the growth of prices and the date of subsequent redevelopments.

2.4.2 The urban cross-section and its evolution

We have so far concentrated on a single location, x , and now show how development varies across places in the city. We interpret x as distance from the city centre and assume the existence of commuting costs such that wages net of these costs, $w(x, t)$, decrease with distance from the centre at exponential rate θ . In Appendix Section A1.1 we give the underlying assumptions that yield this exponential form. Exponential decline with distance and exponential growth through time imply that space-rents are

$$p(x, t) = \bar{p}e^{\hat{p}t}e^{-\theta x}. \quad (20)$$

The switch points in eqns. (14) - (16) can now be interpreted either as giving the date at which place x develops, or the place that develops at date t . The latter interpretation gives the urban cross-section. Proposition 3 states how different stages of development vary across this stylized city as it grows, and Proposition 4 summarises how this cross-section evolves through time.

Proposition 3: If formal sector construction costs and informal sector quality are iso-elastic, space-rents are growing at constant exponential rates \hat{p} , $\rho(\gamma - 1)/\gamma > \hat{p} > 0$ and declining with distance at constant rate $\theta > 0$, and agents have perfect foresight then:

- (i) The distance from the city centre to the edge of new informal development increases through time according to $dx_0/dt = \hat{p}/\theta$.
- (ii) If $D = 0$, the distance from the city centre to the outer edge of formal development increases through time according to $dx_1/dt = \hat{p}/\theta$.
- (iii) The distance between successive formal sector redevelopments, Δx , is constant,

$$\Delta x = \frac{\gamma - 1}{\theta} \ln \left[\frac{\gamma}{\gamma - \rho\Phi} \right].$$

See Appendix Section A1.5 for proof.

¹⁵ See Appendix Section A1.4 for listing of parameters.

These results are illustrated in Fig. 1, where we now fix a date and move along a line parallel to the x -axis. At the city edge land is informal. Moving towards the centre, at the inner edge of informal land there is the most recent formalisation; and, as we move in, further places that have been urban for longer have been through more stages of development, with a jump in housing volume and height at each stage. As the city grows, part (i) of the proposition says new informal sector development is pushing continuously into rural land on the fringe. Similarly, part (ii) says the ring of most recent formal sector development is pushing continuously into the inner edge of the informal sector ring and that the width of the ring of informal area, $x_1 - x_0$ is constant through time. Hence, even in a circular city, the share of urban land area that is informal declines as the city gets larger. Part (iii) says that the width of formal sector rings of development is constant. We also note that as we move from the city edge to the inner boundary of slums land, housing quality declines continuously as more investment is crowded onto each piece of land. Thus, iso-elasticity and exponential growth yield simple patterns in a benchmark city without frictions.

2.5 Spatial heterogeneity and formalisation costs

Cities do not look like the neat pattern described above. Places have idiosyncratic features, and our focus here is on those that affect the ability to formalise land. These include institutional and geographic costs of moving from informal to formal settlement, all of which are summarised in our modelling by the parameter $D(x)$, the fixed cost of formalisation at place x .¹⁶

The implications of this spatial heterogeneity are illustrated Fig. 2 in which $D(x)$ varies with distance from the centre, taking random non-negative values (details in Appendix A1.4). Instead of the sharp edges of Fig. 1, stages of development are now locally diffused because of the variation in $D(x)$. This yields ‘waves of development’ in different grey scale shades in the figure. The leading edge of the formalisation wave is places with low D and, within a particular distance band, this is followed a process of infill, with the fraction of the area covered by formal buildings increasing through time. The city cross-section therefore exhibits areas of informality next to formal structures, some of which may have gone through several phases of redevelopment. In the informal areas volume and crowding increase. This is the pattern that we see in the data, as will be discussed in Section 5.

We know from Proposition 1 that the time interval between each redevelopment at a particular place is fixed, so it follows that if first formal development is later and taller, then so too are subsequent redevelopments.

Proposition 4: If formalisation costs $D(x)$ vary across places, space-rents are growing at constant exponential rates \hat{p} , $\rho(\gamma - 1)/\gamma > \hat{p} > 0$ and declining with distance at constant rate $\theta > 0$, formal sector construction costs are iso-elastic and agents have perfect foresight then:

¹⁶ While we have $D(x)$ as a conversion cost that occurs with formalization whether converting land from slums, farmland or vacant land, there may be an additional cost of clearing slums per se, including eviction costs even from private land. For properties where such costs may be high due to history and specific rights, landholders may skip the slum stage and go from farmland to formalisation directly. This element is not added to the model here to limit complexity but is noted again in Section 5 in interpreting data.

- (i) Formal development is delayed for places with high $D(x)$, but proposition 1 continues to hold for each x .
- (ii) Variation in $D(x)$ means that first formal development and subsequent redevelopments may be undertaken in different places at the same date. For such developments, volume (height) and land-value gradients are

$$\frac{dv_F(x, \tau_i)}{dx} \frac{1}{v_F} = \frac{-\theta}{\gamma - 1}, \quad \frac{dR_F(x, \tau_i)}{dx} \frac{1}{R_F} = \frac{-\theta\gamma}{\gamma - 1}. \quad (21)$$

The spatial gradients in the last part of this proposition are derivatives of eqns. (10) and (11) and will be used in the empirical section of the paper to back-out parameters of the model.

Spatial heterogeneity of formalisation costs $D(x)$ supports the observed hotchpotch of adjacent formal and slum buildings. Some of the costs are incurred in securing private property rights (for either leasehold or freehold land) needed to avoid risk of expropriation, to obtain financing and insurance, and to clarify issues such as inheritance, compensation for takings and the like. Other costs arise from the institutional blockages in the long and tortured development of private property rights in Nairobi, as discussed in the Introduction. Such frictions involving corruption and a dubious legislative history transferring land to the government have been a critical factor in the city's development. These frictions vary widely across areas of the city, as we discuss in Section 6.

Spatial heterogeneity in $D(x)$ may be due to geographical as well as institutional factors. Initial construction costs may vary with terrain, for example the costs of draining a swamp or levelling a rugged site for development. We offer empirical evidence for the presence of such effects as discussed in Section 5.1 and Appendix 3. We show that poorer geography delays initial development. High price land near that city centre undergoing first development has distinctly poorer terrain compared to the city edge, as all better terrain land near the city centre has already been developed. Evidence also suggests that poor terrain affects initial development costs, rather than costs of redevelopment, since it does not reduce building heights on redeveloped formal sector land.¹⁷

3. The built environment of Nairobi:

Our empirical work provides evidence on the built fabric of Nairobi and its evolution through time. We use the data and analytical structure to estimate equations characterising the 2015 cross-section, and these are used in Section 4 to solve for model parameters. In Section 5 we use the model and parameters to make predictions about the dynamics of the city and then in Section 6 to do welfare analysis.

3.1 Data and mapping

As noted in the Introduction and detailed in the Data Appendix, we capture the characteristics of the built environment of Nairobi at a very fine spatial resolution. We work with 150m x 150m grid cells,

¹⁷ Specifically, in Appendix Table A3.2 rebuilding heights are uncorrelated with elevation, presence of water in the neighbourhood, and do not decrease with greater ruggedness.

based on data aggregated from 40 cm resolution to 3m x 3m cells and to the grid squares that we work with. From aerial photographs, we have tracings of all building footprints in 2003 and 2015 and we have LiDAR height data for 2015. LiDAR records on average 1-3 points per square meter on the ground and is accurate in height to within a meter or less. Each building's height is the average of the LiDAR recordings that intersect with the building's 2-D footprint and allows for sloping roofs and differential height in building segments. To infer 2003 heights, we assume that if building footprint is unchanged between dates then so is height. For demolished buildings, we assume 2003 heights equal the average height of unchanged buildings in the eight queen neighbouring grid squares, although that is likely to overstate their height, since demolished buildings are likely less tall than unchanged ones.

In Fig. 3 we show two mappings of slums and define the area of the city we work with. The yellow star in Fig. 3 marks the city centre, from which all gradients we estimate emanate. This centre is the traditional central business district and is defined by the brightest nightlights pixel in Nairobi in 1992. Fig. 3 shows our economic boundaries for the city, a dashed outline for 2003 and a solid one for 2015. These boundaries are based on built cover. For a (150m x 150m) grid square to be in the city, the average roof cover in cells whose centroid is within a 900 meter radius of the cell must be above 10%; and we only keep those cells that contiguously connect to the centre. To focus on private sector development, we remove all grid squares entirely in permanent public uses and amounting to 11% of land in the 2003 city boundary; neighbourhood schools and roads remain (see Appendix 2). There is a large extensive margin expansion between 2003 and 2015 of about 60% in land area. Nairobi's sausage shape and the lack of expansion directly north and south of the centre arises because of permanent fixtures (Harari, 2020): in the south an airport and a national park and in the north a state forest.

Our analysis requires a distinction between formal and informal settlements (slums). We base our slum mapping on a 2011 classification by IPE (2011) under the Kenya Informal Settlements Improvement Program, with details given in Appendix 2. The IPE mapping used satellite imagery, topographic maps, and on the ground surveying to define slums as having aspects of lack of planning and lay-out, low house quality, poor infrastructure, or insecure tenure. They also classify slums by land ownership, in particular private versus government, and some residual categories such as riparian and roadside slums, where riparian also includes large tracts of privately owned land and land under possessory rights. Fig. A2.7 in Appendix 2 shows how ownership and slum classification varies with distance from the centre. An earlier delineation of slums comes from a 2003 land use map prepared by the CSUD at Columbia University. While the effective definitions differ in precise detail across the two studies, we use the 2003 mapping to try to distinguish slum areas that underwent formal sector development between 2003 and 2011. While we rely on the 2011 IPE mapping to define slum areas in empirical work, the categorisation of formal versus slum buildings and the drawing of slum borders can have fuzzy portions. While in Section 2.2.2, there was a clear general distinction in building materials between the sectors, there can be more of a continuum where, for example, some slum buildings are made of load bearing materials. They are just in crowded areas that are irregularly laid out.¹⁸ Any categorisation has issues, and we experiment in robustness checks with other divisions.

¹⁸ There are writings about semi-formal buildings, which are built tall of cheap materials and have a reputation in the press of collapse. This sector does not seem to be growing given the bad publicity, although semi-formal

In Fig. 3, slum areas are marked red if recorded in both studies, light red if only in IPE (2011) and very dark red if only in CSUD (2003). We will look at these very dark areas to show aspects of slum conversion to formal sector use. However, there seems to be little overall slum removal. Areas near the centre with no recorded slums are marked by the dashed and solid rings for respective periods. While the area with no slums expands considerably from about a 0.8 km radius to about a 2 km radius around the centre by 2011, the removed slum areas are tiny. The map suggests that most slum extensive margin expansion (light red) is near the 2003 fringe of the city and beyond. The well-known large slum of Kibera, which we will discuss later in some detail, is the large slum area directly south-west of the centre (ranging from about 3-5 km of the centre).

Data on space-rent and housing characteristics are derived from a georeferenced household data set from the 2012 ‘Kenya: State of the Cities’ survey by the National Opinion Research Center (NORC) and will use these to infer base formal and slum rents. The NORC used the Census mapping of slums noted above but sampled within those areas to get what they thought were true slum versus formal sector properties; we accept their on-the-ground classification for the sample properties. Second, on prices, we have asking prices for vacant land listed in late 2015, obtained by scraping from property24.co.ke, a website that advertises property for sale in Kenya. All listings fall in the formal sector. In Fig. A2.4 of the Appendix, we show a map of the locations of these properties (noting some dots overlap) and the NORC sites. Nairobi has a relatively sophisticated property market. We consider the listing services to give excellent coverage of formal sector land and there is a good scattering of all types of properties in the map.

We have made choices above as to what is the city centre, how slums are defined, and whether there is an important alternative centre. In Section 3.2.3, we perform robustness checks on our results, concerning these choices. Second, we rely on specific sources for land value and rental data. As such there is the issue of selection in the listings for land price data and in NORC surveying of slums, and the potential for omitted variables in regression analysis. We will discuss potential biases below. Third concerns what land value data reflect in markets where there can be bubbles. Appendix Section A2.4 analyses this issue for Nairobi, concluding that bubbles are not an issue in the data we use.

3.2 Characteristics of Nairobi’s built environment in the cross-section

Our cross-section description of the city starts with how the built fabric varies with distance from the centre, from which we derive parameters of the model and evaluate key model assumptions. As well as fitting our model, looking at variation with respect to distance from the centre is standard in the empirical literature (e.g. Combes, Duranton, and Gobillon, 2018; Ahlfeldt and McMillen, 2018). Such regressions take the form

$$\log(y_i) = b_1 + \beta_1 x_i + \beta_c Z_i + \epsilon_i.$$

construction had a modest boom and the well-publicised collapses. There is no mapping of semi-formal areas, but based inspection of satellite images and on Huchzermeyer (2007) we think slum boundaries largely contain these structures, where the Huruma area Huchzermeyer studied is within the IPE slum mapping we use.

Outcomes, y_i , include land prices and measures of the built fabric in formal and slum areas. Where the dependent variable is space-rent or land price observation i denotes the household or vacant lot; where it is feature of the built fabric observation i denotes the 150m x 150m grid square. x_i is distance to the city centre so that β_1 is the gradient. Following the model, but also the literature, regressions are semi-log, so a gradient coefficient can be interpreted as proportionate change per km distance from the centre. b_1 is the value at the city centre. Z_i are outcome specific relevant control variables discussed below. For the error term, omitted variables will be a concern.

Results on estimated relationships are presented in figures and in tables. The figures for this section cover the 2003 bounded area of the city, cut at 10km from the centre, based on observations at the 150m x 150m grid square level. Regressions cover all 6470 grid squares of the 2003 city. For this section, regression results are presented in Tables 1 and 2 where all regressions contain controls on grid square elevation and ruggedness. Regressions for prices have further controls, and tables report the gradients and an “intercept” which is $b_1 + \beta_c \bar{Z}_i$ for Z_i evaluated usually at median values for the city. More details are in table notes and coefficients on all covariates of key regressions are in Appendix Table A2.5.

3.2.1 Formal Sector.

Intensity of land use is measured by volume of built space in cubic meters (m^3) per square meter (m^2) of land area, which we call the built volume to area ratio, BVAR and correspond to $v_i(x, t)$ in the theory. The gradient for BVAR is given in col. 1 of Table 1 and illustrated for the raw data with a smoothed fit in Fig. 4, which shows the confidence intervals and the 25th and 75th percentiles of observations.¹⁹ Formal sector BVAR near the centre is very high, averaging around 8 m^3 of space per m^2 of ground area (where ground area includes roads and minor public spaces) in Fig. 4; it declines at a rate of 4.9% per km from the centre from Table 1, col. 1. The 25th and 75th percentile of observations. Note there is heterogeneity across grid squares in BVAR’s at each distance, reflecting, in part, differential timing of formal sector development of grid squares, based on differential geography and the specific histories of properties’ paths to formalization, as well features such as roads.

BVAR is the product of height and the cover-to-area (CAR, or $c_i(x, t)$ in the model), which have following gradients. In Table 1, col. 3, average heights of buildings decline with distance from the centre at a rate of 7.6% per km. This compares with height gradient numbers for Chicago in the 1960’s or 70’s from Ahlfeldt and McMillen (2018).²⁰ As Fig. 5 shows, there are very tall buildings at the centre in Nairobi; at 0-1 km from the centre average grid square height (based on 3m x 3m pixel) is about 10 storeys (at 3m a storey); the data indicate that 6% of the grid squares are over 16 storeys. For CAR, we assumed in the theory that the formal sector has constant cover to area ratio throughout the city. The gradient in Table 1, col. 2 has a zero slope; and the solid blue line in Fig. 5 indicate that CAR does not

¹⁹ While we have the universe of observations, smoothing requires estimation to show a continuous gradient. As such, the displayed confidence intervals on the mean are extremely tight. To show the dispersion, we plot smoothed estimates of the 25th and 75th percentile of observations at each distance. Smoothing involves grid squares whose centroid is in a 300m moving window going out from the centre.

²⁰ The authors use different functional forms and look just at tall buildings. But at say 5km from the centre their slope estimates are similar. However, for later years they have steeper slopes.

change much across the city and is not decreasing with distance from the centre. The decline in formal sector BVAR is thus due to height rather than cover.

To recover formal sector parameters in line with the theory we focus on the gradient for heights of newly redeveloped buildings, as given in eqn. (21).²¹ We don't know building ages per se, just whether buildings are new since 2003. To pick the most recently redeveloped buildings from the 2003-2015 interval, we first take the sample of redeveloped buildings from 2003-2015. Based on the theory where at any distance heights of new buildings will rise with time, we focus on the 80th percentile, balancing out wanting the most recent and hence taller buildings, against getting extreme outliers. Col. 4 estimates the gradient for redeveloped buildings using a quantile regression. Fortunately, for redeveloped buildings, different quantile and OLS estimates of the gradient slope are very similar.²² From Table 1 col. 4 the slope is -10.1% per km, noticeably larger in absolute value than for the overall stock.

What are identification issues for these regressions on the built environment in cols. 1-4? The data are for the universe within the 2003 city, so there is no sample selection. The concern would be that there are geographic conditions (affecting the D 's in the model) which could be correlated with distance to the centre and affect gradients, which is why we control for elevation and ruggedness. That said, the two key equations we use to parameterise the model, the slum BVAR and the formal sector 80th percentile of height quantile regression, should not be affected by D magnitudes, according to the model. Correspondingly, whether we control for ruggedness and elevation or not, the estimated coefficients are stable.

Turning to price and rent gradients, col. 6 of Table 1 uses scraped land price listings from the fall 2015 property24.co.ke data to estimate the formal sector land price gradient with respect to distance from the centre, with asking prices (per square meter) in USA\$. An exponential form captures the relationship well, and the gradient is steep, with price declining by 17.2% per km of distance from the centre. Such a rate of decline means land prices vary almost six-fold from the centre out to 10 km. This slope is right in the middle of the range of land price gradient slopes across French cities in Combes et al. (2018, Table 3).

For col. 6 of Table 1, identification issues are a concern. There are two issues, sample selection, and bias in estimates within the selected sample due to unobserved property features affecting prices that could be correlated with distance to the centre. On selection, while we believe most lots in the formal sector are listed, we do not know the universe of properties for sale. Fig. A2.4 in the Appendix shows a good scattering of listing points throughout the city. If there are unlisted properties, they may be different, on more challenged land or land with worse amenities, both of which could vary with distance from the centre. For example, if these omitted properties are first formal sector developments, we will show later that these would be on worse quality sites nearer the centre (last to be developed). In this case selection could bias our estimate of the intercept upwards and our estimate of the slope downwards (more negative). On the issue of omitted variables, we control for plot area, ruggedness, elevation, and whether the listings have an exact street address (which many properties do not) as an indicator of a

²¹ Note we cannot do redeveloped BVAR, since we cannot properly assign land area to new vs. old buildings.

²² Gradient coefficients on redeveloped buildings are stable, with OLS and quantile regressions up to the 90th percentile all having gradients in the neighbourhood of -0.10. But intercepts rise, as we raise the percentile level.

quality differential. However unobserved amenities could be better or worse near the centre, and the direction of bias is ambiguous. In short, depending on the relative extent of selection and omitted variable biases, our estimate of the land rent gradient could be biased either way.²³

Table 1: Formal sector							
	Ln formal BVAR	Ln formal CAR	Ln formal height	Ln formal redeveloped height; quantile: 80 th percentile	Ln formal unchanged height; quantile: 20 th percentile	Ln land price: m ² , \$2015.	Ln formal space-rent: m ³ , \$2015.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance to centre	-0.0493 (0.0124)	0.0239 (0.00888)	-0.0763 (0.00791)	-0.101 (0.00521)	-0.0763 (0.00223)	-0.172 (0.0476)	-0.0855 (0.0310)
Intercept ($x = 0$) for typical item	0.747 (0.0829)	-1.672 (0.0571)	2.487 (0.0537)	3.315 (0.0356)	2.131 (0.0150)	7.254 (0.277)	3.148 (0.071)
Ruggedness & elevation	yes	yes	yes	yes	yes	yes	yes
Other Controls	no	no	no	no	no	yes	yes
Observations	5435	5435	5435	4621	5079	136	361
R-squared	0.035	0.043	0.203			0.292	0.244
<p>Note: All columns except 7 are based on 2015 data for observations inside the 2003 extent of the city. Col. 7 is based on NORC data and restricted to observations inside the 2003 extent of the city. Standard errors in parentheses. Standard errors are robust and in columns 1-3 are clustered based on a 750m x 750m grid. Errors in cols. 4 and 5 are not clustered; in col. 6 they are clustered at the neighbourhood area from the on-line listing service; and in col. 7 at the census enumeration area. Details and full results for columns 4, 6 and 7 are in Table A2.5 of the Appendix. Reported intercepts are predicted values at the city centre for median ruggedness and elevation for cols. 1-5. In those columns, for height regressions we only include grids for which there is cover. For BVAR and CAR regressions about 5% of observations have no cover and hence volume (e.g. playing fields, overpasses, small parks and the like). A Tobit including these as 'censored' at 0 yields almost identical slope coefficients. In col. 6, the reported intercept is the predicted sales price at the centre per m² for a lot based on median area, ruggedness and elevation for the sample. In col. 7 the intercept is the rent per cubic meter of space for a house with typical characteristics in the formal sector: mean values for categorical variables and median for continuous for the NORC formal sector, for the regressions where we force the gradient to have the model slope of -0.071 solved in the next section.</p>							

The final column in Table 1 estimates the space-rent gradient with data from the NORC, using hedonic price regressions for space-rents. Col. 7 reports gradients and intercepts in 2015 USA\$ for m³ of

²³ Using reports from HassConsult, an alternative set of vacant land listing prices, the estimated gradient from regressing ln price per square meter on distance gives a similar slope -0.198 (0.0309), within one standard error of that estimated here. See Appendix Section A2.4 for details.

volume.²⁴ We will utilise the intercept only in work below; this space-rent gradient can be sensitive to specification unlike other columns. These regressions control for house characteristics in order to define rents per unit volume for a *typical* formal sector house; full hedonic regressions are reported in Appendix Table A2.5. Table 1, col. 7 gives the gradient of formal sector space-rents, which is flatter than that of land-rents as suggested by the theory. In col. 7, its slope of -8.6% per km is close to what we back out of the model (-7.1%) in Section 4.1 below.

3.2.2 *Slum areas*

Table 2 gives gradients for the slum sector. Slum areas should exhibit a decline in volume with distance from the centre, with the decline driven by diminishing cover per unit area, not by a change in height. Table 2 and Figs. 4 and 5 give results. Column 1 of Table 2 indicates BVAR declines significantly with distance from the city centre at a rate of 9.5% per km. In col. 3, slum heights are flat in terms of distance from the city centre as assumed in the model and as pictured in Fig. 5 by the red dotted line. With constant slum height, gradients for slum BVAR in col. 1 and slum cover to area ratio, CAR in col. 2 should be the same; the estimates of -0.0948 and -0.103 in Table 2 cols. 1 and 2 confirm this.²⁵ Slum CAR near the city centre in Fig. 5 is very high at 50% or so, more than twice the 25% number for the formal sector. This means that slums have little green/open space around houses, with attendant loss of amenity.²⁶

The final column of Table 2 gives the slum space-rent gradient. Note that in the model and in the data, typical formal and slum sector houses are not comparable; and we have separate gradients for each. The slum space-rent gradient is flat, in contrast to that in the formal sector. In eqn. 6 of the model, observed space-rents are constant throughout the city; the insignificant coefficient of -0.0094 validates the modelling that gives this result. The intercepts of the space-rent equations in the formal and slum sectors in Tables 1 and 2 are respectively 3.148 and 1.886. These will be used later in welfare analysis to infer land-rent differentials by sector. The formal-slum differential in space-rents reflects different amenity values, incorporated in the model by the parameters a_i .

Again, there are issues of identification for this NORC sample. On selection, Fig. A2.4 in the Appendix shows a good scattering of points throughout the city for formal and slum housing, covering newer and older slums within the 2003 city. Households were randomly sampled and stratified based on the 2009 Census definition of slums. On bias and unobserved attributes that could vary systematically with distance from the centre for developed formal sector and slum housing, we simply do not know the likely direction of bias. We also note that features like ruggedness and elevation on average are the same across sectors with insignificant differences.²⁷

²⁴ Calculations assume each m² of floor space yields 3.0 m³ of volume (given typical ceiling height) inflates 2012 nominal rents at 8% a year based on reports from <http://www.hassconsult.co.ke>, converts monthly to annual rents and Kenyan shillings at the 2015 exchange rate of 100 KS to a USA\$.

²⁵ Note the slopes for slum CAR and height (-0.103+0.0107) sum to almost that for BVAR, -0.0948.

²⁶ We also have paved road surface from high-resolution SPOT satellite data for 2015. Formal sector roads are about 15% of area near the centre while in slums they maybe hit 5%.

²⁷ Mean [s.d.] log elevation in the slum and formal sector are 7.43 [0.045] and 7.40 [0.028] and log ruggedness are 1.58 [0.54] and 1.43 [0.57] respectively.

Table 2: Slum areas				
	Ln slum BVAR	Ln slum CAR	Ln slum height	Ln slum space-rent m ³ , \$2015.
	(1)	(2)	(3)	(4)
Distance to centre	-0.0948 (0.0172)	-0.103 (0.0171)	0.0107 (0.00871)	-0.0094 (0.0243)
Intercept ($x = 0$) for typical item	1.275 (0.1176)	-0.116 (0.1049)	1.375 (0.0686)	1.886 (0.0581)
Ruggedness & elevation	yes	Yes	yes	yes
Other controls	no	No	no	yes
Observations	958	958	958	439
R-squared	0.104	0.142	0.017	0.391
<p>Note: All columns are based on 2015 data for observations inside the 2003 extent of the city. Robust standard errors are in parentheses. Errors in columns 1-3 are clustered based on a 750m x 750m grid and in col. 4 are clustered at the enumeration area. Details and full results are in Table A2.5 of the Appendix. Reported intercepts are predicted values at the city centre for median ruggedness and elevation for cols. 1-3. In col. 4 the intercept is the rent per m³ of space for a house with typical characteristics in the slum sector: mean values for categorical variables and median for continuous in the NORC slum sector for a regression where we force the gradient to have the model slope of 0.</p>				

A further point of interest in Fig. 4 is that, at 2km and beyond, slums and the formal sector deliver generally the same BVAR. For the opposing views of whether formal sector height trumps slum coverage in providing volume of built space per unit land, in Nairobi, they do equally well on average, albeit at different quality levels. However, while each sector has similar BVAR, slums occupy a smaller fraction of the city's surface area and hence provide just 10% of total building volume.

3.2.3 Robustness

We perform robustness checks on various definitions or classifications we imposed on the data. The first concerns the city centre being chosen as the brightest nightlights' pixel in Nairobi in 1992. Alternatively, using the current nightlights from NASA's Visible Infrared Imaging Radiometer Suite to identify the top 1% of grid-cells readings in Nairobi, we can define the centre as the brightest pixel in the largest cluster of these top 1% pixels (Appendix Section A2.4). That moves the star in Fig. 3 by about 400 meters. The second check concerns the presence of a second major centre in Nairobi around which gradients might orient. The most likely candidate is a large industrial area to the east and somewhat south of the centre. We experiment with specifying this as a second centre influencing or even potentially dominating the gradient from the city centre that we estimate. Finally, there is the definition of slums. Alternative to IPE 2011 is a 2009 Census map of slums classifying census tracts as

slums if they are “unplanned”; this results in a looser definition of slums and larger tracts, as can be seen in Fig. A2.3 in the Appendix. We will experiment with using this definition of slums.

We report robustness results for the five equations that we use below to derive parameters of the model: the land, slum rent and formal rent price equations and the equations for slum BVAR and formal sector redeveloped height. For the slum BVAR and formal sector height equations, these involve adding a control for distance to the main industrial centre to the base equations, using the alternative centre point based on recent (not 1992) nightlights in the equations, and using the 2009 Census definition of slum areas (as unplanned settlements and having density over 3000 per sq. km). For the price gradients, we examine the impact of having a differently placed city centre and having a second centre. We do not change slum area definitions for the price equations. Land listings cover only the formal sector; and we accept that, in their sampling, NORC have correctly defined on the ground allocations of households to slum and formal sector environments. The results are reported in Table A2.6 of the Appendix. The new city centre definition has a negligible impact on all estimated gradients and intercepts for the five equations. Adding distance to the industrial centre to the base specifications adds a covariate that is always insignificant, has a small impact on all intercepts, and only has a noticeable effect on the CDB distance gradient in one case, formal sector rents. However, we do not use the gradient for that outcome to solve the model parameters below, only the intercept which is minimally affected. Finally using the census definition of slums similarly has a negligible impact on gradients and intercepts for slum BVAR and redeveloped heights.

4. The data and the model

We now derive parameters of the model, using 2015 cross-section information from Tables 1 and 2 and relationships in the model. The parameter estimates are of intrinsic interest, and we compare them with other estimates in the literature, where available. In Section 5, we use the parameterised model to predict dynamic changes between 2003 and 2015 and compare these with what we find in the data.

4.1 Solving for the cross-section built environment

The model, in cross-section, is fully determined by parameters describing technology and amenity, γ , α , κ_I , a_I , and (κ_F / Φ) , together with knowledge of space-rent at the centre and its gradient, $p(0,0)$ and θ . (Formal sector choices depend on κ_F / Φ ; here we calculate the ratio, and in Section 4.3 separate the components). Relevant values and model equations are set out in Table 3 which divides into two parts, the upper block giving gradients of variables with respect to distance from the centre and the lower the levels of parameters and variables at the centre, $x = 0$ (i.e. intercepts reported in Tables 1 and 2).

Looking first at Table 3 panel a; for the formal sector we use gradients on newly developed buildings, taking the distance coefficient of new building height (Table 1 col. 4) as the measure of how volume varies with distance, and similarly, land values as the measure of $R_F(x, \tau_i)$, the present value of land-rent that accrues over the life of a structure. For slums, the volume measure is BVAR; there is no land value measure available since slum land is not widely transacted. The three equations in this part of the table give values for parameters γ , α , θ .

Panel b of Table 3 equates intercepts from the regressions with expressions for the level of variables from the theory. These parameters play a role in model simulations where we need levels. For formal sector volume of new buildings, we use height times the median cover of 0.247 throughout the city. This is in preference to the intercept in col. 2 of Table 1 (which gives a CAR of 0.188) but has low explanatory power given the undulating but basically flat relationship for CAR in Fig. 6. For slums with their malleable capital all that matters is BVAR. The final row of the table uses the space-rent data and together with values of γ , and α , they give values of κ_I , a_I , (κ_F/Φ) and $p(0, 2015)$.

Table 3: Model parameters		
a. Model parameters from gradients		
Gradients:	Formal:	Slum:
Volume m ³ , per m ² land.	$\frac{dv_F(x, \tau_i)}{dx} \frac{1}{v_F} = \frac{-\theta}{\gamma - 1} = -0.101$ Eqn (21); Table 1 col. 4	$\frac{dv_I(x, t)}{dx} \frac{1}{v_I} = \frac{-\theta\alpha}{\alpha - 1} = -0.0948$ Eqn (4); Table 2 col. 1
Land-price per m ² land.	$\frac{dR_F(x, \tau_i)}{dx} \frac{1}{R_F} = \frac{-\theta\gamma}{\gamma - 1} = -0.172$ Eqn (21); Table 1 col. 6	
Solutions: $\gamma = 1.703$, $\alpha = 3.983$, $\theta = 0.071$		
b. Levels of variables and further parameters		
Volume m ³ , per m ² land.	$v_F(x, \tau_i) = \left[\frac{p(x, \tau_i)}{\gamma} \cdot \frac{\Phi}{\kappa_F} \right]^{1/(\gamma-1)} = 6.80$ $= \exp(3.315) \times 0.247$ Eqn (10a); Table 1 col. 4 for height, times median formal sector CAR = 0.247	$v_I(x, t) = [a_I p(x, t) / \kappa_I \alpha]^{\alpha/(\alpha-1)} = 3.58$ $= \exp(1.275)$ Eqn (4); Table 2 col. 1
Space-rent per m ³ volume,	$p(x, t) = 23.29 = \exp(3.148)$ Table 1 col. 7	$p(x, t) a(v_I(x, t)) = \kappa_I \alpha = 6.59 =$ $\exp(1.886)$ Eqn (6); Table 2 col. 4
Solutions: $p(x, t) = 23.29$, $a_I = 0.734$, $\kappa_I = 1.655$, $\kappa_F/\Phi = 3.554$. Levels evaluated at $x = 0$, $t = \tau_i = 2015$, in \$US		

4.2 Discussion

Technology and amenity parameters, γ, α , come from cross-section gradients and give diminishing returns in formal sector construction costs and slum crowding respectively. They also imply shares of land-rent and construction cost in spending on housing. The value of $\gamma = 1.70$ implies that the share of land-rent in formal sector revenue, $1 - 1/\gamma$, is 0.41, while $\alpha = 3.98$ implies a corresponding share in the informal sector, $1 - 1/\alpha$, of 0.75. The formal sector share of 0.41 is similar to that in Case (2007) for the USA but higher than the 0.30 for Paris in Combes et al. (2017). For slums, we know of no data

to make comparisons, since typically slum lands are not transacted privately. However, given the low construction costs of slum housing, such a high land-rent share seems reasonable. Note that $\alpha > \gamma$ is the condition for the model to predict that slum land-use at the city edge precedes formal development.

Land-rent is the product of space-rent per unit volume (\$ per m³), space per unit land (BVAR m³ per m²) and the share of space-rent that is attributable to land, $1 - 1/\gamma$ in the formal sector and $1 - 1/\alpha$ in slums. We find that, at the centre in 2015, $x = 0, t = 2015$, $r_F(x, t, t) = \$65.38$ and $r_I(x, t) = \$17.67$ (using eqns. (5) and (12) with numbers from Table 3). Moving away from the centre these imputed land-rents drop at respective proportional rates $\theta\gamma/(\gamma - 1) = 0.17$ per km and $\theta\alpha/(\alpha - 1) = 0.095$ per km, nearly twice as fast in the formal sector as the informal. The wide gap in land-rents near the centre is indicative of inefficient land-use (an argument that we will develop precisely in Section 6), although we note that from equations (15) and (16), efficient switch points are generally not at the point of equality of r_F and r_I .

What are the components of these changing land-rents? In the formal sector the distance gradient is driven principally by commuting costs reducing net wages and hence space-rents, (as given by eqn. 1); land-rents fall faster than space-rents because of declining volume per unit land. For slums, as we saw above, observed space-rent is constant as lower commuting costs near the city centre are offset by greater crowding; the decline in land-rents is therefore due to lower volumes.

There are two other predictions of interest, with details in Appendix Section A1.6. The amenity derived from a unit volume of slum housing at the centre is 28% of a unit of formal volume; slum crowding decreases with distance, so this fraction rises, reaching 58% at 10 km.²⁸ This change reflects the quality continuum aspect of slums. The construction cost of a unit volume of slum housing is 12% that of formal sector construction cost per unit volume at the centre (cost expressed on an annual amortization basis); formal construction costs per unit volume decrease with distance as buildings get less tall, so at 10 kms this fraction reaches 25%. As implied, formal sector unit construction costs per unit volume at 10km are 50% of what they are at the centre.

4.3 Solving for the dynamic environment

The city is characterised by changing land-use and periodic redevelopment, as is modelled in the theory. Quantifying this requires two further pieces of information. One is the rate of interest, ρ , and we use the real interest rate $\rho = 0.057$, which is the average of the World Bank real interest rate (lending rate adjusted for inflation) for Kenya for the 14 years post 2002.²⁹ The other is the expected rate of increase of space-rent, \hat{p} . We have estimates of formal sector land-rents, $r_F(x, t, t) = \$65.38$ ($x = 0, t = 2015$, previous sub-section) and the capital value of land at date of redevelopment, $PV_F(x, \tau_i) = \$1414$ ($= \exp(7.254)$, Table 1 col. 6). Eqn. (19) links this capital value to the level and rate of increase of flow rents, dependent on price growth, and hence enables us to back-out the expected price increase. Using

²⁸ The amenity parameters $a_I = 0.734$ and $a_F = 1$ are constant.

²⁹ <https://data.worldbank.org/indicator/FR.INR.RINR?locations=K>

these numbers and the value for γ from Table 3, we solve for Φ , $\Delta\tau$, and \hat{p} in eqns. (17) - (19), restated here as:

$$\Phi = \int_0^{\Delta\tau} e^{(\hat{p}-\rho)t} dt = \frac{1 - e^{(\hat{p}-\rho)\Delta\tau}}{\rho - \hat{p}}, \quad \Delta\tau = \frac{\gamma - 1}{\hat{p}} \ln \left[\frac{\gamma}{\gamma - \rho\Phi} \right],$$

$$PV_F(x, \tau_i) = r_F(x, \tau_i, \tau_i)\Phi / \left[1 - e^{-\left(\rho - \frac{\hat{p}\gamma}{\gamma-1}\right)\Delta\tau} \right].$$

Solving these equations gives $\hat{p} = 0.0092$, $\Phi = 20.64$, $\Delta\tau = 89.29$ years. Thus, real space-rent appreciation is estimated to be about 1% a year. This is broadly consistent with the space-rent series produced by HassConsult as described in Appendix Section A2.4. The value-to-rent ratio, $\Phi = 20.64$, is close to that implied by an infinite stream discounted at 5.7%, and is in the middle of range of ratios reported on the internet by realtors for US cities.³⁰ The length of life of new buildings, 89 years, is novel and an estimate for which we know of no easy comparison.

5. City dynamics: The data and the model.

As reported in Section 3.1, we have tracings of all building footprints for 2003 and 2015, and LiDAR height data for 2015. In this section we use these data to describe the patterns of change that occurred in the city, and then compare these actual changes with predictions of the model. As we have seen, the model is calibrated on 2015 cross-section data, so using the model to predict the timing and extent of changes that occurred in the city is a demanding comparison.

5.1 Development and redevelopment in the formal sector

We look first at the formal sector, where the model predicts that there is a wave of development, followed by a wave of redevelopment (Section 2.5 and Fig. 2). Fig. 6 is the empirical counterpart, giving formal sector volume change (2003-15) per unit area at each distance and its decomposition into two parts, infill and redevelopment out to 12km of the 2015 city. Two distinct waves of change are apparent. The wave further out, peaking at around 9.5 – 10 km and significantly higher than other rates, is largely infill, defined as new buildings not intersecting any 2003 buildings. This is the empirical counterpart of formalisation; it is construction of formal structures at the edge of the city, some of them on land temporarily occupied by low-density slums, and some on land previously in non-built use.³¹ The wave nearer the city centre peaks at 3-4 km from the centre and is principally redevelopment,

³⁰ E.g., <https://smartasset.com/mortgage/price-to-rent-ratio-in-us-cities>

³¹ This formal sector development near the city edge occurs as the fraction of land covered by formal buildings increases between 2003 and 2015, as discussed in section 2.5. Not all of this land may have been occupied by slums prior to formal development. However, since we rely on one slum map in 2011, we do not know the formal status of all of this land in 2003. Costs of evicting slum tenants may also inhibit landowners from having land in slum usage close to the time of first formal sector development. Note also that “infill” can be on developer assembled lots where vacant land is built on and formally covered land is left as green space, so it is not truly infill but a rearranging of how land is used on assembled lots. There is also a small residual category (not shown) of torn down buildings with no new footprint appearing (yet) over the old.

defined as new buildings that overlap the footprint of 2003 buildings, where redevelopment volume is net of what was demolished.

These two waves are as predicted by the model. How do they line up in terms of exact timing and location? Table 4 gives timings of formalisation and first redevelopment predicted by the model using values of parameters that are reported in the text above. The first row gives date of formalisation, and the model predicts from eqn. (15a) that, if $D(x) = 0$, the 2003-15 peak would occur at around 12-12.5 km; the data in Fig. 6 has peak at 9.5-10 km, i.e. the model is predicting formalisation slightly early. This is consistent with generally positive values of $D(x)$ that delay formalisation. The predicted date of first redevelopment in Table 4 at 3.5 km is about 2030, while the distance band predicted to redevelop during the period 2003-15 is around 0.5 – 1 km from the CBD (Table 4 second row). The model is therefore predicting first formal development later than the wave that occurs in the data. The discrepancy arises as the predicted distance between places experiencing formalisation and first development is around 11.5 km (89 years times \hat{p}/θ), whereas the peaks in Fig. 6 are just 7 km apart. The ‘late’ first redevelopment predicted by the model could – amongst other things – be attributable to using the World Bank discount rate of 5.7%; a higher rate reduces the time interval between developments, $\Delta\tau$, and hence also the distance between successive building peaks. Notwithstanding these remarks, we note that Table 4 gives first formalisation in the city at 1917, surprisingly close to its date of founding, 1899. Of course, such long-run projections are way out of sample and highly speculative, given historical changes through that time period.

Table 4: Dates of Development							
	Distance from centre, x						
	0.5 km	1.5 km	2.5 km	3.5 km	10.5km	12.5km	14.5km
τ_1 : Formalisation $D(x) = 0$	1917	1924	1932	1940	1994	2009	2025
τ_2 : First redevelopment $D(x) = 0$	2006	2014	2022	2029	2083	2099	2114
Model prediction with space-rent path: $p(x, t) = p(0, 2015) \exp(\hat{p}(t - 2015) - \theta x) = 23.29 \exp(0.0092(t - 2015) - 0.071x)$							

A critical aspect of redevelopment in the data and in the model is the change in building height that it brings about. Fig. 7 splits out the formal sector height gradient in Fig. 5 to show how the mean heights of redeveloped buildings compare with those that did not change. In the formal sector beyond 1.5 km redeveloped buildings are significantly higher, nearly double the height of unchanged buildings in the interval 2-4 km. To get a number conceptually comparable to the model, we compare the height of the 80th percentile of redeveloped buildings, mimicking heights of newly completed buildings, to the 20th percentile of unchanged buildings, mimicking the height of those *just about* to be torn down. This is graphed in the figure and regressions on these slopes are reported in Table 1, cols. 4 and 5. From 1-4.5 km, the regression (using intercepts and slopes) and the figure both suggest that building height

increases around 2.75 – 3 times with redevelopment. This compares with a model prediction given by equation (18) where volume increases by a factor of 3.23, at each redevelopment.

5.1.1 Role of terrain.

An element of $D(x)$ which can be observed and that bears on the timing of development (Section 2.5) is the terrain. In Appendix 3 (Land Quality) we report the results of investigating this with water (river, wetland, pond) and ruggedness as independent variables at various spatial scales.³² There are two main findings. Fig. 6 indicates infill occurring at all distances, including close to the centre. Table A3.1 in the Appendix shows that formal sector infill near the centre is on much worse quality land than already developed buildings, indicating delay in the date in which poor terrain land is first developed. The effect falls rapidly with distance so, by 10km out, there is no clear difference between infill and already developed areas. Second, Table A3.2 shows that the height of redeveloped buildings is unaffected by underlying terrain fundamentals. Overcoming terrain involves a sunk cost of land improvement that does not affect the height of structures that are then constructed (i.e. κ_F in eqn. (10) is not affected).

5.2 Slum dynamics

The model predicts that, if the condition given in proposition 2 (i) is satisfied, new slums will form on the edge of the city. That near the city edge there will be conversion of existing slums into formal sector use. That existing slums will become increasingly crowded. And that there will be no slums remaining near the city centre unless there are substantial formalization costs, $D(x) > 0$.

The data is broadly consistent with these predictions. First, there are new slums. We have a proxy measure, which are areas identified as slums in the 2011 mapping but not in the 2003 mapping. As noted above, some of the differences are due to different methodologies employed by researchers in defining slums. Nevertheless, in the pink areas in Fig. 3 we see large tracts of new slum development at and beyond the 2003 city edge. Referring back to dates of first formalisation in Table 4, we see that in the early 2000s these should be at some distance beyond 10 km (i.e. beyond places formalising). In Fig 8 we show the change in slum volume for slums as mapped in 2011 and indicates that at about 8 km there is one peak of infill (new slum cover) and then another at 10.5 km. The difference between these two peaks is interesting. The peak at 8 km is on government owned slum lands where we presume D 's are too high to allow formal use yet and slum usage is intensified as prices rise. The peak at 10.5 km is disproportionately on private land, so it is on 'market driven' slum expansion. Fig. A2.7 in the Appendix displays these ownership patterns.³³

Second, there are some slum conversions we might infer from the 2003 and 2011 mapping of slums. In Appendix Fig. A2.5 we show rates of building teardown for 2011 slums (non-converted) and slums that disappear between the 2003 and 2011 mappings (converted). Teardown rates for these converted slums average nearly 60%, three times greater than in non-converted areas. Appendix Fig. A2.6 also shows that buildings on converted slum lands are generally built significantly taller than those in current slums.

³² Ruggedness is defined as the standard deviation of elevation across the 30x30m grid squares of the 90x90m square neighbourhood which is queen neighbours plus the own square.

³³ In addition to purely private, we are including some riparian lands as graphed in Appendix Fig. A2.7. From Kenya (2018) we know these areas contain private developments, as well as areas with possessory rights.

Third, there is densification of existing 2011 slums within the 2003 city-boundary. In Fig. 8, we show total slum volume changes and their decomposition into “redevelopment” and infill. Up to about 6km, ‘redevelopment’ is akin to rearrangement of the Meccano parts and is conceptually not really distinguishable from infill. Both are part of a process of increasing crowding in slums, as CAR increases. Figs. 5 and 7 indicate, as assumed in the model, that there is essentially no increase in slum height, which remains low and uniform across the city. Volume changes are driven entirely by CAR changes.

Finally, no slums get created near the centre, but there is persistence of some large slum areas, as we discuss in detail in Section 6.

5.3 Total volume and population

What are the implications of these changes for aggregate building stock and population? We have data on each of these for 2003 to 2015. As noted in the Introduction, total built volume within the 2003 boundary increased by 47% between 2003 and 2015, and that in the 2015 boundary by 60%. Based on the 1999 and 2009 censuses for the 2003 area of the city, population grew at 3.8 p.a. which, for the period 2003 – 2015, would project to 57% for 2003 boundary and 68% for the 2015 city boundary.

In the model, volume comes from integrating over all x at a particular date, weighting volume per unit area by area at the place.³⁴ For population, volume at each space is divided by each household’s demand for space, $s_i(x, t)$, $i = I, F$, so takes the form

$$L(t) = \sum_{i=1}^{imax(t)} \int_{x_{i+1}(t)}^{x_i(t)} n(x)v_F(x, \tau_i)/s_F(x, t)dx + \int_{x_1(t)}^{x_0(t)} n(x)v_I(x, t)/s_I(x, t)dx.$$

In this expression $n(x)$ is land area (in a circular city, proportional to distance x from the centre). The first term integrates over land in its i -th generation of development at date t (i.e. land in interval $(x_{i+1}(t), x_i(t))$), and sums over generations up to that which has been redeveloped the most times, denoted $imax(t)$. Applying this equation to volume (i.e. setting $s_i(x, t) = 1$) gives the volume increase along a ray ($n(x) = 1$) over the 12 year period of 38%. Population increase exceeds volume as rising income and space-rents create income and substitution effects in household demand for space, with the latter dominating. This gives a population increase along a ray of 56%. In a circular city each of the rates of increase are lower, at 27% and 34%; this is because the model predicts that redevelopment during this period is close to the centre than it is as noted above, i.e. covering a small land area.

To conclude this section, we remind the reader that parameters of the model were estimated entirely on cross-section data, together with an imported value of the discount rate. The model does well in predicting many aspects of Nairobi’s development through time, albeit with the proviso that the long time period between developments has the effect of making the model predict redevelopment later than is in the data.

³⁴ This is integrating under the surface of Fig. 2 (anti-logged) at a particular place, and weighting by the area of each place x .

6. The cost of delayed formal sector development

Finally, we turn to quantifying some of the costs of persistent slums in central parts of Nairobi. Chief amongst these is Kibera, often described as the largest slum in Africa. The costs and benefits of such slums – and policy towards them – is a complex and contentious issue. We do not seek to quantify all elements (such as community dislocation), and focus on a single element of the equation, namely the real income loss due to inefficient land-use. This is measured by the loss in land values associated with informal settlement on high opportunity cost land.

6.1 Formalisation costs

Nairobi was founded in 1899 and the British colonial government, as was typical throughout Africa, housed the African population in informal settlements without land title (Olima, 2001). After independence in 1963, a series of reforms resulted in over 85% of land in Nairobi becoming privately owned under charges of widespread corruption (Southall, 2005). However, older slums mostly within 6 km of the centre remain ‘government owned’. These areas are not managed by the government but rather by slumlords who operate ‘illegally’ and make high profits. Gulyani and Talukdar (2008) estimate payback periods on an investment in a single room of just 20.4 months. This is consistent with the fact that land is ‘free’ to slumlords; and, by our calibration, land’s share in revenue in the informal sector is 75%. Moreover, slumlords have characteristics that are problematic for formalisation. In Kibera for example, of 120 slumlords surveyed, 41% were government officials, 16% (often the biggest holders) were politicians, and 42% were other absentee owners (Syagga, Mitullah, and Karirah-Gitau 2002 as cited in Gulyani and Talukdar 2008). The political economy issue is that if the government were to auction the land it ‘owns’ (or give it to the tenants), the slumlords would lose their claim to the revenue. Having well-connected bureaucrats and political figures opposed to formalisation presents a political problem.

This problem is magnified when there are historical private claims to the land, as is typically the case. Kibera gives a nice example. The 1000 acres in Kibera were awarded to Nubian soldiers for service in 1912 by the British. They occupied a portion of the land, but at independence their claims (but not tenancy) were revoked and land reverted to the government. The majority area of present day Kibera which is not occupied by Nubians and their descendants was settled by others, on the basis of claims illegally allocated by local chiefs and bureaucrats. The moral right of the Nubian descendants to at least the land they occupy is well recognized but the unwillingness to grant them title has historically been a road-block to redevelopment (Etherton 1971; Joireman and Vanderpoel, 2011). The literature has similar stories for other major slums in Nairobi.³⁵

³⁵ For example, Mathare 3km northeast of the centre, was originally a stone quarry. When the quarry closed the land went to the Department of Defence and the area was occupied by squatters. There then followed a long history of squatters attempting to set up collectives to ‘buy’ the land in competition with land buying companies, dissolution of the cooperatives, allegations of corruption, and competing claims on the land (Medard 2006). Today the majority of the slum part of Mathare (about 1 km square) is private, but significant portions remain under government ownership of some sort (police, central government, Nairobi City Council). Syagga (2011) analyses

6.2 The loss of land value due to slum persistence.

In our open city model the welfare costs of inefficient land use are given by the present value of land-rents foregone. To assess the cost of delayed formalisation we use the model and estimated parameters to calculate the present value of land-rents earned by land in different uses. The gap between the values of land in formal versus slum use measures the real income loss due to inefficient land-use, although it does not capture other costs and benefits outside the model, such as social costs of disruption involved in slum redevelopment and community dislocation, or possible productivity benefits of spatial reorganisation of the city.

For a unit of land at place x we calculate the present value, in 2015, of being held in slum use until conversion to formal use (with $D(x) = 0$) at date z , and denote this value $PV(x, z)$. For perpetual delay versus conversion in 2015 the expressions are respectively

$$\begin{aligned} PV(x, z = \infty) &= \int_{2015}^{\infty} r_I(x, t) e^{-\rho(t-2015)} dt \\ &= \int_{2015}^{\infty} r_I(x, 2015) e^{[\hat{p}\alpha/(\alpha-1)-\rho](t-2015)} dt, \end{aligned} \quad (22)$$

$$\begin{aligned} PV(x, z = 2015) &= \sum_{i=0}^{\infty} R_F(x, z + i\Delta\tau) e^{-\rho i\Delta\tau} \\ &= r_F(x, 2015, 2015) \Phi / \left[1 - e^{-\left(\rho - \frac{\hat{p}\gamma}{\gamma-1}\right)\Delta\tau} \right]. \end{aligned} \quad (23)$$

These present values are given in Table 5 as a function of distance from the centre, x . The first row gives the present value if formalisation occurs at date $z = 2015$, and the second row gives the present value if it never occurs. For each distance reported in the table, 2015 is beyond the efficient date for formalization given in Table 4. We compare the cost of never formalising with formalising in 2015 for each distance, $PV(x, z = 2015) - PV(x, z = \infty)$. These range from \$920 per m² in the 0-1 km distance band to \$249 per m² at 6-7 km from the centre.

We illustrate the gains from conversion by looking at lands 3-4km from the centre, which includes some parts of Kibera. At 3-4km the cost of perpetual informality as compared to switching to formal sector use in 2015 is row 1 minus row 2, or \$774-\$284=\$490 per m². There are 1.07mn m² of slum land in that distance band. There is thus an aggregate gain from converting at 2015 compared to perpetual delay of \$525mn. We give three perspectives on this.

how Kenyan tenure legalization can take decades to implement due to the need to reconcile interests of stakeholders and offers more examples such as the Korogocho slum.

Table 5: The value of land formalised at different dates							
Present values in 2015 in \$2015 per m ² .							
	Distance from centre, x						
Date of formalisation, z	0-1 km	1-2 km	2-3 km	3-4 km	4-5km	5-6km	6-7km
$PV(x, z = 2015)$	1297	1092	919	774	652	549	462
$PV(x, z = \infty)$	377	343	312	284	258	235	213
Slum land, km ² , 2011	0	0.0024	0.24	1.07	2.22	1.9	1.32
No. slum households, 2009	0	0	2920	29,070	45,810	33,100	28,390
Lower bound on D (15a)	493	395	314	248	194	150	113

First, with so much money on the table, why is land at 3-4 km not converting from slum to formal use at a faster rate? We have argued that there are financial costs and political barriers to formalization, and a lower bound on the monetary value of these costs is given the bottom row of Table 5. This number is per m², and is derived by solving eqn. (15a) as an equality. At 3-4 km this is \$248, equal to 32% of the present (market) value of land in formal use and 87% of the value in slum use, clearly non-trivial amounts consistent with our political story on Kibera. If there were slums in the centre the lower bound on D would be \$493 but, since there are none, we know that the pressure to redevelop was so intense as to overcome formalisation barriers.

Second in terms of the one-off gain in land values from conversion at 3-4 km, the surplus of \$525mn is about 10% of Nairobi's GDP in 2015³⁶. If we add all the net gains out to 7 km, that is about \$2.5bn or 45% of Nairobi's GDP.

The third perspective hints at a political solution. Suppose slumlords were fully compensated for conversion by \$284 per m², as if they had the right to utilize the land forever. That still leaves a remaining surplus of \$490 per m² at 3-4 km. For the 29,070 households affected, the gain is about \$18,000 per household. This is a very large sum, for households paying on average less than \$700 a year in rent (from the NORC data and consistent with the rent calculations we use in Section 3.3.1). At 4-5 km, a distance including much more of Kibera, the same type of calculation gives a surplus of about \$19000 each for the 46,000 households.³⁷

In these calculations, we have ignored moving costs for slum residents, broadly defined. For those forced to move there could be losses in terms of job location and social networks, although proper relocation programs could mitigate those. Nevertheless, a key point of the calculations is that, relative to income and space-rents paid, there is an enormous surplus to play with to compensate residents. In one version of a 'just' non-political world this would be solved by giving tenants the land titles and allowing them to sell themselves when and if they are ready.

³⁶ GDP per capita in Kenya was \$1090 in 2015. We set it at \$1300 in Nairobi for a population of 4.2mn.

³⁷ These model-based numbers are substantially higher than preliminary back-of-the-envelope type calculations based on raw data reported in CEPR DP 11211 which give a gain of \$13,000 US per slum household for the core city.

These calculations are subject to bias. First, to be valid, slum lands near the centre must be the same quality as available formal sector lands. That is, it is slum history and formalisation costs driving the current delay in development of slum lands, not geography. The basis of comparison is the vacant land for which we have sales price data. In Appendix Table A3.3, we perform a border experiment at slum boundaries comparing quality of slum lands with formal sector infill (vacant) lands that have just been built upon. We show that at distances out to 6km (where these old slums are), for elevation, ruggedness in the small, ruggedness in the large, presence of water, and being lower than mean neighbourhood elevation, of the 20 cases for slums compared to formal sector, 17 show no differential, two are better, and one worse. It seems differentials in geography do not drive non-conversion of slums near the city centre.

Second, as a downward bias, we have used formal sector residential space-rents as the basis for gain. For slum lands some highest and best use might be commercial which could have higher values nearer the city centre. Finally, as an upward bias, there is the general versus partial equilibrium distinction, in a context with heterogeneous households. In our context, with identical households and an open city, we have captured welfare gains in terms of increased land values. With households heterogeneous in terms of preferences for x 's and a 's we are effectively looking at the marginal household. In a general equilibrium context, there would be intra-marginal household at each location who might be willing to pay less to live in the formal sector and more in the slum. For large-scale conversions, this is a source of upward bias in our calculations.

7. Conclusions

This paper develops a framework for analysing a growing monocentric city in which residents make housing and location choices and developers sink capital in built structures, acting on the expectation of future rental income. We tailor the model to a developing country, so include an informal or slum sector in which structures are cheap and malleable, so capital is not sunk and construction can take place even if land-rights are insecure. We show how the model predicts the conversion of land from informal to formal as the city grows, and how areas go through successive waves of teardown and redevelopment. The model provides a platform for further work, such as exploring the role of expectations in urban growth, endogenising urban productivity, and moving from a monocentric to polycentric urban structure.

The application is to Nairobi, a fast-growing city that, we argue, has many features in common with other developing cities. Based on a unique data set constructed from high-resolution aerial photographs, we study the city in cross-section (2015) and change (2003-2015). We characterise how building height and ground cover, in both the formal and informal sectors, vary with distance from the city centre and use this cross-section to calibrate key parameters of the model. We find – consistent with assumptions in our model – that, closer to the centre, both sectors offer more built volume per unit area; the formal sector by building tall and the informal by crowding (i.e. more cover per unit area). We derive estimates of parameters of the two building technologies, indicating that land-rent (as opposed to construction) cost, is a much higher share of total rental income in the informal sector than in the formal. Turning to

changes through time, the model does well in predicting the pattern of slum growth and two waves of formal sector development that occurred in the period 2003-15. One was redevelopment of formal structures relatively close to the city centre, with redevelopment typically trebling building height as predicted by the model. The other was new formal sector development further out, on the fringe between formal and informal areas.

A feature of Nairobi and other developing cities is the persistence of slums in high land value areas near the city centre. This is rationalised in the model as being due to a package of barriers to formalisation. In Nairobi, many of these barriers arise as 'slumlords', who rent out property at market rates, do not have sound claims on the land, so resist a process of formalisation that might deprive them of their flow of rental income. We use estimates from the model to calculate the losses due to these barriers, and find they are very substantial. Even if slumlords were to be paid-off (compensating for a loss of perpetual control), conversion would yield a surplus of about \$18,000 per slum household in a context where they pay about \$500-700 a year for their housing.

References

- Ahlfeldt, G.M , S.J. Redding, D.M. Sturm and N. Wolf. 2015. “The Economics of Density; Evidence from the Berlin Wall”, *Econometrica*, 83(6) 2127-2189.
- Ahlfeldt, G.M and D. P. McMillen. 2014. “New Estimates of the Elasticity of Substitution of Land for Capital”, processed, LSE.
- Ahlfeldt, G.M and D. P. McMillen. 2018. “Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago, 1870 – 2010”, *Review of Economics and Statistics*, 100(5), 861-875.
- Baruah, N., J.V. Henderson and C. Peng. 2017. “Colonial Legacies: Shaping African Cities”, SERC DP 0226:
- Besl, P., and N. McKay. 1992. “A Method for Registration of 3-D Shapes.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 14(2), 239-256.
- Bird, J., P. Montebruno, and T. Regan. 2017. “Life in a slum: understanding living conditions in Nairobi’s slums across time and space”. *Oxford Review of Economic Policy*, 33(3), 496-520.
- Braid, R. M. 2001. "Spatial Growth and Redevelopment with Perfect Foresight and Durable Housing," *Journal of Urban Economics*, 49(3), 4 25-452
- Brueckner J and S. Rosenthal. 2009. “Gentrification and Neighbourhood Housing Cycles.” *Review of Economics and Statistics*, 91(4): 725–743
- Capozza, D.R. and R.W. Helsley. 1989. “The Fundamentals of Land Prices and Urban Growth”, *Journal of Urban Economics*, 26, 295–306.
- Case, K. E. 2007. "The Value of Land in the United States: 1975-2005; *Land Policies and Their Outcomes*", Gregory K. Ingram and Yu-Hung Hong (Eds.), Cambridge, MA: Lincoln Institute of Land Policy
- Cavalcanti, T., D. Da Mata.and M. Santos. 2019. “On the determinants of slum formation”. *The Economic Journal*, 129, 1971–1991
- Charles, K. 2010. “A negotiated framework for rehabilitation of riparian zones in Nairobi city: the case of Mathare river valley”, 46th ISOCARP Congress
- Combes, P-P, G. Duranton and L. Gobillon. 2018. “The Costs of Agglomeration: House and Land Prices in French Cities”, *Review of Economic Studies*, forthcoming
- Combes, P-P, G. Duranton and L. Gobillon. 2017. “The Production Function for Housing: Evidence from France”, CEPR DP 11669
- CSUD, 2003. “Nairobi Regional Project: GIS Maps”, <http://csud.ei.columbia.edu/projects/nairobi-regional-project/nairobi-gis-maps/>
- Davis, M. 2006. “Planet of Slums”, Verso, London.
- Duranton G and D Puga. 2015. “Urban Land Use” in *Handbook of Regional and Urban Economics*, G. Duranton. J.V. Henderson and W. Strange (eds). Elsevier

- Etherton, D. 1971. Mathare Valley: a case study of uncontrolled settlement in Nairobi. Housing Research and Development Unit, University of Nairobi.
- Ferreira, P., A. Monge-Naranjo, and L. Torres de Mello Pereira. 2016. "Of Cities and Slums," FRB St. Louis Working Paper No. 2016-22.
- Fujita, M. 1982. "Spatial development processes with renewal in a growing city," *Journal of Urban Economics*, 12, 22-52.
- Getcher, M. and N. Tsivanidis, 2018. "The Welfare Consequences of Urban Renewal: Evidence from the Mumbai Mills Redevelopment", working paper.
- Gulyani, S. and D. Talukdar. 2008. "Slum Real Estate: The Low-Quality High-Price Puzzle in Nairobi's Slum Rental Market and its implications for Theory and Practice." *World Development*, 36, 1916-1937
- Harari, M. 2020. "Cities in bad shape: Urban geometry in India", *American Economic Review* (forthcoming).
- Harari, M and M. Wong. 2019. "Slum Upgrading and Long-run Urban Development: Evidence from Indonesia", working paper.
- Huchzermeyer, M. 2007. "Tenement City: The Emergence of Multi-storey Districts Through Large - scale Private Landlordism in Nairobi" *International Journal of Urban and Regional Research*
- IPE Global Private Limited and Silverwind Consultants. 2013. Consultancy Services for City/Municipal Situational Analysis of Conditions of Informal Settlements in 15 Municipalities. Prepared for the Government of Kenya, Ministry of Lands, Housing and Development.
- Joireman, S.F. and R.S. Vanderpoel. 2011. "In Search of Order: State Systems of Property Rights and Their Failings" in *Where There is No Government: Enforcing Property Rights in Common Law Africa*, Sandra F. Joireman (ed.), Oxford University Press.
- Kelly, R., 2003. "Property Taxation on Kenya"
<http://www1.worldbank.org/publicsector/decentralization/June2003Seminar/Kenya.pdf>
- Kenya. 2018. "Report on Riparian Lands Conservation and Management Policy Framework." Kenya National Land commission
- Lall, S., J.V. Henderson, and A.J. Venables. 2017. *Africa's Cities: Opening Doors to the World*. World Bank <https://openknowledge.worldbank.org/handle/10986/25896>
- Marx, B., T. Stoker and T. Suri. 2013. "The Economics of Slums in the Developing World," *J. Econ. Perspect.* 27, 187–210.
- Medard C. 2006. "City Planning in Nairobi: The Stakes, the People, the Sidetracking" in *Nairobi today: The Paradox of a Fragmented City*. D. Rodrigues-Torres (ed). Mkuki na Nyota Publishers, Nairobi
- Michaels, G., A. Dahlstrand-Rudin, D. Nigmatulina, F. Rauch, and T. Regan. 2019. "Planning Ahead for Better Neighbourhoods: Long Run Evidence from Tanzania." Working paper.

- Mwaura, A. M. 2006. Policy Review for Zones 3, 4 and 5, Nairobi, Kenya 42nd ISoCaRP Congress. http://www.isocarp.net/Data/case_studies/886.pdf
- Narubi, A. O. 2015. "An evaluation of property rate collection and enforcement of devolved systems of governance: A case study of Nairobi city-county." MA thesis, University of Nairobi, Department of Real Estate and Construction Management.
- Olima, W. 2001. The Dynamics and Implications of Sustaining Urban Spatial Segregation in Kenya: Experiences from Nairobi Metropolis; www.begakwabega.com/documenti/dynamics-implications-urban-segregation.pdf
- Selod, H. and L. Tobin. 2018. "The Informal City," World Bank Policy Research Working Paper No. 8482.
- Solow, R M. and W.S. Vickrey. 1971. "Land use in a long narrow city," Journal of Economic Theory, Elsevier, vol. 3(4), pages 430-447, December.
- Southall, R. 2005. "The Ndungu report: land and graft in Kenya". Review of African Political Economy 32(103), 142-151.
- Syagga, P. 2011. Land Tenure in Slum Upgrading Projects. Les cahiers d'Afrique de l'Est, IFRA Nairobi, pp.103-113.
- Syagga, P., W. Mitullah, and S. Karirah-Gitau. 2002. "A rapid economic appraisal of rents in slums and informal settlements," Prepared for Government of Kenya and UN-HABITAT Collaborative Nairobi Slum Upgrading Initiative.
- Taubenbock, H., N.J. Kraff and M. Wurm 2018 "The morphology of the Arrival City - A global categorization based on literature surveys and remotely sensed data", Applied Geography 92, 150-167.
- UN. 2018. World Urbanization Prospects, 2018. <https://population.un.org/wup/>
- Un-Habitat. 2011. Ghana Housing Profile
- UN-Habitat. 2016. "Slum Almanac 2015/2016: Tracking Improvement in the Lives of Slum Dwellers".
- Williams, S., E. Marcello and J. M. Klopp. 2014. "Toward Open Source Kenya: Creating and Sharing a GIS Database of Nairobi," Annals of the Association of American Geographers, 104:1, 114-130.
- World Bank. 2006. Where is the Wealth of Nations: Measuring Capital for the 21st Century, World Bank, Washington DC.
- Yeom, J., Y. Kim, and J. Lee. 2015. "Hierarchical ICP Matching to Enable Interoperability of Multisource Building GIS Data." Advanced Science and Technology Letters, Vol. 89 (Architecture and Civil Engineering) pp. 55-60.

Figure 1: Urban development with perfect foresight

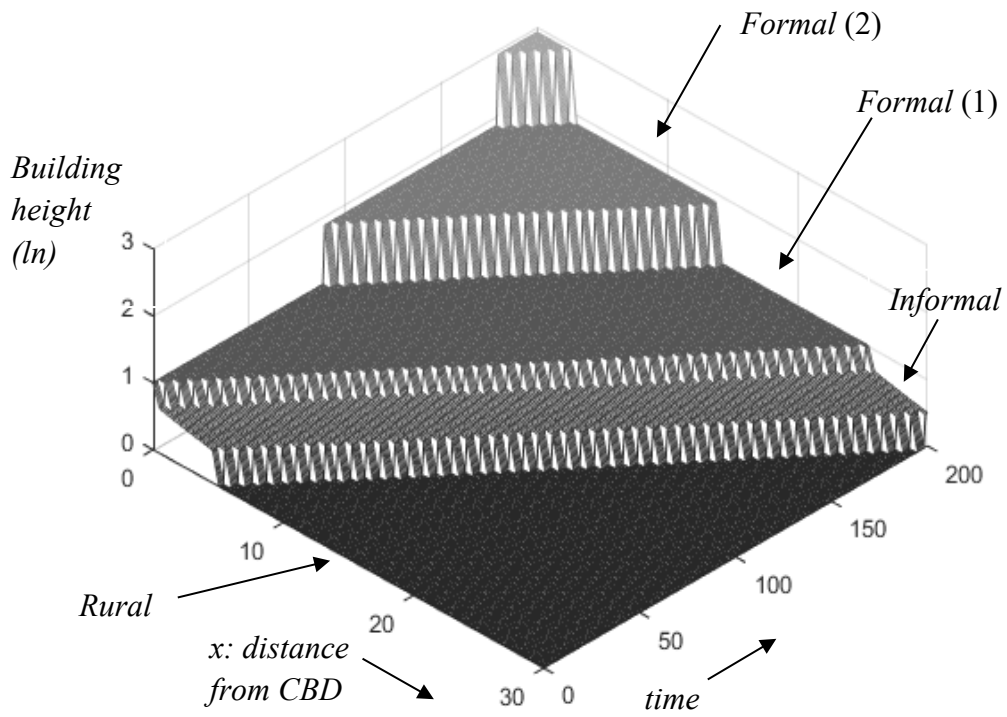


Figure 2: The hotchpotch - random variation in fixed costs of development, $D(x)$

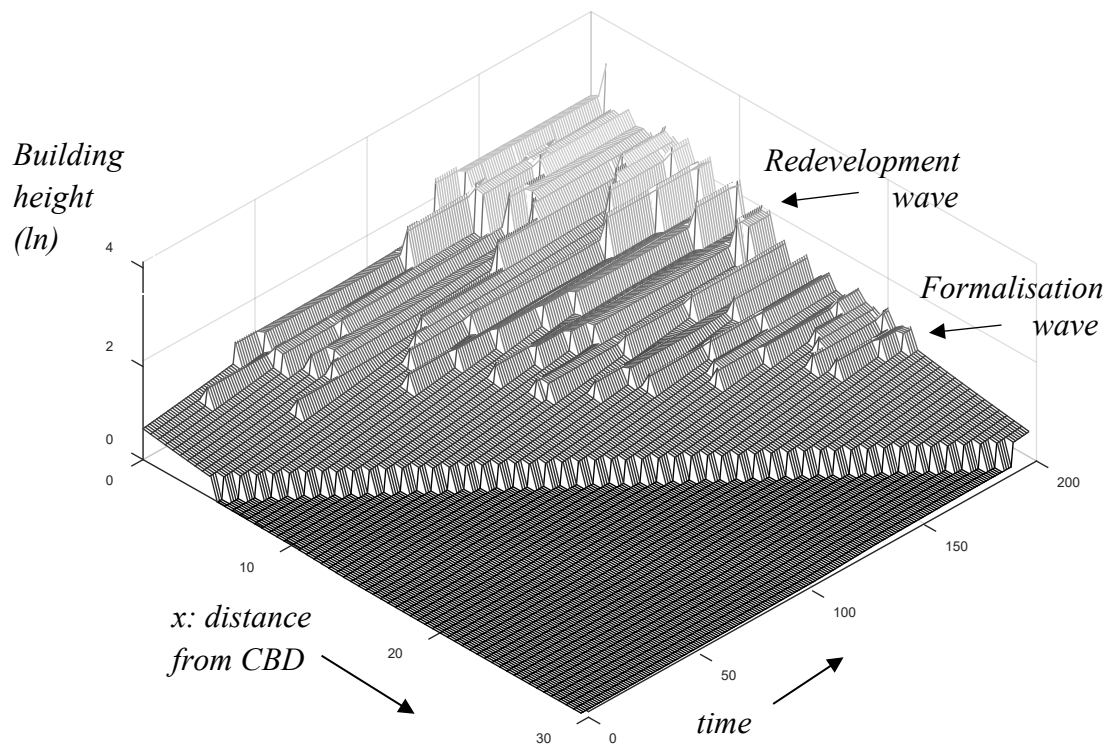
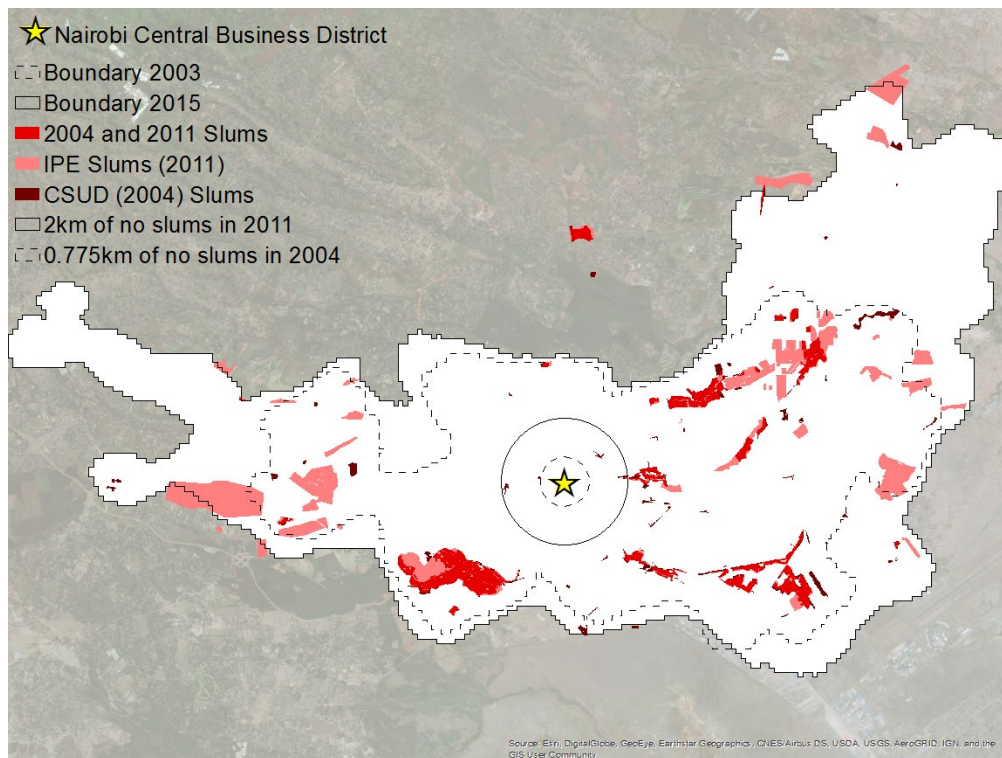
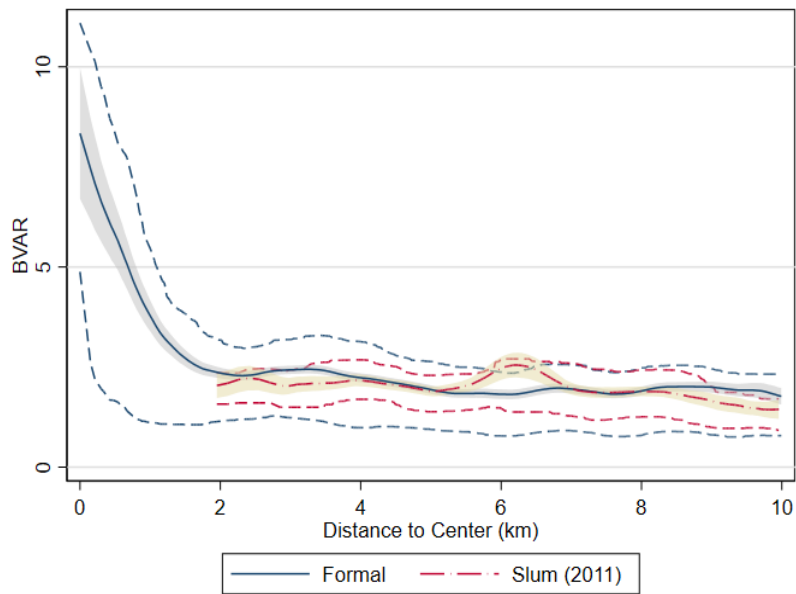


Figure 3: City shape and slums



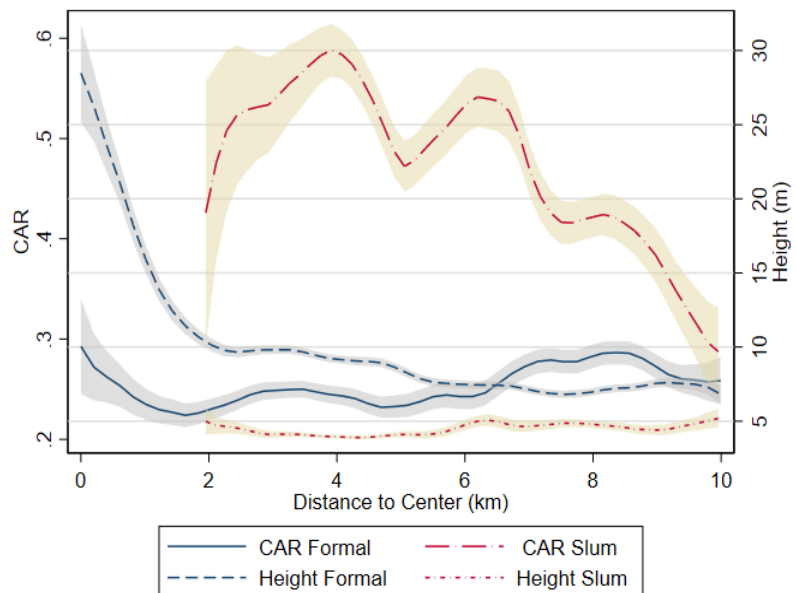
Notes: This figure maps two slum definitions; IPE from 2011 (light red), CSUD from 2003 (dark red), and the overlap of the two (red).

Figure 4: Built volume per unit area (BVAR)



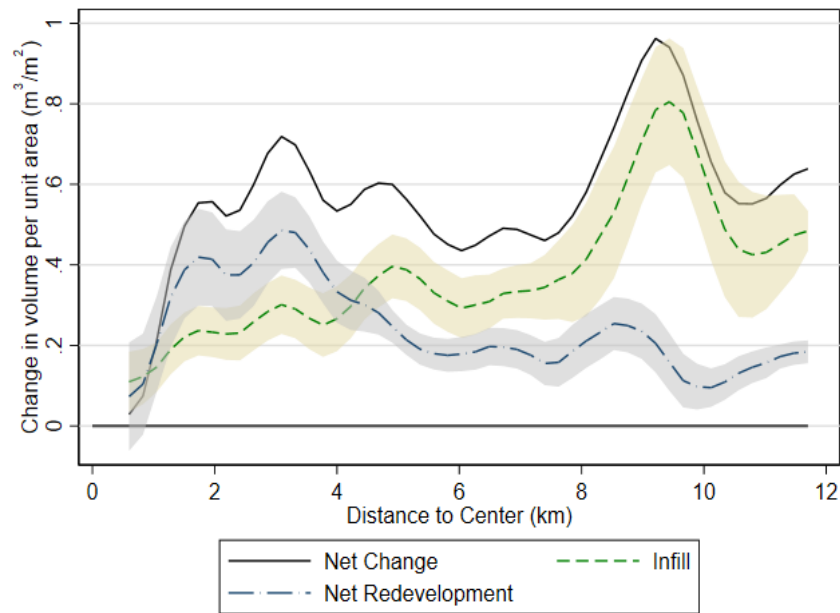
Notes: Built volume to area ratio (BVAR) by distance from the centre for formal and slum sectors. Lines show the local average BVAR, shaded areas show local 95% confidence intervals, and dashed lines show local 25th and 75th percentiles. Local statistics calculated using an Epanechnikov kernel with bandwidth of 300m.

Figure 5: Built cover per unit area (CAR) and height



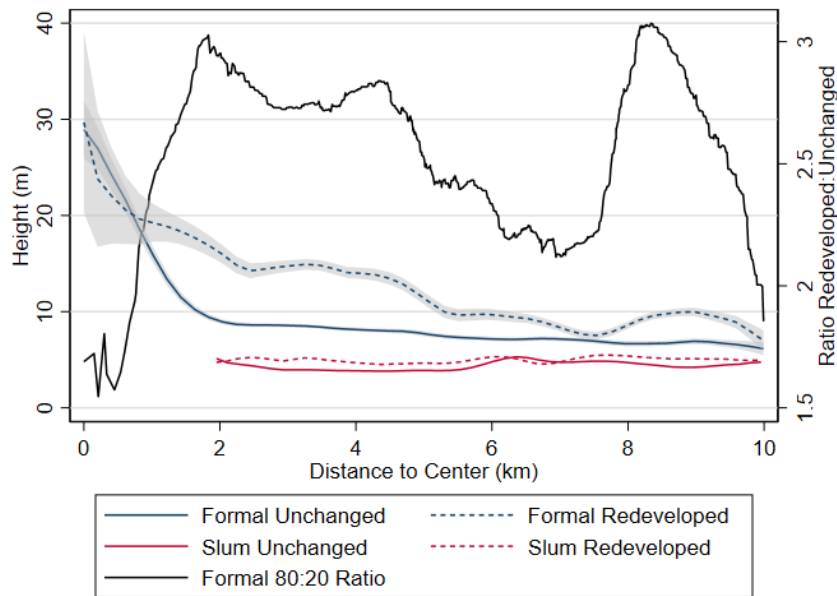
Notes: Cover to area ratio (CAR) and building height by distance from the centre for formal and slum sectors. Lines show the local average CAR and height, and shaded areas show local 95% confidence intervals. Local statistics calculated using an Epanechnikov kernel with bandwidth of 300m.

Figure 6: Formal sector changes in volume per unit area (2015 city to 12km)



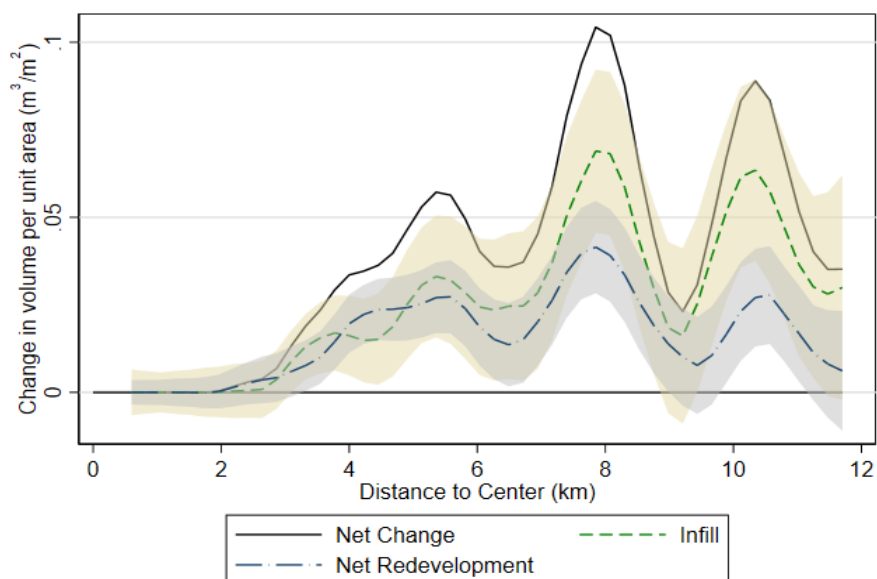
Notes: The change in formal sector volume, by redevelopment, infill and total net change, calculated in bins of 300m inside the 2015 city boundary and divided by the area of each bin. Shaded areas show local 95% confidence intervals for infill and redevelopment only. Local statistics smoothed using an Epanechnikov kernel with bandwidth of 300m.

Figure 7: Mean height - unchanged and redeveloped buildings



Notes: Height of unchanged and redeveloped buildings by distance from the centre for formal and slum sectors. Lines show local average height and shaded areas show local 95% confidence intervals for the formal sector only. The right axis is the ratio of the local 80th percentile of redeveloped height to the local 20th percentile of unchanged height. Local statistics calculated using an Epanechnikov kernel with bandwidth of 300m.

Figure 8: Slum sector changes in volume per unit area (2015 city out to 12km)



Notes: The change in slum sector volume, by redevelopment, infill and total net change, calculated in bins of 300m inside the 2015 city boundary and divided by the area of each bin. Shaded areas show local 95% confidence intervals for infill and redevelopment only. Local statistics smoothed using an Epanechnikov kernel with bandwidth of 300m.

Online Appendices

Appendix 1: Theory

A1.1 Further details of the model:

Details of the open city equilibrium model underpinning the model of the text are as follows:

Households: At date t a representative urban household living at distance x from the centre receives net income $w(t, x)$. The indirect utility function is $U^*(p(x, t), w(x, t))$ and the space-rent $p(x, t)$ must be such that at all occupied places $U^*(p(x, t), w(x, t)) = u_0$. In Sections 2.1-2.3 we assume that housing is normal, $w(x, t)$ is increasing through time, and u_0 is constant. Household demand for quantity adjusted housing volume is, by Roy's identity, $s(x, t)a = -\frac{\partial U^*}{\partial p(x, t)} / \frac{\partial U^*}{\partial w(x, t)} = \beta w(x, t) / p(x, t)$. The last of these equations assumed preferences are Cobb-Douglas, i.e. $U^*(p(x, t), w(x, t)) = p(x, t)^{-\beta} w(x, t)$, with corresponding direct utility function $U(s(x, t)a, c(x, t)) = (s(x, t)a)^\beta c(x, t)^{1-\beta} \beta^{-\beta} (1-\beta)^{\beta-1}$, where $c(x, t) = w(x, t) - p(x, t)s(x, t)a$.

From Section 2.4 onwards the rates of growth of $w(x, t)$ and form of U^* are such that space-rent grows at constant exponential rate through time and declines exponentially with x , interpreted as distance from the centre. The simplest primitives supporting this are (i) gross wage $W(t)$ equal to exogenous labour productivity and growing exponentially at constant rate \hat{w} , so $W(t) = W(0)\exp(\hat{w}t)$. (ii) Commuting costs taking fraction $1 - \exp(-\delta x)$ of gross wage, so net wage $w(x, t) = W(0)\exp(\hat{w}t)\exp(-\delta x)$. (iii) Cobb-Douglas preferences.

With the open city assumption $U^* = p(x, t)^{-\beta} w(x, t) = u_0$. This gives equilibrium space-rents $p(x, t) = \bar{p}\exp(\hat{p}t - \theta x)$ where $\hat{p} = \hat{w}/\beta$, $\theta = \delta/\beta$ and $\bar{p} = (W(0)/u_0)^{1/\beta}$. Notice that \hat{w} can be interpreted as growth of the ratio of gross wages to outside utility.

Labour and population: In the open-city equilibrium population is endogenous. Population at a point is v/s , total volume supplied divided by consumption of floor space per household. Total city population at date t is therefore

$$L(t) = \sum_{i=1}^{imax(t)} \int_{x_{i+1}(t)}^{x_i(t)} n(x)v_F(x, \tau_i)/s_F(x, t)dx + \int_{x_1(t)}^{x_0(t)} n(x)v_I(x, t)/s_I(x, t)dx.$$

In this expression x is distance from the centre, and $n(x)$ is land area (in a circular city, proportional to x). The first term integrates over land in its i -th generation of development at date t (i.e. land in interval $(x_{i+1}(t), x_i(t))$), and sums over generations up to that which has been redeveloped the most times, denoted $imax(t)$. The second term is slum population.

A1.2 Derivation of equations (15) & (16)

Using (7) in (13), excluding dates before τ_0 and after τ_2 , gives

$$PV_{02}(x) = \int_{\tau_0}^{\tau_1} r_I(x, t)e^{-\rho t} dt + v_F(x, \tau_1) \int_{\tau_1}^{\tau_2} p(x, t)e^{-\rho t} dt - [k(v_F(x, \tau_1)) + D(x)]e^{-\rho\tau_1}.$$

Differentiating, noting that volume and other development dates are optimised gives (15) of the text:

$$\frac{\partial PV_{02}(x)}{\partial \tau_1} e^{\rho \tau_1} = r_I(x, \tau_1) - v_F(x, \tau_1)p(x, \tau_1) - \rho[k(v_F(x, \tau_1)) + D(x)] = 0.$$

Similarly for date τ_2 , excluding dates before before τ_1 and after τ_3 , gives (16):

$$PV_{13}(x) = v_F(x, \tau_1) \int_{\tau_1}^{\tau_2} p(x, t) e^{-\rho t} dt - [k(v_F(x, \tau_1)) + D(x)] e^{-\rho \tau_1} \\ + v_F(x, \tau_2) \int_{\tau_2}^{\tau_3} p(x, t) e^{-\rho t} dt - k(v_F(x, \tau_2)) e^{-\rho \tau_2}.$$

$$\frac{\partial PV_{13}(x)}{\partial \tau_1} e^{\rho \tau_2} = v_F(x, \tau_1)p(x, \tau_2) - v_F(x, \tau_2)p(x, \tau_2) + \rho k(v_F(x, \tau_1)) = 0$$

A1.3 Proposition 2: Transition from rural to informal occurs at $\{x, \tau_0\}$ given by: $p(x, \tau_0) = (r_0/(\alpha - 1))^{(\alpha-1)/\alpha} \kappa_I^{1/\alpha} \alpha/a_I$, (eqn. 14a). Transition rural to formal occurs at $\{x, \tau^*\}$ given by: $p(x, \tau^*) = ((r_0 + \rho D)/(\gamma - \rho \Phi))^{(\gamma-1)/\gamma} (\kappa_F/\Phi)^{1/\gamma} \gamma$, (eqn. (15a), replacing $r_I(x, \tau_1)$ with r_0 and using (12)). $p(x, \tau)$ is increasing in τ , so for each place x there is an interval of informality if $\tau_0 < \tau^*$, i.e. $((r_0 + \rho D)/(\gamma - \rho \Phi))^{(\gamma-1)/\gamma} (\kappa_F/\Phi)^{1/\gamma} \gamma > (r_0/(\alpha - 1))^{(\alpha-1)/\alpha} \kappa_I^{1/\alpha} \alpha/a_I$. The ratio τ_0/τ^* is smaller (an interval of informality more likely) the lower is κ_I , higher is a_I , and higher is κ_F/Φ .

A1.4 Parameters for figures: Parameter values in Figs. 1 and 2 are those derived from the Nairobi data as reported in Table 3 and Section 4.3. Units on the horizontal axes can be interpreted as years and kilometres. Date 0 on Figs. 1 and 2 correspond to 1914 in real-time, and the space-rent path underlying these figures has $p(0, 2015-1914) = 23.29$ (Table 3). The figures are constructed with $r_0 = 3.3$; this determines the date at which edge slum development starts (see proposition 2), but has no bearing on dates of formalisation or subsequent redevelopment. Setting $r_0 = 3.3$ gives first slum development at $t = -14$ in Figs 1 and 2 (real time, 1900). In Fig 2, D takes positive parts of a normally distributed random variable with mean 0 and standard deviation = 75.

A1.5 Proof of Proposition 3: Part (i) is the total differential of (20) at value $p(x, \tau_0)$ given by (14a). Part (ii) comes from differentiation of (15a) with (2), (12) and (20); it also uses the fact that Φ is independent of x and t , as given in proposition 1. Part (iii) of the proposition follows from eqn. (17) with the space-rent equation, (20), giving $\Delta x/\Delta \tau = \hat{p}/\theta$.

A1.6 Amenity and construction costs by distance from centre: 2015: Space-rent in 2015 is $p(x, 2015) = p(0, 2015) \exp(-\theta x) = 23.29 * \exp(-0.071x)$, (Table 3).

Amenity is $a_F = 1$ and for the informal sector $a_I[v_I(x, t)]^{(1-\alpha)/\alpha}$. With eqn. (4) and values from Table 3 this becomes $\alpha \kappa_I / p(x, 2015) = 6.59 * \exp(0.071x)/23.29$. Evaluated at $x = 0$ and $x = 10$ gives respective vales 0.28 and 0.58.

Construction costs per unit volume in the informal sector are $\kappa_I = 1.64$.

Construction costs per unit volume in the formal sector are $k(v_F)/v_F = \kappa_F v_F^{\gamma-1}$, with $v_F(x, \tau_i) = \left[\frac{p(x, \tau_i) \Phi}{\kappa_F \gamma} \right]^{1/(\gamma-1)}$ (eqn (10)), giving $k(v_F)/v_F = \Phi p(x, 2015)/\gamma = \Phi(23.29 * \exp(-0.071x)/1.703)$. Evaluated at $x = 0$ and at $x = 10$ this gives respective values 13.7Φ and 6.7Φ . The ratios reported in the text are $1.64/13.7=$ at the centre and $1.64/6.7$ at 10 km out.

Appendix 2: Data

This section has four components. The first discusses and describes the sources for all data used in this paper. The second deals with measures on cover/footprint and volume we use to analysis. The third gives the algorithm used to extract unchanged buildings, redeveloped buildings and infill from the overlay of 2003 and 2015 depiction of building polygons. The last reports some regression and welfare results.

A2.1 Data sources

Building data: We use two cross sections of data that delineate every building footprint in the city of Nairobi. The first is based on tracings of buildings from aerial photo images for 2003 that we received from the Nairobi City Council. Although no explicit metadata was provided, as far as we can tell this data was created by the Japan International Cooperation Agency (JICA) and the Government of the Republic of Kenya under the Japanese Government Technical Cooperation Program, and based on aerial images taken in February 2003 at a scale of 1:15,000. We base this off documentation from the Center for Sustainable Urban Development (CSUD) at Columbia University, who use a highly detailed building density and land-use map from the JICA (Williams et al. 2014). Further, we do our own data quality check by comparing the digital tracings to very high-resolution imagery from Google Earth (2002), (2003), and (2004). By examining areas that changed from 2002-2003 and from 2003-2004 we confirm that our data of building outlines matches those that exist in 2003, but did not exist in 2002, and does not include those that were yet to be built in 2003 and appeared in 2004. The second cross section comes from January 2015, when imagery at (10-20cm resolution) was recorded and digitized into building footprints by a Nairobi based company Ramani Geosystems.

The footprint data describe only the area on the ground that each building occupies while we are interested in the complete volume of each building. To address this need we supplement the 2-dimensional building data with 2015 building height data derived from LiDAR (0.3-1m resolution), also produced by Ramani Geosystems. Without direct measurements of heights in 2003, we interpolate by assigning to each building in a grid square in a sector (slum or formal) the average height of unchanged buildings in the same sector over queen neighbouring grid squares.

Slum and land use maps: We focus on a definition of slums provided IPE Global under the Kenya Informal Settlements program (KISIP). IPE mapping of informal settlements was done using satellite imagery and topographic maps. Their approach was to identify slums as “unplanned settlements” which have some aspects of low house quality, poor infrastructure, or insecure tenure. To incorporate this definition of slums into our database we created shape files by manually digitizing KISIP documentation that contain detailed maps of all identified informal settlements in Nairobi (IPE Global Private Limited and Silverwind Consultants, 2013). There remains an issue of tight delineation of slum areas, where boundaries are drawn to outline the slum areas leaving a lot of empty land residual in the formal sector that we define as the complement to slums. To offset this, we adjust the IPE slum boundaries by first classifying buildings as slum if their centre lies within the original slum boundary, and then assigning each 3m x 3m pixel of non-built land to slum if the nearest building is classified as slum, and formal otherwise.

A secondary set of maps that we use comes from the Center for Sustainable Urban Development (CSUD) at Columbia University. The CSUD maps land-use in 2003, including slums, based on a more

detailed, copyrighted, land-use map created by the JICA and the Government of Kenya under the Japanese Government Technical Cooperation Program which was published and printed by the survey of Kenya 1000 in March 2005 (Williams, et al. 2014). In principle, polygons are categorized as slums if they seemed to contain small mostly temporary buildings that are randomly distributed in high-density clusters. We use this set of slums to offer a descriptive comparison of how slums have changed on the extensive margin, but for our analysis we defer to a single definition based on IPE due to discrepancies in the definition of slum across the data sources. We also make use of the CSUD land-use map to identify areas that we remove from our formal classification. The areas that we chose to remove are listed in Appendix Table A2.4 and are areas in permanent public use.

Household Survey: In order to get estimates on slum and formal household rents we use a cross section of georeferenced household level data from the 2012 ‘Kenya: State of the Cities’ survey by the National Opinion Research Center (NORC) (Zinnes et.al. 2012). This is the first survey to record *household* rent (with detailed house and some neighbourhood characteristics) for a sample that is stratified between slum and formal areas (based on the 2009 Census) covering Nairobi. We use the survey responses for monthly household rent and total household floor space to get a measure of the monthly rent per square meter of floor space. We then convert to annual rents in 2015 USD by appreciating 8% per year for 3 years, by 12 months a year, and at an exchange rate of 100 Kenyan shillings to the US dollar. Also included in this survey were geo-coordinates taken at the time of survey, however we found these to be imprecise when compared to the location of the enumeration area (EA) that the household was recorded to reside in. We correct household coordinates if they fall outside of their EA by replacing them with the EA’s centroid coordinates.

Vacant land price listings: We also require data on land values in order to calibrate the model, for this we rely on property values that have been scraped from property24.co.ke over the period September 2014 to November 2015. This data source provides us with vacant land listings recording information on asking price and plot area and location, all of which are provided for in over 80% of the listings. Prices are listed as 1,000 Kenyan shillings per acre and we convert these to 2015 USD by 4047 m² to an acre, and 100 Kenyan shillings to a US dollar. The locations are descriptive and so we entered geo-coordinates by manually searching the addresses and location descriptions. These listings are only found in the formal sector.

SRTM elevation: Elevation and ruggedness measures used in regression tables are calculated from the Shuttle Radar Topography Mission (SRTM), a grid of 1 arc-second wide cells (or roughly 30 metres in Nairobi) published by the USGS (2005). Elevation is simply the mean of these cells in each of our 150x150m gridcells, while we measured ruggedness as the standard deviation in elevation within each 150x150 metre gridcell.

SPOT Imagery: We also use high resolution SPOT5 and SPOT6 images of Nairobi for 2004 and 2015 respectively. The raw imagery was created by Airbus Defence and Space and we used it as reference to manually trace roads and define their widths in order to come up with estimates of the extent of road coverage in both the early and late time periods. Alternative sources, like Open Streetmap, were unsuitable as they did not allow us to make the comparison across time.

A2.2 Measures of cover and volume

Our unit of analysis is 150x150m grid squares. For calculating cover within the grid square in a usage, each of these is broken into 2500 3x3m cells and use type classified by what is at the centroid of the 3m

square in each period. There are three uses: vacant land, slum area and formal. For each 150x150 square we sum across the 2500 cells to get total use of each type. Most 150x150 squares are either all slum or all formal sector. However, there are about 12% which are mixed grid squares, for which we record the cover or volume of slum and formal separately.

Having summed the total area of use of each type in 3x3 squares in each 150x150 meter square, these are averaged for 150x150m squares whose centroid falls in a narrow distance ring. That sum is then divided by the total number of 150x150 grid squares in that distance band. For volume for 2015, for each 3x3m square which is formal sector, we have the height of the building at the centroid of that square. Volume for that 3x3 square is 9 times the height in meters of the building from LiDAR data. We then sum across the grid squares occupied with formal usage for 150x150m grid squares in each distance ring and then average by the total number of 150x150 m grid squares in the ring. For 2003 we have no height data. To infer 2003 heights, we use what we think is an upper bound on height: the height of unchanged buildings, where we presume demolished buildings between 2003 and 2015 are likely to be of lower height than those which survive. To assign a height to a 3x3m square in 2003 in formal sector usage, we take the average height in 2015 of all buildings that were there in 2003 for all 3x3m formal sector unchanged buildings in the own 150x150m grids square and its 8 queen neighbours. Height is the height assigned to each 3x3m square in usage in a distance ring from the centre averaged over all such cells, to effectively get a coverage weighted average of individual building heights.

How do we measure change between 2003 and 2015? For demolition, at the 3x3m level the square is defined as demolition if its centroid is covered by a 2003 building that has become open space. Demolished coverage is lost 2003 cover; demolished volume is assessed as before using the average height of unchanged buildings in the neighbourhood. Infill is new buildings that do not overlap with any 2003 buildings; a 3x3m square is infill if its centroid is covered by such a building on 2015 where there was no building in 2003. Infill cover and volume are assessed from 2015 data. Net redevelopment in coverage takes coverage in the new 2015 buildings and subtracts the coverage of old 2003 buildings. So for each 150x150m meter square we have for redeveloped buildings, we have total coverage in 2003 measured at the 3x3m level (centroid covered by the old 2003 building(s)) and we have total coverage in 2015 measured at the 3x3m squares (centroid covered by the new replacement 2015 building(s)). Net redevelopment at the 150x150m square is the difference. In general, the same buildings are drawn in 2015 to have modestly more coverage than in 2003 so coverage change is likely to be an upper bound. Net volume change again assigns heights in 2003 to the 3x3m coverage based on neighbourhood averages for unchanged buildings and uses 2015 height information on the new buildings.

A2.3 Overlaying buildings

We match buildings across time by overlaying 2015 and 2003 building polygon data in order to track the persistency, demolition, construction and reconstruction of buildings over time. Since buildings are not identified across time our links rely on a shape matching algorithm. For each building, the algorithm determines whether it was there in the other period, or not, by comparing it with the buildings that overlap in the other time period.

This task is not straightforward, since the same building can be recorded in different ways depending on the aerial imagery used, whether building height was available, and the idiosyncrasies of the human digitizer.

Data and definitions: For 2003 we use the building dataset received from the Nairobi City Council with digitized polygons for every building, roughly 340,000 in the administrative boundary of Nairobi. For 2015 we use the dataset that was created by Ramani Geosystems using imagery (10-20cm resolution). The nomenclature we use is as follows. First, a *trace* is the collection of polygon vertices that make up its outline. A *shape* is the area enclosed by the trace, and can be thought of as a representation of the rooftop of a building. A *cavity* is an empty hole completely enclosed in a shape. A *candidate pair* is the set of any two shapes in different time periods which spatially intersect. A *link* is the relationship between a set of candidates in one period to a set of candidates in the other time period.

Pre-processing: Before running our shape matching algorithm we clean up the data sets. First we take care of no data areas. There are some areas that were not delineated in 2003, including the Moi Air Base, and Nairobi State House. We drop all buildings in these areas for both 2003 and 2015, amounting to roughly 1,500 buildings from the 2015 data, and 100 buildings from 2003. Next we deal with overlapping shapes, an issue arising in the 2015 data, although not that for 2003. This is most often the same building traced multiple times. We identify all such overlapping polygons and discard the smaller version until no overlaps remain; about 1,400 buildings from the 2015 data this way. We also drop small shapes, in part because the 2015 data has many very small shapes, while the 2003 data does not. In order to avoid complications of censoring in the 2003 data, we simply drop all shapes that have an area of less than 1m^2 . We drop 2 small buildings in 2003, and 462 small buildings in 2015.

Another issue is that buildings are often defined as contiguous shapes in 2003, but broken up in 2015. For the majority of buildings we cannot aggregate the broken up pieces in 2015 since it is hard to identify such cases in general. To match these cases across time we rely on our one to many, and many to many matching algorithms defined below. However, in the specific case where a building is completely enclosed in another the task is much easier. First, we find all cavities present in each period, then we take all building shapes that overlap with the cavities in the same time period. After identifying all shapes that intersect a cavity, we redefine both shapes, the original shape containing the cavity and the shape intersecting it, as a single new shape.

Shape Matching Algorithm: After the pre-processing of each cross-section is complete, we run our shape matching algorithm to establish links between buildings across time periods. For any given building we consider 5 possible scenarios; that it has a link to no building, that it has a link to one building (one to one match), that it has a link to multiple buildings (one to many), that it is part of a group of buildings that match to one building (many to one), or that it is a part of a group of buildings that matches to a group of buildings (many to many). We follow an approach similar to Yeom et al (2015) however, due to the inherent difficulty of inconsistent tracings we contribute to their method by introducing the one to many and many to many approaches. We assign each link a measure of fit that we call the overlay ratio. We then choose optimal links based on the overlay ratio. Finally, we categorize links as matched or not using a strict cut-off on the overlay ratio of 0.5. Other cut-offs such as 0.4, 0.6 and 0.7 produced more errors in categorization.

Candidates: For all buildings A in the first time period, and B in the second time period we identify the set of candidates:

$$CP = \{(A, B); \text{Area}(A \cap B) \neq 0\}$$

For each candidate pair we find the ratio of the intersection area over the area of each shape, so if shapes A and B intersect, we find $r_{AB} = \frac{Area(A \cap B)}{Area(A)}$ and $r_{BA} = \frac{Area(A \cap B)}{Area(B)}$. We link all shapes which do not belong to a candidate pair to the empty set.

One-to-One Matching: First we consider candidate pairs to be links on their own. For each pair, we calculate the overlay ratio as the intersection area over union area, so if A and B are candidate pair, we find:

$$R_{AB} = \frac{Area(A \cap B)}{Area(A \cup B)} = \frac{Area(A \cap B)}{Area(A) + Area(B) - Area(A \cap B)}$$

One-to-Many Matching: For each time period separately, we identify all candidate pair links for which their intersection to area ratio is above threshold θ . For shape A we define a group = $\{B; r_{BA} \geq \theta\}$. Now we calculate the overlay ratio of one to many links as the intersection area over union area ratio:

$$R_{AG} = \frac{Area(A \cap \cup_{B \in G} B)}{Area(A \cup \cup_{B \in G} B)} = \frac{\sum_{B \in G} Area(A \cap B)}{\sum_{B \in G} Area(A \cup B)}$$

Many-to-Many Matching: Here we have two cases, one when the shapes are fairly similar, which we capture in previous sections (one to one, or many to one). The other is inconsistent shapes that form the same structure. To capture these we consider both time periods at the once, we clean the candidate pair list, keeping links for which either ratio is above a threshold θ_1 :

$$LC = \{(A, B); r_{AB} \geq \theta_1 \text{ or } r_{BA} \geq \theta_1\}$$

Then we condition to only keep shape for which the total ratio intersection is above threshold θ_2 , so shape A will be included if $\sum_{B \in \{x | (A, x) \in LC\}} r_{AB} \geq \theta_2$. Now we are left with a new candidate list, which we convert to sets $LC = \{(\{A\}, \{B\})\}$ and start merging them:

$$\text{if } G_i \cap G_j \neq \emptyset \text{ or } H_i \cap H_j \neq \emptyset: LC = \{(G_i \cup G_j, H_i \cup H_j)\} \cup LC / \{(G_i, H_i), (G_j, H_j)\}, i \neq j$$

We keep doing this until we can no longer merge any two rows. At this point we calculate the overlay ratio of many to many links as the intersection area over union section ratio:

$$R_{GH} = \frac{Area(\cup_{A \in G} A \cap \cup_{B \in H} B)}{Area(\cup_{A \in G} A \cup \cup_{B \in H} B)}$$

ICP Translation: We encounter a problem when the two shapes or groups of shapes are similar but do not overlap well, this usually stems from the angle at which the images were taken, and is especially prevalent with tall buildings. To address this issue, we translate one trace towards the other, and then recalculate the overlay ratio. As in Besl and McKay (1992), we use the iterative closest point (ICP) method to estimate this translation. To perform the ICP we ignore any cavity points as we found they often cause less suitable translation. We found that for similar shapes this will optimize the intersection area.

Optimal Linking: In the end, we rank all links by their overlay ratio. We iteratively keep the link with the highest overlay ratio, or discard it if at least one of the buildings in the link has already been confirmed in a separate link. From the list of optimal links, we define a link to be a match if its overlay ratio, or the overlay ratio after ICP translation is above 0.5. We then define all matched candidates as

unchanged, and the remaining candidates as redeveloped. All buildings that were not considered as candidates are defined as infill, if from 2015, and demolished, if from 2003.

Accuracy Assessment: In order to assess the performance of the polygon matching algorithm we manually classified links between 2003 and 2015 for a random sample of buildings. We sampled 48 150x150m grid cells, stratifying over slum, non-slum within 3km, non-slum within 6km, and non-slum further than 6km to the centre. The sample consists of over 2,250 buildings in 2003 and 3,500 buildings in 2015.

Results: We first break down matches by their mapping type. There are five types of manual link: redeveloped/infill/demolished (0), one to one match (1), one to many match (2), many to one match (3), and many to many match (4). For the algorithm we further split (0) into infill/demolished (-1) and redeveloped (0). Appendix Table A2.1 shows the correspondence between the two mappings by building (a) and roof area (b). We can see that most errors come from the one to one matches, however, the many to many matches have the worst performance. Overall the diagonal values are quite high, which means not only are we matching buildings well, but also the algorithm is recognising the clumping of buildings as a human does (bear in mind that, for example, the one to one matches which we ‘misclassify’ as many to many will still be classified as match in the final data). Finally, we have perfect correspondence for demolition and in 2015 nearly perfect for infill.

Next we compare buildings that were matched by the algorithm and those matched manually. For now we use a cut-off of the overlay ratio of 0.5, later we explore the effect of different cut-offs on performance. As seen in Appendix Table A2.1 infill and demolition are classified with almost perfect correspondence. For this reason we ignore buildings with these mappings and focus on accuracy of redevelopment and unchanged. In Appendix Table A2.2 we condense mappings 1, 2, 3, and 4 into category 1, while redevelopment, or category 0, remains the same.

We define precision P (negative predictive value NPV) as the fraction of buildings classified as unchanged (redeveloped) by the algorithm that are correct, recall R (true negative rate TNR) as the fraction of buildings classified as unchanged (redeveloped) by hand that the algorithm gets correct, and the F1 score (F) as the weighted average of the two.

$$P = \frac{\text{True Positive}}{\text{Positive Predictions}}, \quad NPV = \frac{\text{True Negative}}{\text{Negative Predictions}}, \quad R = \frac{\text{True Positive}}{\text{Positive Condition}}$$

$$TNR = \frac{\text{True Negative}}{\text{Negative Condition}}, \quad F = \frac{2 * P * R}{P + R}$$

The confusion matrix in Table A2.2 is done across all sampled buildings in 2003 and weights observations by buildings (1) and roof area (2). The F1 score is high in both cases, but in part this is due to relative success classifying unchanged buildings: precision for buildings that were classified as redeveloped by the algorithm is 76% of buildings and 72% of roof area, while recall of true redeveloped buildings is 83% of buildings and 74% of roof area

In our first attempt we arbitrarily picked 50% as a cut off of the overlay ratio. Here we take a closer look at this choice. Using our manually classified links we can maximize the F1 score with respect to the cut off. In Appendix Fig. A2.1 we plot the F1 score weighted by roof area against cut-offs of the

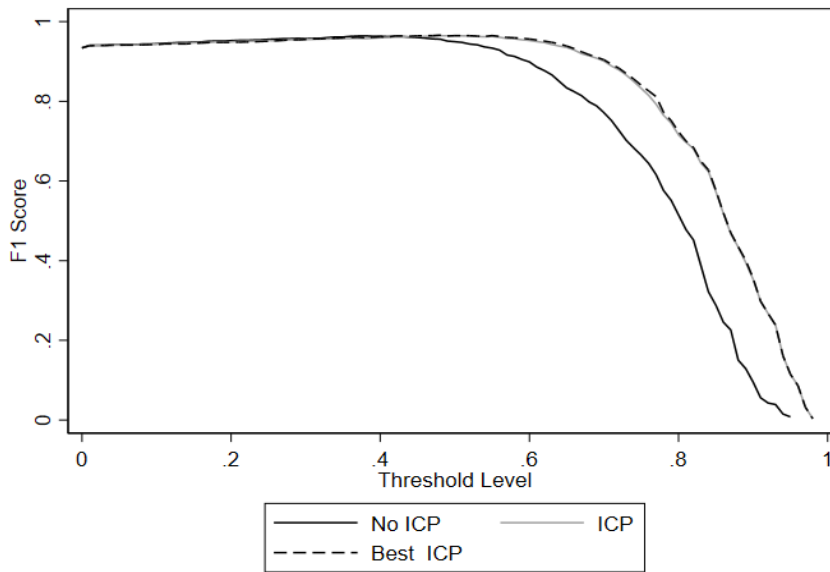
overlay ratio for the 2003 data. We find that the highest F1 score comes just below 50% suggesting our first estimate was not far off.

In Fig. A2.1 we plot lines for each method of calculating the overlay ratio: without ICP, with ICP, and the maximum of the two. Around 50% we can see that the maximum performs best, but with only a very slight improvement over the ICP alone, which is in turn marginally better than without the ICP.

Table A2.1: Mapping Correspondence 2003						
a) Weighted by Building						
	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4
Manual=0	280	413	41	24	30	10
Manual=1	0	12	729	20	3	3
Manual=2	0	10	13	315	0	6
Manual=3	0	13	3	0	145	1
Manual=4	0	82	69	57	47	84
b) Weighted by Area (m ²)						
	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4
Manual=0	12708	27194	3303	3295	1982	686
Manual=1	0	691	112262	6055	347	242
Manual=2	0	2343	1545	94617	0	515
Manual=3	0	347	202	0	15562	23
Manual=4	0	5308	5543	10978	4429	8950
Mapping definitions: -1 demolition or infill; 0 redevelopment; 1 one to one match; 2 one to many match; 3 many to one match; 4 many to many match.						

Table A2.2: Matching all areas 2003			
a) Weighted by Building			
	Algo=0	Algo=1	Recall
Manual=0	413	105	0.8
Manual=1	117	1495	0.93
Precision	0.78	0.93	F=0.93
b) Weighted by Area (m ²)			
	Algo=0	Algo=1	Recall
Manual=0	27194	9266	0.75
Manual=1	8689	261270	0.97
Precision	0.76	0.97	F=0.97

Figure A2.1: F1 Scores under ICP Routines



Notes: This figure shows average area-weighted F1 scores by threshold level of the overlay ratio, and by ICP translation routine. The solid dark line shows results using the overlay ratio determined without the ICP translation adjustment, the solid gray line uses the overlay ratio after ICP translation, and the dark dashed line uses the larger of the two overlay ratios (essentially in-line with that using ICP only).

Table A2.3: List of public uses

<p>Recreational</p> <p>a) Impala club, Kenya Harlequins, and Rugby Union of East Africa (0.14kmsq)</p> <p>b) Golf Course (0.9kmsq)</p> <p>c) Arboretum (0.25kmsq)</p> <p>d) Central park, Uhuru park, railway club, railway golf course (0.5kmsq)</p> <p>e) Nyayo stadium (0.1kmsq)</p> <p>f) City park, Simba Union, Premier Club (1.1kmsq)</p> <p>g) Barclays, Stima, KCB, Ruaraka, Utali clubs, and FOX drive in cinema (0.3kmsq)</p> <p>Undeveloped</p> <p>a) Makdara Railway Yard (1kmsq)</p> <p>b) John Michuki Memorial Park (0.1kmsq)</p> <p>Special use (includes poorly traced areas)</p> <p>a) State House</p>	<p>Public utility</p> <p>a) Dandora dump (0.5kmsq)</p> <p>b) Sewage works (0.25kmsq)</p> <p>c) Kahawa Garrison</p> <p>Public use</p> <p>a) Communications Commission of Kenya (0.1kmsq)</p> <p>b) Langata Womens prison (0.2kmsq)</p> <p>c) Nairobi and Kenyatta hospitals, Milimani Police Station, Civil Service club</p> <p>d) Mbagathi hospital, Kenya Medical Research Institute, Monalisa funeral home</p> <p>e) National museums of Kenya</p> <p>f) Kenya convention centre and railway museum</p> <p>g) Industrial area prison</p> <p>h) Mathari mental hospital, Mathare police station, traffic police, Kenya police, Ruaraka complex, and National youth service</p>
--	---

b) Ministry of State for Defence c) Forces Memorial Hospital and Administration Police Camp d) Langata Army Barracks e) Armed Forces f) Moi Airbase	i) Jamahuri show ground <i>Educational (not primary and secondary schools)</i> a) University of Nairobi and other colleges b) Kenya Institute of Highways & Built Technology c) Railway Training Institute d) Kenya Veterinary Vaccines Production Institute e) Moi Forces Academy f) NYS engineering, Kenya Institute of Monetary Studies, KCA university, KPLC training, Utali college
---	--

A2.4 Context issues: Monocentricity, and real estate data

Monocentricity

We look at a set of cities in the developing versus developed world in terms of their degree of monocentricity using a traditional indicator: steepness of the population density gradient, representing how sharply density declines from the city centre. A flatter slope indicates a lack of centrality, or force of pull of the centre. The sample cities is based upon the dataset of World Urbanization Prospects: The 2014 Revision, produced by the Population Division of the UN. The selection procedure follows three rules: (1) Select 10 cities from SSA and another 10 from Europe or USA; (2) Given the population of Nairobi is 3.7 million in the dataset, we select cities in the range from 1.5 million to 5.5 million; (3) Prioritize cities with population closer to the population in Nairobi to the sample.

Population is measured by Landscan which captures the ambient population at a 1km grid square level. We exclude grids with zero population and go out to 10km from city centre, cutting out water areas. To define the city centre, we

1. Find all light grids with light radiance readings in the top 1% ranking within the city.
2. Define light clusters based on spatial rook contiguity for just these 1%.
3. Find the largest cluster with the highest sum of light radiance.
3. Define the grid with the highest radiance in the largest cluster as the center.

We then run a regression where the dependent variable is ln of grid pop. With a 10 km radius, we can have up to about 315 sq per city. Regressions have city fixed effects and the basic explanatory variable is distance from the city centre. In Table A2.4 we show the results for developed vs developing country cities. Results in columns 1 and 2 show a distinctly steeper slope for developing country cities. In developing country cities, Nairobi with gradient slope of -0.273 has 5th steepest gradient out of 10. In developed countries only San Francisco at -0.346 is steeper.

Table A2.4: Monocentricity of population in developing and developed country cities

	Developing country cities	Developed country cities
	(1)	(2)
Distance to centre (km)	-0.330*** (0.018)	-0.138*** (0.013)
R ²	0.17	0.13
Observations	2721	2473

Notes: All observations are 1x1km grid cells. The dependent variable is always log population as measured by Landsat. The developing country cities (col 1) are Luanda, Dar es Salaam, Abidjan, Nairobi, Kano, Dakar, Addis Abba, Ibadan, Yaound, Douala. The developed (col 2) are Atlanta, DC, Phoenix/Mesa, Montreal, Rome, Detroit, San Francisco, San Diego, Athens, Lisbon. City fixed effects are included in all columns. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Real estate data and robustness

We use quarterly reports on the Nairobi housing market from HassConsult, a market leader in the Kenyan Real Estate sector.³⁸ These reports have been an important source of information for Kenyan real estate investors and homeowners for over ten years. The primary source of data is based on listings available in public sources (newspapers, magazines, social media, and online property portals). These listings are cleaned by HassConsult to remove properties and vacant land that are deemed atypical or duplicates, and then averaged in a way that ensures that price changes are not driven by composition effects.

The Hass Composite Sales and Lettings indices go back to 2000, and over its first 17 years used 163,000 cleaned listings after having dropped 15% due to duplications and 4% as outliers. Assuming that the collection rate has increased since the initial years, we can give a lower bound to the number of property listings used each year of 10,000. The Hass Land Composite Index goes back to 2007, and over its first ten years used 58,000 cleaned land listings. Assuming that the collection rate has increased since the initial years, we can give a lower bound to the number of land listings used each year of 5,800. The indices are broken down geographically into 18 Nairobi Suburbs³⁹ which is what we use and 14 Nairobi Satellite Towns⁴⁰. We geolocate these neighbourhoods by a manual google maps search. Each neighbourhood had a distinct polygon which we were able to trace and georeference neighbourhood boundaries. We use the centroid of these boundaries for any spatial analysis of the HassConsult data.

³⁸ All original reports are available in the supplementary data and were last downloaded from the HassConsult report archive on May 5th, 2020: <https://hassconsult.co.ke/real-estate/hass-index/16-report-archives>

³⁹ The HassConsult suburbs are: Donholm, Eastleigh, Gigiri, Karen, Kileleshwa, Kilimani, Kitisuru, Langata, Lavington, Loresho, Muthaiga, Nyari Estate, Parklands, Ridgeways, Spring Valley, Upperhill and Westlands.

⁴⁰ The HassConsult satellite towns are: Athi River, Juja Town, Kiambu Town, Kitengela, Limuru, Mlolongo, Ngong Town, Ongata Rongai, Ruaka, Ruiru, Syokimau, Thika Town and Tigoni.

We first looked for bubbles. Based on Glaeser, Huang, Ma and Shleifer (2017), bubbles are hard to define; but we looked at two common indicators for which there are data: overbuilding and price paths. The evidence suggests there is no overbuilding, and price increases are not high and see quite steady around the 2015 time period. First on sustained overbuilding, as noted in the Introduction, built volume increased over the 12 years by 59%, while population growth increased by 67%. Second, we investigated the rate of price increases using data from HassConsult. For looking at bubbles, researchers look at property value data, to see if rates of increases are very high or if a bubble has burst. For the full set of communities, prices rose 2.38 fold from 2007 to 2019 for an average annual increase of 7.5% (HassConsult, 2019). For years around our Q4 2015 data from the Q4 2014, Q4 2015 and Q4 2016 archived reports, the prior year's annual rate of property value changes are respectively 8.3, 9.6 and 7.6, so 2015 looks like a typical year (HassConsult; 2014, 2015a, and 2016). The official inflation rate varies a lot annually but averaged 7.4% p.a. from 2010 to 2019, with about a 6% rate around 2015. The property price changes are not indicative of any bubble; they seem steady and not high.

We also used the HassConsult reports on rental prices to look at real rates of increase in space-rents. One index uses the full sample of communities, where over time more properties in outlying lower quality housing areas are added. Rental prices for this sample of communities have risen 1.85 fold over 2007-2019, which gives a 5.3% average annual rental price increase (HassConsult, 2019). Another uses more traditional central communities without this bias.⁴¹ House rental price increases for each of these communities range from 1.96 to 3.25 fold over the period 2007-2019, or 5.8% to 10.3% average annual rental price increases (HassConsult, 2019). Inflation was averaging 7-8% p.a., so the 0.94% annual rate of real space-rent increase that we use falls between these numbers.

Finally, using our georeferenced neighbourhoods and the HassConsult report on land prices for Quarter four of 2015, we take the average listing price per square meter of vacant lots in the 18 Nairobi Suburbs mentioned above to look at land price gradients (HassConsult, 2015b). This serves as a robustness check of the analysis of property24 data for which we have microdata, but may not cover as broad of a sample as does the HassConsult data. We measure the distance from the neighbourhood centroid to the city centre. The estimated gradient from the simple regression of ln price per square meter on distance to centre in kilometres gives a similar slope -0.198 (0.0309); within one standard error of that estimated with property24 data.

⁴¹ The central neighbourhoods we choose are: Lavington, Kileleshwa, Kilimani, Westlands, and Muthaiga. From 2007 to 2019 rental prices in these areas rose by 2.69, 2.30, 1.96, 3.25, and 2.66 fold respectively, so the average annual rate of change in rent for each was 8.6%, 7.2%, 5.8%, 10.3%, and 8.5% respectively.

A2.5 Other key regressions and figures:

Table A2.5: Full key regressions in Tables 1 and 2						
	Ln land sales price (USD per m ²)	Ln Formal redeveloped height; quantile: 80th percentile	Ln Slum BVAR		Ln formal space-rent per m ³ vol. in \$2015	Ln slum space-rent per m ³ vol. in \$2015
	(1)	(2)	(3)		(4)	(5)
Distance to centre (km)	-0.172*** (0.0476)	-0.101*** (0.00521)	-0.0948*** (0.0172)	Distance to centre (km)	-0.0855*** (0.0310)	-0.00943 (0.0243)
SD Elevation (km)	-0.0114 (0.0531)	0.0222* (0.0116)	-0.0138 (0.0136)	Ln SD Elevation	0.00226 (0.144)	-0.138 (0.120)
Elevation (km)	0.00535*** (0.00178)	321.7 (197.9)	-1142.5*** (404.3)	Ln Elevation	2.145 (2.122)	6.294*** (1.204)
Lot size	-0.0302 (0.103)			No written tenancy	-0.428*** (0.133)	-0.320 (0.209)
Lot size ²	-0.00108 (0.00159)			No piped water	-0.335** (0.152)	-0.278** (0.109)
Coordinates estimated	-0.468 (0.372)			Ln # Floors	0.287** (0.108)	0.126 (0.126)
				# <i>Bathrooms</i>		
				One	-0.162 (0.144)	-0.0714 (0.125)
				Two+	0.00980 (0.220)	0.296** (0.132)
				<i>Structure type</i>		
				Single-story Shared facil.	0.0893 (0.175)	0.260** (0.116)
				Multi-storey	0.165 (0.200)	0.347** (0.133)
				Room in house		-0.603*** (0.169)
				<i>Type of walls</i>		
				Some mud	0.404* (0.215)	-0.839*** (0.211)
				Wood only	-0.370 (0.225)	-0.228 (0.234)

				Iron or tin	0.400** (0.196)	-0.375* (0.192)
Constant	-1.328 (3.234)	2.744*** (0.329)	3.163*** (0.700)	Constant	-12.46 (15.68)	-43.72*** (8.862)
Month F.E.s	Yes					
Observations	136	4621	958	Observations	361	439
R ²	0.292		0.104	R ²	0.244	0.391

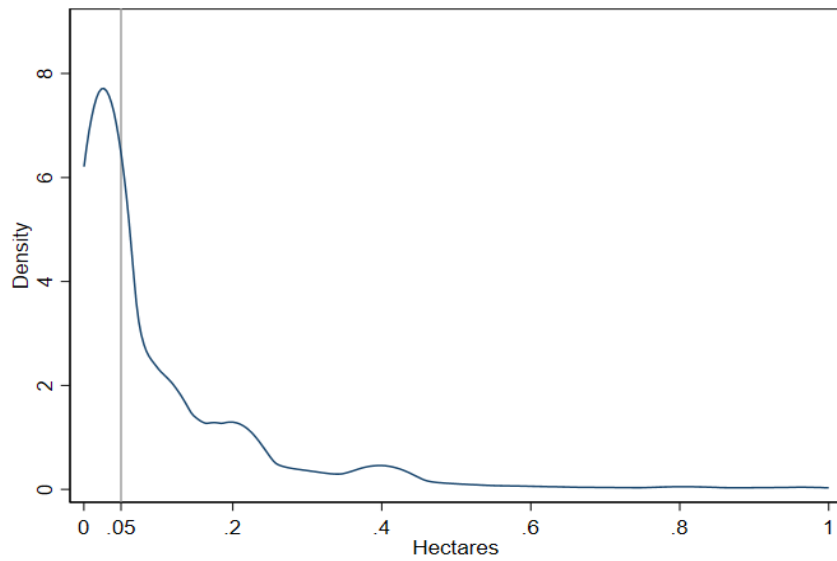
Notes: Columns 1-3 are based on 2015 data for observations inside the 2003 extent of the city. Columns 4 and 5 are based on NORC data and restricted to observations inside the 2003 extent of the city. Standard errors in parentheses. Errors in col. 1 are clustered at the neighbourhood area from the on-line listing service; in col. 2 are not clustered; in col. 3 are clustered based on a 750m x 750m grid; and in cols. 4 and 5 at the census enumeration area. In column 2, for height regressions, we only include grids for which there is cover. For the BVAR regression, column 3, about 5% of observations have no cover and hence volume (e.g. playing fields, overpasses, small parks and the like). A Tobit including these as 'censored' at 0 yields almost identical slope coefficients. Monthly fixed effects in col. 1 refer to the month of first listing. In predicting the intercept in Table 1 we assign the last 3 months a weight of 1/3 each and the rest 0. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Table A2.6: Robustness				
(1) Ln land price (\$2015 per m2)	Standard	Adding distance to industrial	New centre definition	Census slum definition
	(1)	(2)	(3)	(4)
Distance to centre	-0.172 (0.0476)	-0.187 (0.0534)	-0.17 (0.0474)	n/a n/a
Distance to industrial centre		0.046 (0.0974)		n/a n/a
'Intercept' for typical item	7.254 (0.277)	7.308 (0.279)	7.242 (0.275)	n/a n/a
Observations	136	136	136	n/a
R ²	0.183	0.177	0.183	n/a
(2) Ln for. Redevelop. height; quantile: 80th percentile				
Distance to centre	-0.101 (0.0052)	-0.107 (0.0095)	-0.101 (0.0047)	-0.09 (0.0049)
Distance to industrial centre		0.007 (0.0107)		
'Intercept' for typical item	3.315 (0.036)	3.368 (0.072)	3.298 (0.032)	3.225 (0.033)
Observations	4621	4621	4621	4133
(3) Ln slum BVAR				

Distance to centre	-0.095 (0.0172)	-0.082 (0.0242)	-0.099 (0.0181)	-0.094 (0.0205)
Distance industrial centre		-0.016 (0.0203)		
'Intercept' for typical item	1.275 (0.118)	1.162 (0.199)	1.286 (0.120)	1.12 (0.152)
Observations	958	958	958	1590
R ²	0.101	0.101	0.105	0.087
(4) Ln formal space-rent m3 in \$2015				
Distance to centre	-0.086 (0.0310)	-0.048 (0.0571)	-0.086 (0.0307)	n/a n/a
Distance to industrial centre		-0.051 (0.0750)		n/a n/a
'Intercept' for typical item	3.148 (0.071)	3.142 (0.072)	3.128 (0.071)	n/a n/a
Observations	361	361	361	n/a
R ²	0.216	0.216	0.217	n/a
(5) Ln slum space-rent m3 in \$2015				
Distance to centre	-0.009 (0.0243)	0.001 (0.0416)	-0.014 (0.0245)	n/a n/a
Distance to industrial centre		-0.013 (0.0450)		n/a n/a
'Intercept' for typical item	1.886 (0.058)	1.859 (0.087)	1.886 (0.058)	n/a n/a
Observations	439	439	439	n/a
R ²	0.371	0.37	0.371	n/a

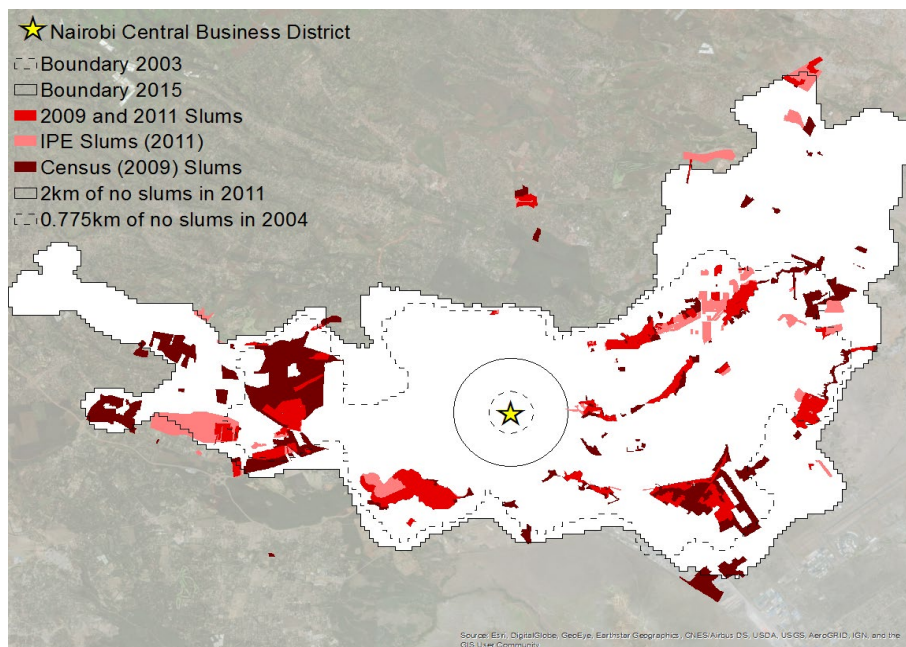
Notes: This table runs four robustness checks where each panel corresponds to a column in Table A2.5. Column 1 repeats the regressions from Table A2.5, column 2 adds a control for the distance to the industrial centre in kilometers, column 3 uses an alternative definition of the city centre based on recent nightlight values rather than 1992 values, and column 4 uses an alternative definition of slums based on the 2009 Kenyan Census unplanned enumeration areas. Panels 1-3 are based on 2015 data for observations inside the 2003 extent of the city. Panels 4 and 5 are based on NORC data and restricted to observations inside the 2003 extent of the city. Standard errors in parentheses. Errors in panel 1 are clustered at the neighbourhood area from the on-line listing service; in panel 2 are not clustered; in panel 3 are clustered based on a 750m x 750m grid; and in panels 4 and 5 at the census enumeration area. In panel 2, for height regressions, we only include grids for which there is cover.

Figure A2.2: Distribution of plot sizes: formal sector



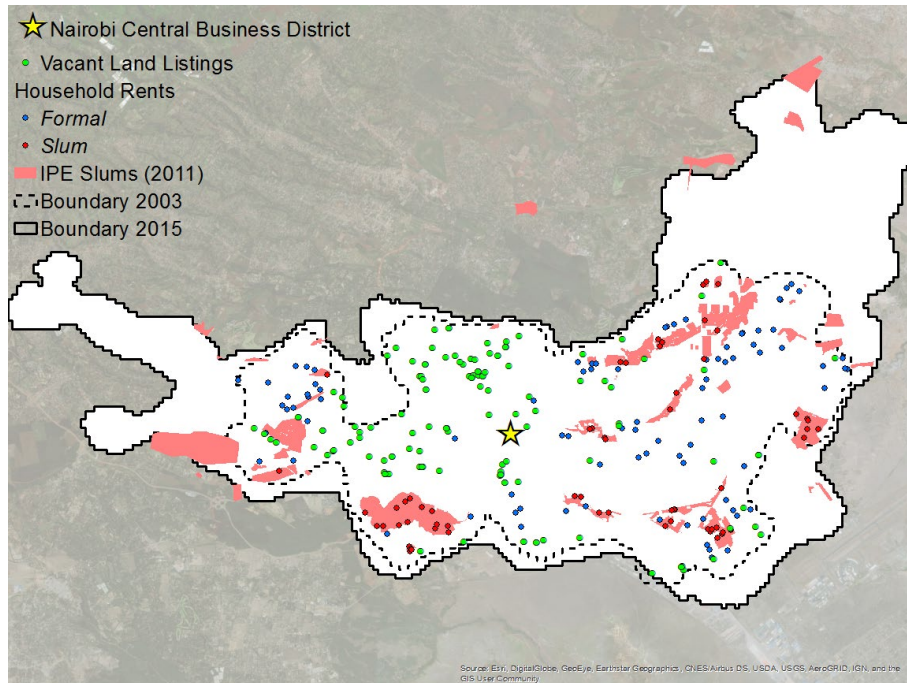
Notes: This figure shows the distribution of formal plot sizes from the cadastre. Plots above one hectare are truncated from the graph. The vertical line denotes 0.05 hectares. Of all recorded plots, 54% are smaller than 0.05 hectares, and modal density is at 0.018 hectares. Density is calculated using an Epanechnikov kernel with bandwidth of 0.025 hectares.

Figure A2.3: Census 2009 slums versus IPE slums



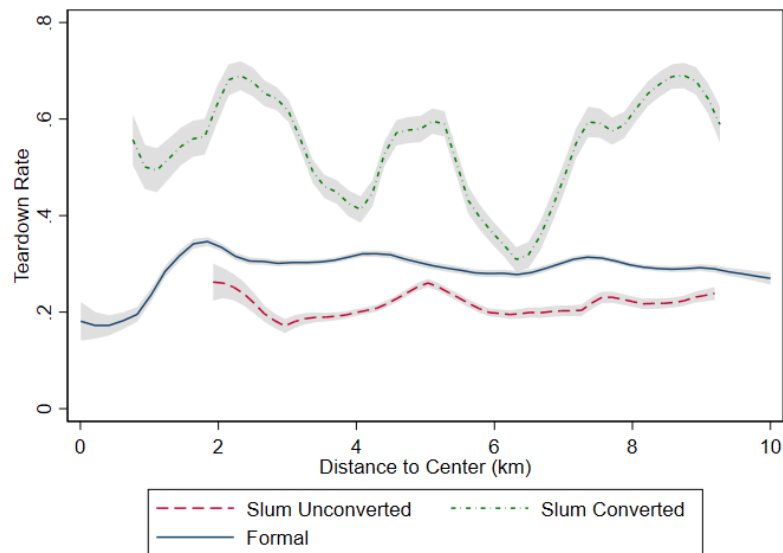
Notes: This figure maps two slum definitions; IPE from 2011 (light red), Census from 2009 (dark red), and the overlap of the two (red). The Kenyan National Census Bureau defines all Enumeration areas as either planned or unplanned, even if the are unpopulated. Our 'Census' definition here maps all Enumeration areas that are unplanned and have at least 3,000 people per km².

Figure A2.4: Location of vacant land listings and of NORC surveying on house rents



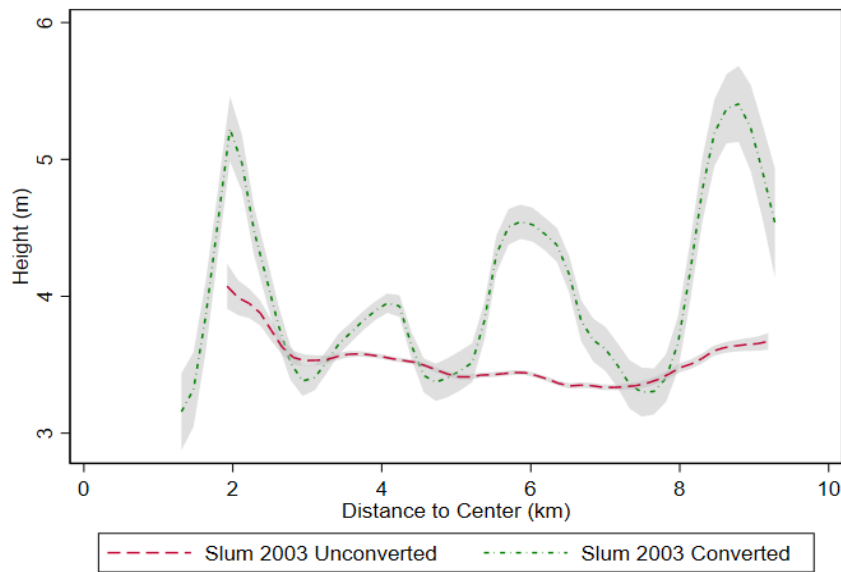
Notes: This figure maps formal (blue) and slum (red) household rents from the NORC survey and vacant land listings (green).

Figure A2.5: Teardown converted vs non-converted slums



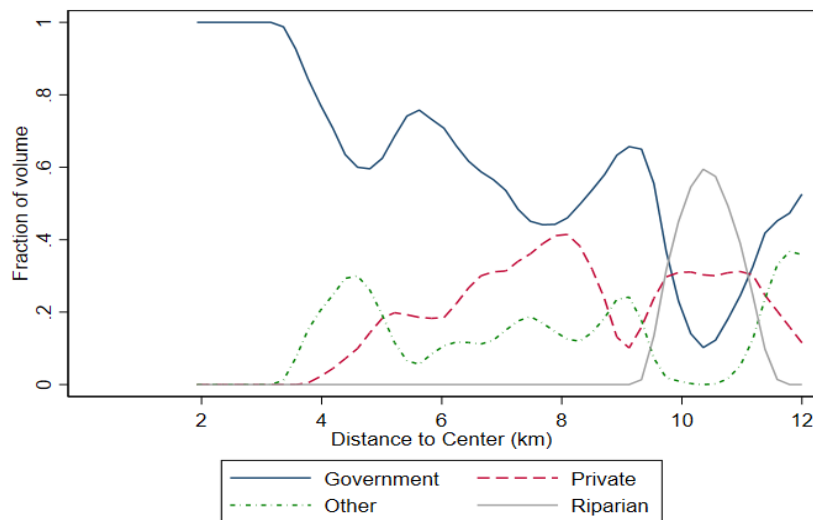
Notes: This figure shows teardown rates of 2003 buildings along distance to the city centre for the formal sector, and the 2003 slum sector. The 2003 slum sector is further broken down into unconverted slums (areas that remained classified as slum in 2011 by the IPE) and converted slums (areas that were not classified as slum in 2011 by the IPE). Teardowns are buildings that were either redeveloped or left unreplaced. The lines show the local average teardown rate, and the shaded areas show local 95% confidence intervals for this estimate. Local statistics are calculated using an Epanechnikov kernel with bandwidth of 300m.

Figure A2.6: Heights in converted vs non-converted slums



Notes: This figure shows heights of 2015 buildings along distance to the city centre for the 2003 slum sector. The 2003 slum sector is further broken down into unconverted slums (areas that remained classified as slum in 2011 by the IPE) and converted slums (areas that were not classified as slum in 2011 by the IPE). The lines show the local average building height, and the shaded areas show local 95% confidence intervals for this estimate. Local statistics are calculated using an Epanechnikov kernel with bandwidth of 300m.

Figure A2.7: Ownership of slums



Notes: This figure shows the fraction of 2015 building volume along distance to the city centre for slum land by land tenure categories. Land tenure of slums is based on IPE (2013) and author calculations on volume. Government refers to slum land where tenure rights are held by a government entity, excluding the National Housing Corporation public housing. Private refers to slum land held exclusively by private individuals or entities. Other includes land that has a mixture of private and government ownership. Riparian refers to slums along rivers. From Kenya (2018) it seems that ownership of riparian slums is a mixture of mostly private property and possessory rights, with some government owned tracts. Local shares are calculated using an Epanechnikov kernel with bandwidth of 300m.

Appendix 3: Land Quality.

In order to address concerns about heterogeneity in land quality driving some of our empirical results this appendix examines the role of geographical characteristics. Our underlying data has a fine resolution, based on 30m x 30m cells. We analyse how geography varies over small spatial scales for slum boundaries and infill, or greenfield developments. Cells are defined as exclusively slum or formal based on where their centroid lies. Elevation is measured using the SRTM data at the 30x30m cell, and its mean and standard deviation are calculated over moving windows of 90x90m, 150x150m, and 450x450m (USGS, 2005) around the own cell to measure ruggedness at different scales. To determine the relative local elevation we calculate the difference between own elevation and mean elevation in the 150x150m and 450x450m windows. Similarly using the CSUD land-use map we digitize water bodies (rivers, lakes, ponds, etc.) and distinguish whether a cell contains water in its own cell or 90x90m window (Williams et al., 2014). Further, cells are classified as infill if the cell contained only infilled buildings in 2015 and no buildings in 2003. With a range of window sizes, we focus on the ‘small’ 90x90m and the ‘large’ which is taken as 450x450m except in order to avoid excessive overlap when considering boundary analysis where we use 150x150m.

First we consider the role of land heterogeneity in explaining infill in areas near to the city centre, which according to the model should have been developed before those further out. In Table A3.1, for all formal sector land, we show that infill occurs on lower quality land, and this differential is largest near the centre. Near the centre, infill tends to occur on land with higher elevation, nearer to water, and that is more rugged. Infill also tends to be on land that is lower than other cells in its 450x450m window (effect dissipating over distance), however the differential gradient is insignificant and so only suggestive. Together this suggests that infill occurs in places that are generally higher up, but locally downhill. Proximity to water and ruggedness are more easily interpreted as raising costs to development and here we find strong and significant results.

Next in Table A3.2, we look at whether heights of redeveloped buildings vary with land quality. If during the initial formal development, a sunk cost is paid to prepare the land; draining swamps, levelling the land, etc., then during successive periods of redevelopment the height of buildings should not vary at a given distance. In the table we show that only one of our five land quality measures significantly affect redeveloped building heights. There the standard deviation of elevation in the small scale (90x90m) is associated with higher redeveloped buildings, rather than reduced height.

Finally, we are concerned that results in our welfare analysis may be driven by land quality. That is there may be a correlation between government owned slum land (where formalisation costs are high) and low quality land (where construction costs are high for natural reasons). In particular, if central slum land was on worse land than neighbouring greenfield formal this could be partially responsible for the gap in land rents. We focus on infill land because, over the past period, it is the relevant comparison for what could have been slum redevelopment. In Table A3.3 we look at the sample of cells that are within 300m of a government slum boundary, restricted to either government slum cells or formal infilled land. We run a fixed effects regression so the analysis focuses on variation within arbitrary 300x300m blocks. Within these neighbourhoods we compare cells in and outside of slums at distance bins from the city centre for different land quality outcomes. Results show that, especially for slums inside of 6km, where we focus our welfare analysis, there is very little difference in land quality between slums and neighbouring formal land. Inside 6km slums are only found to be nearer to water in the bin from 4-5km, and in the bin 2-3km slums are actually less rugged and on higher local land in the

450x450m window. So for the 20 possible coefficients there is only one suggesting significantly lower quality in slums compared to neighbouring formal areas.

Table A3.1: Quality of infilled land					
	Elevation (m)	Water indicator 90x90m	S.D. elevation in 90x90m	S.D. elevation in 450x450m	Elevation difference to mean of 450x450m
	(1)	(2)	(3)	(4)	(5)
Distance to Center (km)	-0.764 (1.169)	0.000449 (0.00162)	-0.0131 (0.0137)	-0.0239 (0.0452)	0.0141*** (0.00470)
Infilled =1	35.24*** (5.134)	0.0971*** (0.0217)	0.421*** (0.106)	1.392*** (0.280)	-0.827*** (0.103)
Distance to Center (km) * Infilled =1	-7.552*** (0.220)	-0.00678*** (0.00112)	-0.0386*** (0.00576)	-0.207*** (0.0107)	0.0224 (0.0141)
Constant	1687.7*** (4.734)	0.0566*** (0.00962)	1.717*** (0.0728)	4.721*** (0.233)	0.0303 (0.0261)
Observations	120299	120299	120299	120299	120299
R ²	0.015	0.005	0.002	0.004	0.003

Note: All columns are based on 30x30m grid observations on formal land inside the 2003 boundary and within 10km of the city centre. Standard errors in parentheses are robust and clustered based on a 450x450m grid. The dependent variables are; the absolute elevation in meters above sea level (col 1); an indicator of the presence of rivers, streams, or lakes inside the surrounding 90x90m block (col 2); the standard deviation in elevation across cells in the surrounding 90x90m block (col 3); the standard deviation in elevation across cells in the surrounding 450x450m block (col 4); and the relative elevation compared to the mean of cells in the surrounding 450x450m block in meters (col 5). 'Infilled' indicates cells that are infill in 2015 and greenfield in 2003, i.e. the building did not previously exist and has not replaced another building. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Table A3.2: Redeveloped height and land quality

	Ln formal redeveloped height
Distance to center (km)	-0.0660*** (0.00796)
Ln elevation (m)	-0.000137 (0.000265)
Water indicator 90x90m	-0.00499 (0.0547)
S.D. elevation in 90x90m	0.0266*** (0.00682)
S.D. elevation in 450x450m	0.00862 (0.00697)
Elevation difference to mean of 450x450m	0.00410 (0.00288)
Constant	2.331*** (0.439)
Observations	33701
R ²	0.053

Note: This table based on 30x30m grid observations on formal redeveloped land inside the 2003 boundary and within 10km of the city centre. Standard errors in parentheses are robust and clustered based on a 450x450m grid. The dependent variable is the log of redeveloped building height. The rows are; distance to the city centre in kilometres; the absolute elevation in meters above sea level; an indicator of the presence of rivers, streams, or lakes inside the surrounding 90x90m block; the standard deviation in elevation across cells in the surrounding 90x90m block; the standard deviation in elevation across cells in the surrounding 450x450m block; and the relative elevation compared to the mean of cells in the surrounding 450x450m block in meters. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Table A3.3: Comparison of slum land on one side of border to formal sector infill land on the other

	Elevation (m)	Water indicator 90x90m	S.D. elevation in 90x90m	S.D. elevation in 150x150m	Elevation difference to mean of 150x150m
Distance to Center=3-4	0.806* (0.474)	-0.0866 (0.0933)	0.468*** (0.169)	0.397*** (0.0899)	0.145 (0.498)
Distance to Center=4-5	0.487 (1.919)	-0.420** (0.207)	0.439 (0.626)	-0.117 (0.701)	0.667 (0.624)
Distance to Center=5-6	1.978 (2.902)	-0.249 (0.231)	0.458 (0.725)	0.504 (0.799)	-0.102 (0.725)
Distance to Center=6-7	0.236 (3.393)	0.201 (0.300)	0.935 (0.683)	0.999 (0.723)	-0.0776 (0.732)
Distance to Center=7-8	-2.433 (3.297)	0.0997 (0.298)	1.645** (0.729)	2.403*** (0.837)	0.533 (0.798)
Distance to Center=8-9	-1.291 (3.748)	0.0558 (0.314)	2.308** (0.937)	3.038*** (1.036)	0.306 (0.860)
Distance to Center=9-10	-2.200 (3.747)	0.0562 (0.310)	1.825** (0.844)	2.298** (1.025)	0.0277 (0.907)
2-3 # Slum in 2015=1	-0.449 (1.353)	0.161 (0.149)	-0.0351 (0.129)	-0.261** (0.116)	0.725* (0.372)
3-4 # Slum in 2015=1	-0.799 (1.237)	0.0844 (0.0828)	0.0265 (0.301)	-0.125 (0.296)	0.0183 (0.217)
4-5 # Slum in 2015=1	-0.468 (0.603)	0.291*** (0.0819)	-0.144 (0.449)	0.321 (0.554)	-0.157 (0.312)
5-6 # Slum in 2015=1	-0.561 (1.434)	0.0597 (0.0738)	-0.0597 (0.320)	-0.183 (0.426)	0.315 (0.280)
6-7 # Slum in 2015=1	0.211 (1.120)	-0.194*** (0.0666)	0.326** (0.132)	0.408*** (0.153)	0.510*** (0.170)
7-8 # Slum in 2015=1	0.525 (0.561)	-0.0246 (0.0720)	-0.103 (0.207)	-0.337 (0.405)	-0.0156 (0.181)

8-9 # Slum in 2015=1	-1.215 (1.688)	0.107 (0.0729)	-0.696 (0.589)	-0.903 (0.593)	0.330 (0.351)
9-10 # Slum in 2015=1	-0.317 (1.604)	-0.0542 (0.0452)	-0.157 (0.286)	-0.136 (0.369)	0.638 (0.399)
Observations	10167	10167	10167	10167	10167
R ²	0.995	0.416	0.369	0.584	0.081

Note: All columns are based on 30x30m grid observations on land inside the 2003 boundary and within 10km of the city centre. Further, observations are restricted to be within 300m of a government owned slum boundary, and for formal land only include cells that are infill in 2015 and greenfield in 2003. Standard errors in parentheses are robust and clustered based on a 450x450m grid. The dependent variables are; the absolute elevation in meters above sea level (col 1); an indicator of the presence of rivers, streams, or lakes inside the surrounding 90x90m block (col 2); the standard deviation in elevation across cells in the surrounding 90x90m block (col 3); the standard deviation in elevation across cells in the surrounding 150x150m block (col 4); and the relative elevation compared to the mean of cells in the surrounding 150x150m block in meters (col 5). Rows show estimates of the mean for each kilometre bin from the city centre on both the formal and slum sides of the boundary. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Appendix References

- Besl, P., and Mckay, N. 1992. "A Method for Registration of 3-D Shapes." IEEE Transactions on Pattern Analysis and Machine Intelligence. 14(2), 239—256
- Glaeser, E, W. Huang, Y. Ma, A. Shleifer. 2017. "[A Real Estate Boom with Chinese Characteristics](#)," Journal of Economic Perspectives, 31(1), pages 93-116
- Google Earth V 7.3.0.3830. (February 2, 2002). Nairobi, Kenya. 1° 18' 19.09"S, 36° 49' 16.26"W, Eye alt 4.61km. DigitalGlobe 2017. <http://www.earth.google.com> [August 18, 2017].
- Google Earth V 7.3.0.3830. (December 15, 2003). Nairobi, Kenya. 1° 18' 19.09"S, 36° 49' 16.26"W, Eye alt 4.61km. DigitalGlobe 2017. <http://www.earth.google.com> [August 18, 2017].
- Google Earth V 7.3.0.3830. (February 14, 2004). Nairobi, Kenya. 1° 18' 19.09"S, 36° 49' 16.26"W, Eye alt 4.61km. DigitalGlobe 2017. <http://www.earth.google.com> [August 18, 2017].
- HassConsult, 2014. "House Price Index Quarter Four Report 2014". Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].
- HassConsult, 2015a. "House Price Index Quarter Four Report 2015". Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].
- HassConsult, 2015b. "Land Price Index Quarter Four Report 2015". Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].
- HassConsult, 2016. "House Price Index Quarter Four Report 2016". Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].
- HassConsult, 2019. "House Price Index Quarter Four Report 2019". Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

IPE Global Private Limited and Silverwind Consultants, 2013. "Consultancy Services for City/Municipal Situational Analysis of Conditions of Informal Settlements in 15 Municipalities." Prepared for the Government of Kenya, Ministry of Lands, Housing and Development.

Sarah Williams, Elizabeth Marcello & Jacqueline M. Klopp, 2014. "Toward Open Source Kenya: Creating and Sharing a GIS Database of Nairobi," *Annals of the Association of American Geographers*, 104:1, 114-130.

USGS (2005), Shuttle Radar Topography Mission, 1 Arc Second scene SRTM_f03_s002e036, Finished Version 3.0, Global Land Cover Facility, University of Maryland, College Park, Maryland.

Yeom, J., Kim, Y., and J. Lee, 2015. "Hierarchical ICP Matching to Enable Interoperability of Multisource Building GIS Data." *Advanced Science and Technology Letters*, Vol. 89 (Architecture and Civil Engineering) pp. 55-60.

Zinnes, C., et al., 2012. Kenya Urban Program Baseline Study (NORC, University of Chicago, Chicago); www.norc.org/Research/Projects/Pages/kenya-urban-program-baseline-study.aspx