7/11/2018

Building the city: urban transition and institutional frictions

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

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. 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. Institutional frictions may hinder conversion of slums to formal usage. Using unique data on Nairobi for 2003 and 2015 we develop a novel set of facts and calculate welfare losses of such frictions. In older slums, even after buying out slumlords, formalisation yields gains amounting to about $17-18,000 per slum household, 25-30 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 Ilia Samsonov and Piero Montebruno. Thanks to seminar participants at LSE, CURE, Berkeley, Pennsylvania, Lausanne, Oxford, Helsinki, Luxembourg, Bristol, USC, NBER and RIETI Tokyo.

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

1. Introduction

In Sub-Saharan Africa the populations of many large cities are increasing by about 50% every ten years. This, along with growing incomes, implies a huge demand for increased building volumes in cities. World Bank (2006) suggests that about two thirds of a country's non-governmental capital stock is in buildings, and urban construction and maintenance are a 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 spread of new slums and densification in existing slums. 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 addresses these issues by doing four novel things. First, we develop a model of the built environment of a growing city with both formal sector and slum housing and institutional and geographic frictions. While Braid (2001) has a dynamic monocentric model with durable capital, no dynamic model deals with slums and frictions, and no work estimates relevant parameters. Second, we assemble a rich data set of the built fabric of Nairobi. While there is work on the USA using census data of building ages (Brueckner and Rosenthal 2009), we know of no work which utilizes city-wide data on individual buildings, with demolition, redevelopment, and infill to detail the changes in the urban landscape. Third, we deliver a new set of important facts about development and redevelopment of the built environment for an arguably representative major developing country city. Finally, we develop a methodology and parameter estimates to calculate the welfare cost of institutional frictions that delay formalisation. This highlights the role of policy for fast growing cities with major market land failures that deter investment.

The first stage is development of a dynamic model of a growing city. Urban expansion involves growing land area, intensified land use within the city, and increasing building heights. Our model captures these features, with a key distinction between formal and informal, or slum sectors. Formal buildings involve sunk capital costs, can be built tall and cannot be modified once constructed, with investment decisions based on expected future rents. As the city and housing and land prices grow, formal sector buildings are periodically demolished and redeveloped to a greater height. In contrast, informal sector (or slum) buildings are made from malleable, non-weight-bearing materials as we will see in the data. 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 already high cover-to-area ratios. New slums appear near the expanding city edge and within the city there is conversion of slums to formal development.

We add frictions to this model, with a focus on explaining why slums persist on potentially extremely high value land near the city centre. First, these older slums may be on land with very troubled property rights, which would require what we call ‘formalisation costs’ to correct. Second older slums may be on land of poor quality. We will incorporate both aspects into the modelling and empirical work, and show that both inhibit and delay formal sector development. We will try to distinguish when and where each aspect matters, suggesting formalisation costs play the key role for older slums near the centre.

The second element of our study is 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.[1] We develop an algorithm to overlay the 2003 polygons with those for 2015 to determine which building footprints are unchanged since 2003, which buildings were demolished and/or redeveloped and where and to what extent infill occurs. Each building in 2015 has a known height measured from LiDAR data. While there are no data on 2003 heights, we infer them from heights of 2015 buildings and the patterns of demolition and redevelopment. We also have high resolution SPOT satellite images of Nairobi for 2004 and 2015, which we use to visually assess areas. Finally, we have formal sector vacant land prices for 2015 and housing rent data for the formal and informal sectors in 2012. More details on the datasets and our data methodology are in the Data Appendix.

With these data we can construct a rich picture of the built environment of Nairobi, both in the spatial cross-section and its evolution through time. The third element is to derive a set of entirely new facts about the evolution of the city and to estimate key parameters of the model. The final element is to use these estimated parameters and price information to calculate what the real income losses are from delayed conversion of older slums to formal sector buildings.

What are the key findings from these two elements of the paper? First we note that Nairobi, while not circular, appears to be primarily ‘monocentric’ with building height, volume, and land prices all declining with distance from the centre. Nairobi’s monocentricity is in line with other African cities such as Addis Ababa, Dar es Salaam, and Kigali, and in contrast to Paris, London, Barcelona, or Atlanta (Lall, Henderson, and Venables 2017, p 21-23). In Fig. 1 we show this for 2015 height with a 3-D map of the city for the 2003 built area, or core of the city. In Section 3, we focus on this core area, noting that the area of the city expands by about 50% to the 2015 built area boundary, mapped later. Nairobi’s sausage shape in both time periods arises from being bounded to the south by an airport and a national park and to the immediate north of the centre by a state forest. The highest buildings are in the city centre (defined to be the brightest lit pixel in night lights data in the early 1990’s). The map gives the average height of all buildings in public or private use in each 150m x 150m grid

[1] The images define features at no more than 40 cm resolution which are then mapped to 3m x 3m cells and aggregated to a grid of 150m x 150m for ease of analysis. For the 2003 built area of the city there are 6470 such grid cells.

square.[2] The city appears monocentric with tall but variable height at the centre, which then diminishes with distance from the centre. Slum areas in red are based on a mapping described later, and generally have low height. In the far north-east the map reveals modest misclassification problems; SPOT satellite images indicate that some tall red areas do not seem to be slums.

Today the greater metropolitan area has a population in excess of 6.5mn. Based on the 1999 and 2009 censuses population is growing at about 4.3% p.a. for the 2015 built area of the city. The 29% share of the population living in slums was just slightly smaller between the two censuses. Our estimates indicate that the built volume of the 2015 city increased at 3.9% pa, expanding by 59% between 2003 and 2015. Within just the 2003 core, overall volume growth was an astounding 48%. Volume growth within the 2003 core accounts for 65% of total volume growth within the 2015 city, so that fringe growth, while at a high rate, is not dominant. Growth within the core involves mostly formal sector net urban redevelopment. For example, at 3kms from the centre net development increased volume by 35% (with infill another 18%), with 35% of buildings being torn down over 12 years and reconstruction increasing building heights by about threefold.

In Nairobi, at otherwise similar locations, slums and the formal sector have similar intensities of land uses measured by built volume to area ratios, but accomplished very differently. Given their low height, slums deliver volume with high built cover to area and little green space. The formal sector has typically half the cover to area ratio but builds high. At the city centre, grid squares *average* 10 stories with 5% of grid squares over 16 stories.

Slum volume growth in the 2015 city at 55% was modestly slower than the formal sector at 60%. New slums appear on the city edge often on land with full private property rights. However, the conversion of older slums nearer the city centre to formal sector usage is very limited, hindered by unresolved property right issues. As a consequence, Nairobi exhibits the oft-photographed hotchpotch of land uses in Africa or parts of Asia, with tall formal sector buildings bordering on pockets of single story corrugated iron sheet slum housing in the urban core. In Fig.1, Kibera the large slum area to the southwest of the centre epitomises these problems and in Section 4 we will give a brief history of the evolution of land rights in Nairobi and why Kibera and other older slums present thorny political and institutional problems. For these older slums at 3-5kms from the centre, we estimate that, even *after* paying illegal slumlords operating in these slums for the value of their land in perpetual slum use, formalisation would bring a gain of about $17-18,000 per slum household, in a context where slum households spend on average about $500-700 pa on housing. We estimate a

[2] Calculations are discussed below and details are in the Appendix. Land is assigned to informal and formal usage by where the centroid of the grid square lies. Blank areas are those which have censored data in 2003 (e.g. the Moi airbase) and large areas that have no cover (e.g. the Royal Nairobi golf course).

lower bound on formalisation costs which is very high, typically about the value of land held in slum use in perpetuity.

As a final note, in Nairobi, similar to many Africa cities,[3] the vast majority of all residents and 90% of slum dwellers rent. In some work (e.g. Cavalcanti and da Mata 2017 on Brazil) slum dwellers are 'owner occupiers' with insecure land tenure who make a choice between being in a regulated formal sector and an unregulated informal sector with insecure land rights. But the issues are not fundamentally different. There, formal sector redevelopment of high value slum lands near the centre is inhibited by high formalization costs: the cost to squatters of securing more formal title in order to sell their lands.[4]

In the paper, the basic model and core theoretical results are set out in section 2. Section 3 presents data, estimates key empirical relationships to back out parameters of the model and analyses the evolution of the built environment. Section 4 develops the welfare analysis to derive costs of misallocation of land to older slums. Section 5 concludes.

2. Theory

In this section we develop the key aspects of a model a growing city, focusing on investment decisions and consequent patterns of land-use and building volume. The analysis assumes that house-rent increases at an exogenous rate through time, and in the Theory Appendix we show how this can be endogenised in an open city equilibrium. Section 2.1 analyses building decisions associated with the slum and formal sector technologies. Section 2.2 focuses on a particular point in the city and examines 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 2.3 discusses how this path varies across points in a stylized city, giving a description of both the cross-section of the city and its evolution through time. Section 2.4 analyses the effects of frictions and spatial heterogeneity in formalisation costs and terrain affecting construction costs. Together, these analytical sections provide the relationships upon which empirics, estimation of key parameters and welfare analysis are based.

2.1 Building technology and housing supply

[3] Addis Abba in Ethiopia, Kisumu in Kenya and Kumasi in Ghana have 60, 82 and 57% renters overall respectively in the early 2000's (UN, 2011), while from recent World Bank LSMS data, Kampala is at 58%.

[4] Formalisation costs can be broadened to include things like the housing regulatory and property tax costs of being in the formal sector versus slums (Cavalcanti and da Mata, 2017), although these seem to be less relevant in Nairobi. But for Nairobi, while there are market responsive land use regulations on the books (Mwaura, 2006), penalties for non-compliance seem limited; and property taxation of formal sector lands has limited implementation (Kelly, 2003).

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, the proportion of land covered by building footprint. 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, $v_i(x,t) = h_i(x,t)c_i(x,t)$, $i = I, F$. For now, *x* denotes any location, but in Section 2.3 will become distance from the city centre.

2.1.1 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 increase the proportion of each unit of land that is covered with buildings, so $v_I(x,t) = c_I(x,t)$. 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 has the effect of reducing the quality of housing. We capture this by supposing that the observed rent (and willingness to pay) for a unit of informal housing is the product of two elements, rent at unit quality and a quality adjustment. The rent of a unit of housing of unit quality, in both the formal and informal sectors, at place *x* and date *t* is denoted *p*(*x, t*). The observed rent for a unit of informal sector housing is $p(x,t)\,a\left(v_I(x,t)\right)$, the normalised rent adjusted for informal quality. $a\left(v_I(x,t)\right) \le 1$, which is diminishing and convex in crowding (as measured by volume = cover per unit land). With this, land-rent (i.e. house-rent minus construction cost times volume per unit land), is

$$r_I(x,t) = \left[\, p(x,t)a\left(v_I(x,t)\right) - \kappa_I \,\right] v_I(x,t). \tag{1}$$

The volume of housing supplied is chosen to maximise land-rent, taking *p*(*x, t*) as exogenous and internalising the effect of crowding on quality.[5] The first order condition equates marginal revenue to marginal cost,

$$\partial r_I(x,t)/\partial v_I(x,t) = p(x,t)a\left(v_I(x,t)\right)\left\lfloor 1 + v_I(x,t)a'\left(v_I(x,t)\right)/a\left(v_I(x,t)\right)\right\rfloor - \kappa_I = 0. \tag{2}$$

In most of what follows we assume that informal house quality is iso-elastic in cover,

$$a(v_I(x,t)) = a_I v_I(x,t)^{(1-\alpha)/\alpha}, \quad \alpha > 1. \tag{3}$$

[5] A simplification is that we do not explicitly model this as an externality: one developer's choice of crowding affects neighbors.

Several implications follow. First, the optimally chosen volume of housing at *x* is

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

The degree of diseconomy in crowding, α, is a key parameter. It will be recovered in the empirics based on how informal sector volume in (4) varies with house price. Second, informal sector house-rent adjusted for quality is constant throughout the city so that $p(x,t)a\left(v_I(x,t)\right) = \alpha\kappa_I$. Essentially, and as we will see later in the house-rent data, increased crowding near the city centre will offset the advantage of improved access to the centre. Third, iso-elasticity means that maximised land-rent is share (1-1/α) of house-rent per unit land, and construction costs are share 1/α. The expression for maximised land-rent which we will use in Sections 3 and 4 comes from substituting (3) and (4) into (1) giving,

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

2.1.2. Formal sector

The formal sector differs in a number of respects. First, buildings are 'putty-clay', malleable at the date of construction but not thereafter. For simplicity, but also based on the data in the cross section, we assume that formal sector land cover is uniform throughout the city at $c_F = 1$, and that volume is achieved by choice of height. Height is chosen at date of construction, denoted τ_i, so $v_F(x,\tau_i) = h_F(x,\tau_i)$, fixed for the life of the structure, i.e. until demolition at date τ_{i+1}, where subscript *i* = 1,2... is used to denote successive redevelopments of formal structures. Construction costs per unit land are one-off and sunk, and are an increasing and convex function of height (= building volume) on that land, $k(v_F(x,\tau_i),\ k', k'' > 0$. Demolition incurs neither costs nor benefits, as materials cannot be recycled back to putty.

This sunk cost of construction differs fundamentally from the flow cost in the slum sector, and we think captures key differences in construction technology. In Nairobi, from the 2009 Census, formal and slum sector wall materials are distinctly different. In slums, the majority (about 55%) of housing walls are corrugated iron sheets which can be easily reconfigured like Meccano parts; most other slum housing involves mud construction (about 20%) and other material with short duration. Neither of these materials are sufficiently load bearing to allow much in the way of height; 85% of slum buildings are under 5 meters tall. In contrast, over 90% of formal sector housing is made of stone or some type of brick/block, many with substantial height.[6] We note that we model slums as a technology choice, not an issue of land rights per se (c.f., Cavalcanti and da Mata, 2017), although the two can be interrelated. However, about 22% of slum house volume in 2015 is on private land, near the city edge.

We assume there is no amenity loss or gain from building tall so the house-rent of a unit of formal sector building volume, $p(x, t)$, is 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 costs $k\left(v_F(x,\tau_i)\right)$ sunk and volume fixed at the date of construction this is given by

$$R_F(x,\tau_i) = \int_{\tau_i}^{\tau_{i+1}} p(x,t) v_F(x,\tau_i) e^{-\rho(t-\tau_i)} dt - k\left(v_F(x,\tau_i)\right). \tag{6}$$

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

$$\Phi(x,i) \equiv \int_{\tau_i}^{\tau_{i+1}} \left[p(x,t) / p(x,\tau_i) \right] e^{-\rho(t-\tau_i)} dt \ , \tag{7}$$

so $R_F(x,\tau_i) = p(x,\tau_i)\Phi(x,i) v_F(x,\tau_i) - k\left(v_F(x,\tau_i)\right)$. 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 (7) is cut at τ_{i+1}).

The first order condition for choice of volume at place x and development date τ_i is

$$\partial R_F(x,\tau_i) / \partial v_F(x,\tau_i) = p(x,\tau_i)\Phi(x,i) - k'\left(v_F(x,\tau_i)\right) = 0 \ . \tag{8}$$

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[7]

$$v_F(x,\tau_i) = \left[\frac{p(x,\tau_i)\Phi(x,i)}{\kappa_F \gamma} \right]^{\frac{1}{\gamma-1}} \ , \ R_F(x,\tau_i) = \kappa_F(\gamma-1)\left[\frac{p(x,\tau_i)\Phi(x,i)}{\kappa_F \gamma} \right]^{\frac{\gamma}{\gamma-1}} . \tag{9}$$

The diseconomy in building taller, γ, is another key parameter which will be estimated based on how the height of newly constructed buildings and hence volume in (9) varies with location as house rent varies.

[6] Some on-going studies classify slums by the use of corrugated iron for roofs. This would not work in Nairobi; 50% [85%] of formal [slum] sector residential buildings have corrugated iron sheet roofs.

[7] 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).

As well as the present value of land-rent at intervals of τ_i, 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 amortization is 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 $\mu p(x,\tau_i)\Phi(x,i)v_F(x,\tau_i) = k\left(v_F(x,\tau_i)\right)$. With iso-elastic cost function and (9) the amortization rate is then $\mu = 1/\gamma$.[8] Flow land-rent net of amortisation is therefore fraction (1-1/γ) 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

$$r_F(x,t,\tau_i) = (1-1/\gamma)\,p(x,t)v(x,\tau_i) = (1-1/\gamma)\,p(x,t)\left[\frac{p(x,\tau_i)\Phi(x,i)}{\kappa_F\gamma}\right]^{\frac{1}{\gamma-1}}. \qquad (10)$$

Notice that land-rent is fraction (1-1/γ) of house-rent, as in the informal sector land-rent where the fraction (1-1/α) in (eqn. 5). We will use these relationships in Sections 3 and 4. Notice also that, comparing (9) and (10) at date $t = \tau_i$,

$$r_F(x,\tau_i,\tau_i) = R_F(x,\tau_i)/\Phi(x,i). \qquad (10a)$$

2.2. Land development and construction phases

Continuing to focus on a particular unit of land, *x*, we now look at the choices of when to develop informal structures and when to develop or redevelop formal structures. At some date (say time 0) the present value of land-rent at location *x* that has not yet been developed is

$$PV(x) = \int_0^{\tau_0} r_0 e^{-\rho t}dt + \int_{\tau_0}^{\tau_1} r_I(x,t)e^{-\rho t}dt + \left[R_F(x,\tau_1) - D\right]e^{-\rho\tau_1} + \sum\nolimits_{i=2} R_F(x,\tau_i)e^{-\rho\tau_i}. \qquad (11)$$

The first term is the present value of rent from 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 gives the present value of rent from informally developed land during interval τ_0, τ_1. The first formal sector development, occurring at date τ_1 yields rent and incurs a potential one-time fixed cost *D* of overcoming any remaining frictions, such as obtaining private property rights or correcting for poor geography such as the need to level land or drain a swamp. The final term in (11) gives the discounted value of land-rents earned over the lives of consecutive formal sector buildings, constructed at dates $\tau_2, \tau_3 ...$ The land-

[8] Using $k\left(v_F\right) = \kappa_F v_F^{\gamma}$ the condition $\mu p v_F \Phi = k\left(v_F\right)$ becomes $\mu p \Phi = \kappa_F v_F^{\gamma-1}$ and using the first eqn. in (9) to substitute for v_F gives $\mu = 1/\gamma$.

rent terms in this expression depend on base house-rent per unit volume $p(x,t)$, which we assume to be exogenous to the city (see Theory Appendix, 1.4) and monotonically increasing in t.

Dates of development and redevelopment are chosen to maximise $PV(x)$. For the first development (which we assume for the moment to be informal), the optimal τ_0 simply equates flow land-rents on undeveloped and informal land, and is implicitly defined by

$$\frac{\partial PV(x)}{\partial \tau_0} = e^{-\rho\tau_0}\left[r_0 - r_I(x,\tau_0)\right] = 0, \tag{12}$$

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

$$r_0 = \kappa_I(\alpha-1)\left[a_I\, p(x,\tau_0)/\kappa_I\alpha\right]^{\frac{\alpha}{\alpha-1}}. \tag{12a}$$

Since house-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 Theory Appendix 1.1)

$$\frac{\partial PV(x)}{\partial \tau_1} = e^{-\rho\tau_1}\left[r_I(x,\tau_1) - p(x,\tau_1)v_F(x,\tau_1) + \rho\left\{k\left(v_F(x,\tau_1)\right) + D\right\}\right] = 0. \tag{13}$$

Using (9) and (10) this can be written as

$$r_I(x,\tau_1) = r_F(x,\tau_1,\tau_1)(\gamma - \rho\Phi(x,i))/(\gamma-1) - \rho D. \tag{13a}$$

With iso-elasticity, this gives a unique switching point from informal to formal if $\alpha > \gamma$, (from eqns. 5 and 10), a condition which we will see is satisfied in the data for Nairobi.

The first redevelopment of formal land is at date τ_2 satisfying (see Theory Appendix 1.1)

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

Generalising this for all redevelopments gives:

$$p(x,\tau_{i+1})\left[v_F(x,\tau_{i+1}) - v_F(x,\tau_i)\right] = \rho k\left(v_F(x,\tau_{i+1})\right), \quad \text{for } i \geq 1. \tag{14}$$

With iso-elasticity and using (9) and (10) this can be written as

$$\frac{r_F(x,\tau_{i+1},\tau_{i+1})}{r_F(x,\tau_{i+1},\tau_i)} = \left[\frac{p(x,\tau_{i+1})\Phi(x,i+1)}{p(x,\tau_i)\Phi(x,i)}\right]^{\frac{1}{\gamma-1}} = \frac{\gamma}{\gamma - \rho\Phi(x,i+1)} \tag{14a}$$

Intuition on these switch points can be seen from inspection of (14).[9] 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; similar intuition applies to eqn. (13). Eqns. (12) – (14) 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. (7), form the basis of the analysis of the next sub-section.

2.3. Analysis

What do we learn from the characterisation of development stages given above? A benchmark case in which house-rents are growing at constant exponential rates, $\hat{p} > 0$ yields analytical results. The full general equilibrium model that supports constant exponential price growth is noted in the Theory Appendix, but here we simply assume these house-rent paths. We look first at urban dynamics, the time series development of a particular place x, and then at the urban cross-section.

2.3.1 Urban dynamics at any location

This section looks at successive redevelopments of a unit of land in a formal area of the city.

Proposition 1: If formal sector construction costs are iso-elastic in height (with elasticity γ), house-rents are growing at constant exponential rate $\hat{p}$, $\rho > \hat{p} > 0$, 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 $\Delta\tau$,

$$\Phi = \int_0^{\Delta\tau} e^{(\hat{p}-\rho)t} dt = \frac{1-e^{(\hat{p}-\rho)\Delta\tau}}{\rho-\hat{p}}, \qquad \Delta\tau = \frac{\gamma-1}{\hat{p}} \ln\left[\frac{\gamma}{\gamma-\rho\Phi}\right] \tag{15}$$

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

$$\frac{v_F(x,\tau_{i+1})}{v_F(x,\tau_i)} = e^{\frac{\hat{p}\Delta\tau}{(\gamma-1)}} = \frac{\gamma}{\gamma-\rho\Phi} > 1. \tag{16}$$

[9] In equations (13) and (14) the switch dates are not when flow land-rents are equalised. This is because land-rents (net of amortization) jump and follow time paths with different gradients; for example, in (14a) rent jumps up at formalisation but then, given fixed volume, increases less rapidly during the life of a formal development than it would under informal development. The terms in γ and $\rho\Phi$ make the appropriate adjustments.

(iii) If the rate of growth of prices is the same in all locations, *x*, then Φ, Δτ, and volume growth are the same in all locations.

The first part of this proposition comes from integrating eqn. (7), using it in (14a), and noting that there is a unique solution solving the two parts of (15) with constant Φ and Δτ over time and space. The second part follows by using this in the first order condition for volume, (9). The third comes from noting that (15) and (16) do not depend on *x*. While volume ratios and time intervals do not vary with *x*, the actual dates of redevelopment do, as discussed below.

It follows from this proposition that the capital value of a unit of land already in the formal sector at location *x* at newly redeveloped at time τ_i is given by

$$PV_F(x,\tau_i) = r_F(x,\tau_i,\tau_i)\Phi / \left[1 - e^{-(\rho - \hat{p}\gamma/(\gamma-1))\Delta\tau}\right]. \tag{17}$$

This is derived using $R_F(x,\tau_i) = \Phi(x,i) r_F(x,\tau_i,\tau_i)$ in (10a) once formal sector development occurs and extending to infinitely repeated cycles of redevelopment.[10] This relationship between current rent and capital 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.

What about the earlier stages of informal development? The first transition we assumed is from rural to informal settlement. For land at *x* this occurs at date τ_0 when $p(x,t)$, the quality un-adjusted informal sector house-rent, reaches the trigger value given by (12a).[11] The transition from informal to formal settlement is given by date τ_1 that solves (13a). There is a unique transition date satisfying the second order condition, providing that $\alpha > \gamma$. We assume the condition to be satisfied, as it is in the data. The implications are 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 (eqns. (5), (10)).

Fig. 2 illustrates our results in a stylized benchmark city without frictions, using model parameters estimated in Section 3. 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 CBD, 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,

---

[10] $r_F(x,\tau_i,\tau_i)$ is growing at rate $\hat{p}\gamma/(\gamma-1)$ and is discounted at ρ. The sum is finite if $\rho - \hat{p}\gamma/(\gamma-1) > 0$.

[11] A period of informal settlement exists only if the return to informality at date τ_0 is greater than commencing formal settlement, $r_I(x,\tau_0) > p(x,\tau_0)v_F(x,\tau_0) - \rho\{k_F(v_F(x,\tau_0)) + D\}$ (see eqn. (11)). If not, then initial development will be formal, with date τ_1 implicitly defined by $r_0 = p(x,\tau_1)v_F(x,\tau_1) - \rho\{k_F(v_F(x,\tau_1)) + D\}$.

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 (although slightly), 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 interval $\Delta\tau$ 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.

2.3.2 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 CBD, and assume that house-rents, $p(x, t)$, remain exogenous and decrease with distance from the CBD at exponential rate θ. In the Theory Appendix we give the underlying assumptions that yield this exponential forms. [12]

Exponential decline with respect to distance together with exponential growth through time imply house-rents are

$$p(x,t) = \overline{p}e^{\hat{p}t}e^{-\theta x}. \tag{18}$$

Given these, the trigger house-rent for informal development in eqn. (12a) depends on both date and place according to

$$p(x,t) = \overline{p}e^{\hat{p}t}e^{-\theta x} = \frac{\kappa_I \alpha}{a_0}\left[\frac{r_0}{(\alpha-1)\kappa_I}\right]^{(1-1/\alpha)}. \tag{12b}$$

This can be interpreted either as giving the date at which place x develops or the place that develops at date t, i.e. the historical induction of place x into development, $t = \tau_0(x)$, or the date t city edge, $x_0(t)$. Similarly, eqns. (13) and (13a) can be interpreted as giving the distance at which first formalisation occurs at date t, $x_1(t)$, and (13a) the distances between

[12] We use a representative consumer model in order to highlight the role of technology and formalisation costs. In a more general case, if consumers differed just by income, under log linear preferences, that would not induce separation in sector choice by income, although there obviously are preference specifications which do. The data suggest strong overlap with some separation in the tails, but that is not the focus of this paper. The number graduating from high school is the same proportion (24%) of *both* slum and formal sector residents in the 2009 Census. Slums do have fewer household heads with college degrees (7% versus 21% in the formal sector) and more who terminate education after primary (standard 8, with 23% in slums versus 13% in formal housing).

areas of redevelopment occurring at date t, $x_i(t)$, $i > 1$. Proposition 2 states how different stages of development (building types and heights) vary across this stylized city as it grows.

Proposition 2: If formal sector construction costs and informal sector quality are iso-elastic, house-rents are growing at constant exponential rates $\hat{p}$, $\rho > \hat{p} > 0$, 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 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,

$$\Delta x = \frac{(\gamma - 1)}{\theta} \ln\left[\frac{\gamma}{\gamma - \rho\Phi}\right]. \tag{19}$$

See Theory Appendix 1.2 for proof.

These results are illustrated in Fig. 2, where we now fix a date and move along a line parallel to the x axis. At the city edge land is informal and, moving towards the centre, locations that have been urban for longer have been through more stages of development and offer greater building volume per unit land. 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 first 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 falls with time as the city gets larger. Finally part (iii) says that the width of formal sector rings of development is constant. Iso-elasticity yields simple patterns in a benchmark city without frictions.

2.4 Spatial heterogeneity

Cities do not look like the highly regular structure in Fig. 2 and Proposition 2 although, as we will see, overall, Nairobi is consistent with these benchmark patterns. The patterns in Fig. 2 can be modified by frictions which are spatially heterogeneous. We focus on two frictions which capture the fixed cost of formal sector development, D, and which vary by location. These are the costs of obtaining secure property rights necessary to formalise development, and variation in first time formal sector construction costs due to factors such as land quality and terrain.

Formal sector development requires private property rights (on either leasehold or freehold land) to avoid risk of expropriation, to obtain financing and insurance, and to clarify issues such as inheritance, compensation for takings and the like. The long and tortured development of private property rights in Nairobi, as discussed later, suggests that frictions in securing these rights has been a critical factor in the city's development, and that these frictions vary across areas of the city. In the model, this can create positive *D*'s at locations without full private property rights, with the following implications. First, a higher value of *D* in a particular place means that higher formal sector house-rents are needed to trigger formalisation, so has the effect of postponing the first formal sector development in eqn. (13a). During this extended period slums persist, even though neighbouring areas (at the same distance but with lower *D*) are formalised; volume per unit area and crowding increase in these slums. Second, effects persist through waves of redevelopment. We know from Proposition 1 that the time interval between each redevelopment at a particular place is fixed. It follows that if first formal development is later and taller, then so too are subsequent redevelopments. Third, we know that formal sector developments at different distances at the same date will have less volume the further they are from the city centre, since eqn. (9) implies

$$\frac{d\log(v_F : \tau_i = t)}{dx} = \frac{-\theta}{\gamma - 1}, \qquad (20)$$

This relationship will be used in the empirical work to identify the diseconomies of building high.

Fig. 3 illustrates these points in a stylised case in which *D* varies with distance from the centre, set in the figure at random non-negative values. While the city overall still displays lower intensity of use further from the centre and each place has increasing volume through time, the city cross section (along a particular ray, and more generally around a city ring) exhibits areas of informality next to formal structures, some of which may have gone through several phases of redevelopment. Such patterns are the hotchpotch we see in the data.

The second form of spatial heterogeneity arises from variations in the construction cost of preparing land for first formal sector development. Even in areas of new formal development at similar locations, not all construction occurs at exactly the same date. As we will see in the data, there is a process of 'infill' with the fraction of the area covered by buildings increasing through time. This is consistent with constant long run c_F if there are fixed costs of first formal construction that vary across places. These costs may vary systematically with terrain, for example the costs of draining a swamp or levelling a rugged site for development. We offer empirical evidence (Land Quality Appendix 3) for the presence of such effects, and for the proposition that they are best thought of as fixed rather than marginal construction costs, since poorer terrain qualities do not reduce building heights on redeveloped formal sector land.[13] The consequences of such costs are analogous to the variations in *D* discussed above. Even in non-slum areas the process of formal construction will take time, will have long-lasting effects, and will show up in the data as 'infill' and increasing cover to area ratios. The duration of such an infill process will depend on the level and distribution of these costs; the data we present later suggests that infill process has largely played out on land close to the centre, but is still important on more recently formalised land further out.

3. The geography 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 focus on three issues. First, what new facts about the built environment emerge, concerning the cross-section of a developing city and the dynamics of its evolution over time? Second, do the data support assumptions made in the model, and are they consistent with its predictions? Third, we estimate equations needed to recover key parameters of the model ($\alpha,\ \gamma,\ \theta,$ and $\hat{p}$) which are both of intrinsic interest and are used in the welfare analysis in Section 4. We first describe the data.

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 generally work with 6470 150m x 150m grid cells in the 2003 area of the city, based on data aggregated from 40 cm resolution to 3m x 3m cells to the grid squares we work with. From aerial photographs we have tracings of all building footprints for 2003 and 2015 and we have LiDAR height data for 2015. To infer 2003 heights, we assume that if building footprint is unchanged between dates then so is height (i.e. we ignore the possibility of adding floors to a structure). For demolished buildings, we assume 2003 heights equal the average height of unchanged buildings in the 8 queen neighbouring grid squares. Both assumptions are likely to overstate relevant 2003 heights and thus understate volume changes since demolished buildings are likely less tall than unchanged ones. To focus on private sector development, we remove all grid squares entirely in permanent public uses listed in the Data Appendix amounting to 11% of land in the 2003 city boundary, and 25% at the centre (0-1 km, including major and the Presidential palace); neighbourhood schools and roads remain.

Besides data on the built fabric of the city, asking prices for vacant land in 2015 are obtained by scraping from property24.co.ke, a website that advertises property for sale in Kenya. Listings are only found for the formal sector. Data on house-rent, housing characteristics,

[13] Specifically, in Appendix Table 3.2 rebuilding heights are not affected by elevation, presence of water in the neighbourhood, and are not decreased by increases in ruggedness.

and slum or formal usage are derived from a georeferenced household data set from the 2012 'Kenya: State of the Cities' survey by the National Opinion Research Center (NORC).

Our analysis requires a distinction between formal and informal settlements (or slums as they are called in classification studies). Empirically, we base this on a 2011 slum mapping by IPE Global Limited under the Kenya Informal Settlements program, with details given in the Data Appendix. The IPE mapping used satellite imagery and topographic maps defining slums as "unplanned settlements" with some aspects of low house quality, poor infrastructure, or insecure tenure. An alternative delineation of slums comes from a 2003 land use map prepared by the CSUD at Columbia University. 2004 satellite imagery (SPOT) indicates that it misses some emerging slum areas which appear in the 2011 slum mapping. While it is clear that the effective definitions differ in precise detail across the two studies, we do use the 2003 mapping to try to distinguish slum areas that underwent formal sector development between 2003 and 2011.

In Fig. 4 we show these two mappings of slums and define the area of the city we work with. The centre is marked by a yellow star. We adopt a conservative definition of the urban boundaries for 2003 and 2015 based on built cover. For a (150m x150m) 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 which contiguously connect to the CBD. Fig. 4 shows the dashed city outline in 2003 and the solid outline in 2015.

Slum areas are marked green if recorded in both studies, yellow if only in IPE (2011) and blue if only in CSUD (2003). We will look at these blue areas in particular to show aspects of slum conversion to formal sector use. However, there seems to be little overall slum removal, the issue of concern in Section 4. Areas near the CBD with no recorded slums are marked by the dashed and solid rings for respective periods. The area with no slums expands considerably between dates, from about a 0.8 km to about a 2 km radius around the centre by 2011, although the removed slum areas are tiny. The map suggests considerable slum expansion (yellow) at the 2003 fringe of the city and beyond, as predicted in the model. We again note the large slum of Kibera directly south-west of the centre (ranging from about 3-5 km of the centre), which we will discuss in some detail. [14]

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 key parameters of the model and evaluate key model assumptions. 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,

[14] In application of Fig. 4, as detailed in the Data Appendix, we adjust the tightly drawn 2011 slum boundaries to allocate unused land to the nearest formal or slum usage.

2018). We examine land prices, the built fabric in slum areas and then in formal areas. Results are presented in figures and in the regressions in Table 1. Figures are for the 2003 city, generally cut at 10kms from the centre, based on observations at the 150mx150m grid square level. Regressions cover all of the 2003 city. Following the model, regressions are semi-log, so the gradient coefficient can be interpreted as proportionate change per km distance from the centre. All regressions contain controls on average elevation and ruggedness for each 150mx150m grid square. The regression in the first column of the table has a few additional listed controls. More details and coefficients on all covariates for the key cols. 1, 3 and 8 are in Data Appendix Table 2.4.

3.2.1 Land prices

Cross section and cross time patterns of the built environment are driven by underlying fundamentals reflected in prices. The regression in col. 1 uses the 2015 property24.co.ke data to estimate the formal sector land price gradient with respect to distance from the centre, with intercept of ln asking price (per square meter) in USA$. An exponential form (as in eqn. 18 and proposition 2) 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). Second, this slope yields an estimate of $-\theta\gamma/(1-\gamma)$, derived from eq. (9) with (18), which combined with a formal sector height gradient below will allow us to solve the rate of house rent decline with distance from the centre, θ, and the diseconomy in building high γ.

3.2.2 Slums

This steep price gradient drives intensity of land use. 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, corresponding to v in the theory section. For slums, BVAR declines significantly with distance from the city centre at a rate of 9.5% per km (Table 1 col. 3). This measures the gradient slope $(dv_I/dx)(1/v_I) = -\theta\alpha/(1-\alpha)$, derived from eq. (4) and (18), which will be used below to recover slum diseconomies, α. Fig. 5 shows the slum BVAR gradient in red with confidence intervals on the mean and the 25th and 75th percentile of observations. [15]

BVAR is height times cover to area. In the theory we assumed constant slum height throughout the city. In Table 1 col. 7, we see that slum heights are not decreasing with

---

[15] 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. This is STATA local mean smoothing with an Epanechnikov kernel and default settings.

distance from the centre and are almost flat, also as pictured in Fig. 6 by the red dotted line. With constant slum height, gradients for slum BVAR and slum cover to area ratio, CAR (c in the model) should be the same; the estimates of -0.0948 and -0.103 in Table 1 cols. 3 and 5 confirm this.[16] Slum CAR near the city centre in Fig. 6 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.[17]

Table 1: Spatial Gradients

	(1)	(2)	(3)	(4)	(5)
	Ln land price ($2015 per m^2.)	Ln formal BVAR	Ln slum BVAR	Ln formal CAR	Ln slum CAR
Distance to centre	-0.172 (0.0476)	-0.0493 (0.00531)	-0.0948 (0.0104)	0.0239 (0.00428)	-0.103 (0.00966)
"Intercept" for typical item	7.254 (0.277)	0.747 (0.0351)	1.275 (0.0698)	-1.672 (0.0278)	-0.116 (0.0642)
Controls, apart from ruggedness & elevation	yes	no	no	no	no
Observations	136	5435	958	5435	958
R-squared	0.292	0.035	0.104	0.043	0.142

	(6)	(7)	(8)	(9)
	Ln formal height	Ln slum height	Ln formal redeveloped height; quantile: 80th percentile	Ln formal unchanged height; quantile: 20th percentile
Distance to centre	-0.0763 (0.00246)	0.0107 (0.00405)	-0.101 (0.00521)	-0.0763 (0.00223)
"Intercept" for typical term	2.487 (0.0166)	1.375 (0.0308)	3.315 (0.0356)	2.131 (0.0150)
Controls for ruggedness & elevation	no	no	no	no
Observations	5435	958	4621	5079
R-squared	0.203	0.018		

Note: Regressions are based on observations for the 2003 extent of the city. Standard errors in parentheses. Column 1 adds in controls for plot area, month & year of sale and whether the plot has recorded GPS coordinates. Details and full results results for columns 1, 3 and 8 are in Table 2.4 of the Appendix. Reported intercepts are predicted values of a typical lot at the city centre. For example, the reported intercept in col. 1 is the predicted price at the centre for a city wide typical plot based on median area, ruggedness and elevation. For height regressions we only include grids for which there is cover and so can be calculated. For BVAR and CAR regression overall about 5% of observations have no cover and hence volume; some may be playing fields, overpasses, small parks and the like. A Tobit including these as 'censored' at 0 yields almost identical slope coefficients. Standard errors are robust and clustered based on a 2250m x 2250m grid.

3.2.3 Formal Sector.

[16] Note the slopes for slum CAR and height (-0.103+0.0107) sum to almost that for BVAR, -0.0948.
[17] We also have road length and width data extracted from high resolution SPOT satellite data for 2015. Much more coverage by way of paved roads is in the formal than the slum sector, where in the formal sector roads are about 15% of coverage near the centre while in slums they maybe hit 5%.

While slums are malleable, the formal sector is not and we must distinguish stock and flow. For the stock, formal sector BVAR near the centre very high, averaging around 8 m^3 of space per m^2 of ground area (including roads) in Figure 5; and then declines at a rate of 4.9% per km from the centre from Table 1, col. 2. Heights decline with distance from the centre, at a rate of 7.6% per km (Table 1 col. 6). This compares with height gradient numbers for Chicago in the 1960's or 70's from Ahlfeldt and McMillen (2018).[18] We assumed in the theory that the formal sector has constant CAR, or c, throughout the city. Table 1, col 4 and the solid blue line in Figure 6 indicate that CAR does not change much across the city and is not decreasing with distance from the centre. Formal sector BVAR thus varies with height.

For recovering formal sector parameters, based on the theory, we focus on the flow of newly redeveloped buildings.[19] The relevant height gradient is that given in eqn. (20), $\left(dv_F / dx\right).\left(1/v_F\right) = \left(dh_F / dx\right).\left(1/h_F\right) = -\theta/(\gamma-1)$, reflecting the rate of house rent decline and diseconomies in height. To get more recent buildings from the 2003-2015 interval we use a quantile regression for just redeveloped buildings. We settled on the 80th percentile, balancing out wanting more recent hence taller buildings, against getting extreme outliers at any time. That said for r*edeveloped* buildings different quantile and OLS estimates are very similar.[20] From Table 1 col. 8 the slope is -0.101, large in absolute value than for the overall stock.

3.2.4 Key parameters

We have identified three gradients, for land prices, slum BVAR and formal sector redeveloped height which are used to solve for model parameters $\{\gamma, \alpha, \theta\}$. The relevant relationships from Sections 3.2.2 and 3.2.3 are repeated in Table 2 and the bottom row shows the solutions to the three equations. The value of $\gamma = 1.70$ implies that the share of land-rent in formal sector revenue is 0.41, while $\alpha = 3.98$ implies a corresponding share in the informal sector of 0.75 (eqns. (10) and (5) respectively). 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 there are no data we know of to make comparisons, since slum land is generally not officially transacted. However, given the low construction costs of slum housing such a high land-rent share seems reasonable. Note that $\alpha > \gamma$ is required for the pattern of development with new slums at the city edge in the model. Finally Table 2 tells us the slope of the house price gradient, $\theta = 0.071$. This is similar to estimates for the top quartile of French cities (Combes et al, 2018).

---

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

[19] Note we can't do redeveloped BVAR, since we can't properly assign land area to new buildings (vs old)

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

Table 2. Gradients and parameters		
Gradients:	Formal:	Informal:
Volume (m^3 per land m^2)	$\frac{dv_F}{dx}\frac{1}{v_F}=\frac{-\theta}{\gamma-1}$=-0.101	$\frac{dv_I}{dx}\frac{1}{v_I}=\frac{-\theta\alpha}{\alpha-1}$=-0.0948
Land-rent (per m^2 land)	$\frac{dPV(\tau\geq\tau_1)}{dx}\frac{1}{PV}=\frac{dr_F}{dx}\frac{1}{r_F}=\frac{-\theta\gamma}{\gamma-1}=-0.172$	$\frac{dr_I}{dx}\frac{1}{r_I}=\frac{-\theta\alpha}{\alpha-1}$ *
Solutions: γ = 1.703, α = 3.983, θ = 0.071		

* For completeness, the bottom left cell of Table 2 allows us to back out the gradient of informal sector land-rent. We do not have empirical observations on this, but the model implies that the informal land-rent gradient is the same as the volume gradient (since house-rent per unit volume is constant across space), or -0.0948 from the cell above. This is flatter than that in the formal sector (-0.172), a consequence of $\alpha > \gamma$.

3.2.5. Other important findings

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) in Fig. 6 is about 10 stories (at about 3m a storey) and 5% of the grid squares are over 16 stories. This despite the view sometimes expressed that height in African cities is constrained, e.g. by unreliability of power for elevators. Second, at 2km and beyond, slums and the formal sector deliver essentially the same BVAR in Figure 5.[21] 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. Finally there is huge heterogeneity across grid squares in BVAR's at each distance. In part this reflects where roads, playgrounds, small parks and parking lots are since they cover area but offer no built volume. And the theory tells us that throughout the city, variability will come from the differential timing of formal sector development, based on differential geography and the specific history of a property's path to formalization.

3.3 City dynamics

This section on dynamics is divided into three sections on empirical findings and then a final section where we return to model parameterization and back out the rate of housing and land price appreciation.

As noted in the Introduction, total built volume in non-public use within the 2003 [2015] boundary increased by 48 [59] % between 2003 and 2015. The theory suggests that volume

[21] At 6.5 km, the BVAR in slums does bump up; but, as we noted earlier for Fig. 1, that may be due to misclassification in the northeast of the city.

changes need to be differentiated by sector and type. Here we present novel facts about this process, which we will argue are consistent with the theory. We look first at the formal and then the slum sectors within the 2003 city boundary, before turning to the extensive margin beyond the 2003 boundary.

3.3.1 Formal sector redevelopment and infill:

In the urban core, we expect to see a process of redevelopment and infill in the formal sector, and increasing cover to area ratios in the slum sector. The formal sector exhibits extremely large overall volume changes, as indicated in Fig. 7, which is cut at 8kms to see clearly what happens nearer the city centre. There are 35- 55% increases in total volume from 2-8 km, the increase being greatest around 4-5 km from the centre. There is less change in the first 2 kms which has already experienced a wave of redevelopment and sky-scraper construction in the late 20th century, and where there is also lock-in by historical buildings.

The growth in formal volume is a combination of redevelopment, infill and demolished sites. Redevelopment involves new buildings which overlap the footprint of 2003 buildings which were torn down; infill is new buildings not intersecting any 2003 buildings; demolition (not yet redeveloped) is a residual, with small changes throughout. Figure 7 shows that volume growth up to 4.5 km from the city centre is driven principally by net redevelopment (new volume minus old volume torn down), while beyond that infill dominates. The scale of redevelopment is enormous. From 1 to 4km about 35% of buildings are torn down and redeveloped over the 12 year period. To benchmark the demolition rate of 35%, in the USA, the American Housing Survey data gives demolition and removal by disaster (fire, hurricane and the like). Depending on the year of the data, 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 typically 3-4 times that.

The benchmark model predicts that each distance from the city centre experiences a wave of redevelopment at a particular point in time. In Section 4 we will show numbers for the model's predicted times of redevelopment. Here we note that, even in the face of spatial heterogeneity discussed in Section 2.4, this element comes through in Fig. 7, where, in our time interval, redevelopment peaks at 3km, significantly higher than the central area which was redeveloped earlier and higher than further out which was developed later. However, spatial heterogeneity in geography affecting the cost of first formal sector development ('levelling the grade' or 'draining the swamp') implies that formal construction takes place over a protracted period of time. In longer established inner city areas with high land prices, the theory says the infill process has played out with only the most marginal land left and redevelopment (demolition and reconstruction) dominating. In more recently developed areas further out infill is still occurring, and redevelopment has not yet been triggered. This is exactly the pattern shown in Fig. 7.

The importance of terrain can be directly established by looking at land quality measures such as ruggedness, elevation and water/ wetlands. In the Land Quality Appendix, Table 3.1, we report the results of investigating this at fine spatial scale. Independent variables are ruggedness in the 90x90m square neighbourhood (the standard deviation of elevation for the own 30x30m grid square of the 90x90m square neighbourhood which is queen neighbours plus the own square); water (river, wetland, pond) in your 90x90m neighbourhood; and ruggedness in the larger 450x450m neighbourhood. The results show, as predicted, that formal sector infill near the centre is on much worse quality land than already developed buildings, but the effect falls rapidly with distance so, by 10kms out, there is no clear difference between infill and already developed. Correspondingly Table 3.2 in the Land Quality Appendix shows that the height (after levelling the grade) of redeveloped buildings is unaffected by underlying terrain fundamentals (i.e. κ_F in eqn. (9) is not affected for redeveloped buildings).

Formal sector volume increases involve some rise in cover to area ratios throughout the city during the process of infill, but is largely due to increased building height.[22] Fig. 8 splits out the formal sector height gradient in Fig. 6 to show how the mean heights of redeveloped versus the stock of unchanged buildings compare, already implied by differential gradients estimates. It also shows the same for slums. In the formal sector, beyond 1.5km, redeveloped buildings are on average significantly higher; as would be expected, the effect is larger closer to the centre, nearly doubling the height of unchanged buildings in the interval 2-4km. This building higher drives the large volume changes between 2003 and 2015 in the mid-portions of the city. To get a better sense of height increases we compare the height of the 80th percentile of redeveloped buildings mimicking heights of new buildings just finished in the 2003-2015 interval with the 20th percentile of unchanged buildings, trying to mimic the height of those about to be torn down. This is graphed in the Appendix Figure 2.2. From 1-4.5 kms, the graph suggests building height increases on average by around 2.75-fold with redevelopment.

3.3.2 Slum dynamics in the city core

What happened to existing 2011 slum areas within the 2003 city-boundary? Fig. 8 indicates, in line with assumptions of 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. In Fig. 9, we show total slum volume changes in the 2 – 8 km interval and their decomposition into "redevelopment" and infill. Both seem equally important throughout,

[22] We note height increases involve an increase in average footprint size of 100% at 3 km, rising to 200% by 6km. In reality, scale economies and construction efficiency require more footprint as height increases to allow for elevators, staircases, and reinforced construction.

reflecting increasing CAR or infill, but also house rearrangement consistent with the Meccano parts assumption.

On conversion of slums, we do have a proxy measure, which are areas identified as slums in the 2003 mapping but not in the 2011 mapping. As noted above, some of the difference is due to different methodologies employed by researchers, and some is likely to be real change. In Appendix Figure 2.3, we show rates of building teardown for 2011 slums (non-converted) and slums which 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 Figure 2.4 also shows that buildings on converted slum land generally are built significantly taller than those in current slums.

3.3.3 The extensive margin

As noted in the introduction, volume in the margin between the 2003 and 2015 boundaries (see Fig. 4) increased by 120%. To see what happens at the extensive margin, Fig. 10 looks at rates of total volume change by sector for the entire 2015 city (not just the 2003 city) out to 12 kms. Nearer the city edge in both sectors this is almost all due to infill. Formal sector rates of volume growth dominate slums until 10kms and peak at 9.5 kms at a rate of 150%. For slums the rate is over 400% at 12 kms and continues to rise beyond that to 1500% (not shown) at 14kms, of course from a small base 2003 volume. At the very city edge, slum rates of development totally dominate formal sector rates.

3.3.4 Model dynamics

What do these findings imply for remaining variables of the model which we do not directly observe, such as land-rents, price appreciation, the time interval to redevelopment, $\Delta\tau$, and the value to rent ratio Φ?

The value of land-rents can be established using the house-rent data from the NORC, together with knowledge of the share of land-rent in house-rent ($1-1/\gamma$, $1-1/\alpha$ for formal and slums respectively). We use the NORC data to estimate hedonic price regressions for formal and slum rents; Table 3 reports gradients and intercepts in 2015 USA$ for m^3 of volume.[23] These regressions control for house characteristics in order to define rents per unit volume for *typical* formal and slum sector houses; full hedonic regressions are reported in Appendix Table 2.4.

Based on the components of (10) and (5), we use the intercepts of these regressions, together with information on building volume per m^2 land area, and γ, α to calculate land-rents

[23] This assumes each sq meter of floor space yields 3.0 cubic meters 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$.

projected to the origin, that is to the city centre x = 0 and year 2015. Values are r_F = \$65.09 and r_I = \$17.5.[24] Moving away from the centre and through time, the paths of land rent are therefore, for the formal sector, (eqns. 10, 10a, 17),

Table 3. House-rent estimation		
	Ln formal house-rent m³ in \$2015	Ln slum house-rent m³ in \$2015
Distance to centre	-0.0557 (0.0308)	-0.00509 (0.0288)
Intercept (x = 0) for typical item	3.144 (0.0663)	1.879 (0.0576)
Controls, apart from ruggedness & elevation	yes	yes
Observations	361	442
R-squared	0.307	0.406
The intercept for slum and formal sector house-rents is based on two regressions. For slums, we redo the regression in the table setting the distance coefficient to 0, as in the model (i.e. there is no distance variable in the regression). For the formal sector we set it to θ solved for in the model of -0.071. The impact of these forced gradients on estimated house characteristics is minimal, given the within sector differences in gradients are so small. Standard errors are robust and clustered at the enumeration unit level for these residences. We also evaluate at typical characteristics as detailed in the Appendix.		

$$r_F(x,t,t) = (1-1/\gamma)pv_F = r_F e^{[\hat{p}(t-2015)-\theta x]\gamma/(\gamma-1)}, \qquad r_F = \$65.09, \qquad (10a)$$

and for the informal sector, (eqn. 5),

$$r_I(x,t) = (1-1/\alpha)pa(v_I)v_I = r_I e^{[\hat{p}(t-2015)-\theta x]\alpha/(\alpha-1)}, \quad r_I = \$17.55. \qquad (5a)$$

Finally, the theory provides three further equations from which we calculate the rate of price appreciation, $\hat{p}$, as well as Φ and $\Delta\tau$.[25] These are the value-to-rent ratio, $\Phi = \left[1-e^{(\hat{p}-\rho)\Delta\tau}\right]/(\rho-\hat{p})$ and the equilibrium building life, $\Delta\tau = ((\gamma-1)/\hat{p}_F)\ln(\gamma/(\gamma-\rho\Phi))$ both in eqn. (15), as well as the relationship between the rental of land and its capital value, $PV_F(x,\tau_i) = r_F(x,\tau_i,\tau_i)\Phi/\left[1-e^{-(\rho-\hat{p}\gamma/(\gamma-1))\Delta\tau}\right]$ in (17). These expressions include the discount rate ρ and the capital value of land, PV. For the former we use the real interest rate ρ = 0.057, which is the average of the World Bank real interest rate numbers for Kenya for

---

[24] Henceforth r_F, r_I and PV denote values of variables evaluated at x = 0 and $t = \tau$ = 2015. Building volume in eqn. (5) per unit area in the informal sector at the city centre comes from Table 1 column 3. The predicted value of $PV_F(0,2015)$ is \$1414 from the intercept in Table 1 col. 1. For , $r_F(0,2015) = \$24.80$, we have the intercept from Table 3 for $p(0,2015)$ and BVAR from Table 1 col. 8 giving the intercept height on redeveloped buildings which is then multiplied by the median CAR of 0.247 for all grid squares in the 2003 city (assumed uniform throughout the city). Note again we do not observed plots assigned to buildings and hence BVAR on redeveloped buildings.

[25] Values of all other parameters are computed in Appendix Table 1.1.

the 14 years post 2002.[26] The latter is the intercept from the land value estimates of Table 1 column 1. Solving the equations gives

$$\hat{p} = 0.0094,\ \Phi = 20.68,\ \Delta\tau = 88.54 \text{ years.}$$

The value-to-rent ratio, Φ = 20.68, is close to that implied by an infinite stream discounted at 5.7%, and in the centre of the range of ratios reported by realtors for US cities. The length of life of new buildings is novel and an estimate for which we know of no easy comparison.

There are two other validity checks on the model from the data here. First observed slum rents (base house rent multiplied by the disamenity measure) is predicted from the theory to be constant throughout the city. In Table 3 col.2, this is confirmed, as the slum house-rent gradient is zero. Second, our parameter values also imply that, at each redevelopment, volume increases by a factor of 3.25, from eqn. 16 where $\frac{v_F(x,\tau_{i+1})}{v_F(x,\tau_i)} = \frac{\gamma}{\gamma - \rho\Phi}$. This compares with the directly estimated 2.75-fold increase in height reported at the end of section 3.3.1.

4. The cost of delayed formal sector development

In this section we discuss why older government owned slums are so persistent in the more central part of Nairobi. Then we use the model to calculate the loss in land values associated with this inefficient land-use on high opportunity cost land.

4.1 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 by 2015 (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 particular features which are problematic for formalisation. In Kibera for example, of 120 slum lords 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

[26] https://data.worldbank.org/indicator/FR.INR.RINR?locations=KE

economy issue is that if the government were to auction the land it 'owns' for formal use (or give it to the tenants), the slumlords would have no claim to the revenue since they have no legal claim to the land. They would simply lose profitable businesses. 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 of present day Kibera that is not occupied by Nubians and their descendants was settled by others and had 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 is yet another road block to redevelopment (Etherton 1971; Joireman and Vanderpoel, 2011).[27] The literature has similar stories for other major slums in Nairobi.[28]

4.2 The loss of land value due to slum persistence.

In our open city model 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 at date *s* of a property which is converted from informality to formality at date *z*. We denote this present value $PV(x,s,z)$ and, from eqn. (11) with *D*'s set at zero, it is given by

$$PV(x,s,z) = \int_s^z r_I(x,t)e^{-\rho(t-s)}dt + e^{-\rho(z-s)} \sum_{i=0}^{\infty} R_F(x, z + i\Delta\tau)e^{-\rho i\Delta\tau}$$

[27] Further documentation on the Nubian settlers and their claims in Kibera can be found online at http://www.nubiansinkenya.com.

[28] For example, Mathare 3km northeast of the centre, was originally a stone quarry managed by an Indian businessman in the early part of the 20th century. When the quarry closed, the land went to the Department of Defence. Over time that land and surrounding villages were 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, long claims of corruption, and there being 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 how Kenyan tenure legalization can take decades to implement due to the needs of reconciling the various interests of stakeholder and offers more examples, such as the Korogocho slum.

$$= r_I(x,s)\frac{1-e^{-(\rho-\hat{p}\alpha/(\alpha-1))(z-s)}}{\rho-\hat{p}\alpha/(\alpha-1)} + r_F(x,s,s)\frac{\Phi e^{-(\rho-\hat{p}\gamma/(\gamma-1))(z-s)}}{1-e^{-(\rho-\hat{p}\gamma/(\gamma-1))\Delta\tau}}. \qquad (21)$$

The second term in the second line comes from eqn. (10a) and (17) on the value of land once it is in formal sector use, and the first from integration of rents in the informal sector until conversion to formal. Expressions for land rentals are given by eqns. (5a) and (10a), and using these in (21) we can solve for present value *PV*(*x*,*s*,*z*) .[29]

Table 4. The value of land formalised at different dates Present values at *s* = 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
PV (*x*, 2015, *z* = 2015)	1297	1092	920	774	652	549
PV (*x*, 2015, *z* = 2104)	430	387	349	314	283	256
PV (*x*, 2015, *z* = ∞)	376	342	311	283	258	234
Slum land, m^2, 2011	0	2,394	238,032	1,068,075	2,224,503	1,900,341
No. slum households, 2009	0	0	2920	29,070	45,810	33,100
D (lower bound)	488	391	311	245	192	148
Date of efficient formalisation	1919	1926	1934	1942	1949	1957

Values of *PV*(*x*,*s*,*z*) are given in Table 4 as a function of distance *x* and formalisation date *z*, with present value evaluated at *s* = 2015. These values can be visualised as a surface on {*x*,*z*} space, with a line of maxima giving, for each *x*, the date *z* at which it is efficient for first formalisation to occur. These dates are given in the bottom row of the table, noting that this backcast assumes technology and price appreciation rates have never changed over Nairobi's history. By adding 89 years to those dates we get the projected dates of the first wave of redevelopment in Nairobi; they are a little later than what is reality but not out of line.

The body of the table gives present values if formalisation occurs at dates *z* = 2015, 2104, and infinity.[30] For each distance reported in the table these are beyond the efficient date, so *PV*(*x*,*s*,*z*) decreases with *z* as well as with *x*. We compare the cost of never formalising with formalising in 2015 for each distance, *PV* (*x*, 2015, *z* = 2015) - *PV* (*x*, 2015, *z* = ∞). This

[29] With 2015 as the reference data the second line of (21) becomes

$$r_I e^{[\hat{p}(s-2015)-\theta x]\alpha/(\alpha-1)}\frac{1-e^{-(\rho-\hat{p}\alpha/(\alpha-1))(z-s)}}{\rho-\hat{p}\alpha/(\alpha-1)} + r_F e^{[\hat{p}(s-2015)-\theta x]\gamma/(\gamma-1)}\frac{\Phi e^{-(\rho-\hat{p}\gamma/(\gamma-1))(z-s)}}{1-e^{-(\rho-\hat{p}\gamma/(\gamma-1))\Delta\tau}}$$

[30] 2104 chosen as one development cycle after 2015.

ranges from \$921 per m^2 in the 0-1 km distance band to \$315 per m^2 at 5-6 km from the centre. Note these differentials decline dramatically if formalization is delayed to 2104.

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-\$283=\$491 per m^2. There are 1.13mn m^2 of slum land in that distance band, of which we estimate an additional 10% is in roads and public schools, so about 1.07 mn m^2 are available for redevelopment. There is thus an aggregate gain from converting at 2015 compared to perpetual delay of about \$525mn. For a perspective, suppose slumlords were compensated for conversion by \$283 per m^2, as if they had the right to hang on forever. That leaves the remaining surplus of \$491 per m^2. For the 29,000 households affected, the gain is about \$18,100 per household. This is a very large sum, for households paying on average under \$700 a year in house-rents. At 4-5 km which includes much more of Kibera with many more households, the same type of calculation gives a surplus of about \$17,200 per household.[31]

The penultimate row of Table 4 gives a lower bound estimate of D per m^2. Conceptually this is the value of D at which it becomes worthwhile to formalise in 2015, and it is computed from eqn. (13a) set as an equality. Even lower bound D's are very high, in the range of the value of land held in slum use forever, indicating how difficult the problem is.

For these calculations to be valid, it must be the case that slum lands near the centre are 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 Land Quality Appendix Table 3.3, we perform a border experiment at slum boundaries comparing slum lands with formal sector infill (vacant) lands that have just been built upon. We show that out to 6kms (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.

What are the biases in our estimates? For downward bias, we have used formal sector residential house-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. On the other side, we have ignored moving costs for slum residents. For those forced to move there could be losses

[31] These numbers are higher than preliminary back-of-the-envelope type calculations based on raw data reported in CEPR DP 11211 from April 2016, where we have a gain of \$13,000 US per slum household for the core city. But even these are still very large numbers. Here we have an estimate of real price increases and an appropriate discount rate.

in terms of job location and social networks, although proper relocation programs could mitigate those. But a key point of the calculations is that relative to income and house-rents paid, there is an enormous surplus to play with to compensate residents. Perhaps in a 'just' world this would all be solved by giving tenants the land titles and allowing them to sell themselves when and if they are ready.

5. Conclusions

This paper examines building development and redevelopment in a growing city and the welfare costs of institutionally created land market frictions. We model the dynamics of a growing city in which formal buildings are durable, but informal are not. We develop propositions about the timing and spacing of developments in the city. Building volumes decline with distance from the centre; but increase over time, as successive redevelopments in the formal sector involve building taller. We take the model to a unique data set on the built environment of Nairobi for 2003 and 2015 and estimate key parameters to calibrate the model. We then formulate a measure of the welfare costs of institutional frictions in land markets, which plague many cities in the developing world.

For a large fast growing city like Nairobi, we find that in the core part of the city there is major redevelopment of 2003 formal sector buildings into taller new buildings, driving 50-60% increases in volume. However, while this dynamic development is occurring, there remain persistent slums in the mid-city, and development of slum into formal sector housing over the 12 years is very limited.

Applying our model indicates that the cost of this inefficient land-use is high. Even slumlords were to be paid off (compensating for perpetual control), conversion today would yield a surplus of $16-17,000 per slum household in context where they pay about $500-700 a year for their housing. Poorly functioning land market institutions dramatically alter the built fabric of the city, creating a hotchpotch of building heights and uses and significant welfare losses.

References

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, *forthcoming*

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 Neighborhood 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. and D. Da Mata. 2017. "On the determinants of slum formation". https://sites.google.com/site/danielddamata/research

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

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.

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

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.

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

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

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. “The Ndungu report: land and graft in Kenya”. Review of African Political Economy 32(103), 142-151, (2005).

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.

S. Williams, 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. 3-D image of Nairobi (2003 city)

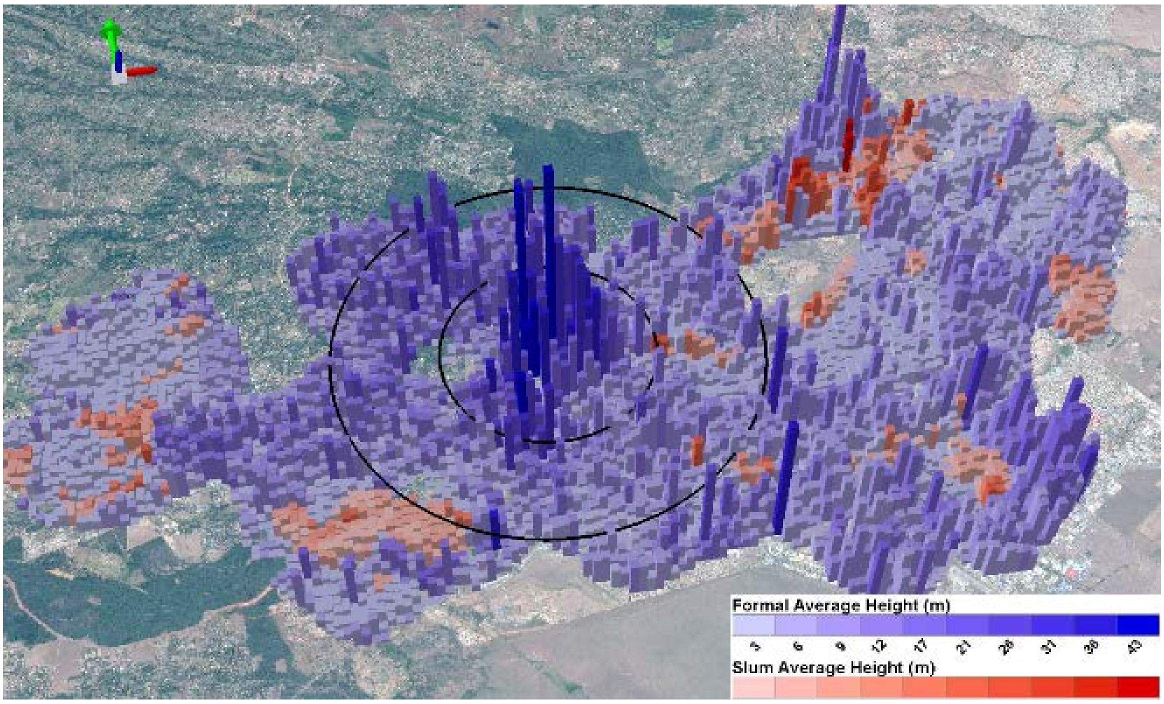


Figure 2: Urban development with perfect foresight

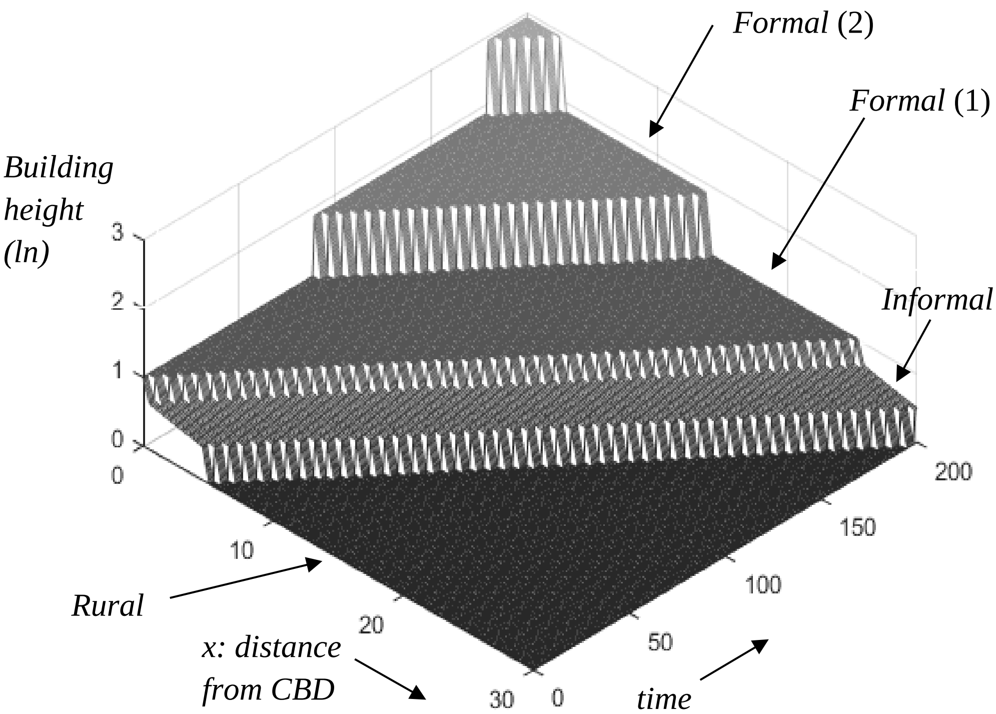


Figure 3: The hotchpotch: random variation in fixed costs of development, *D*

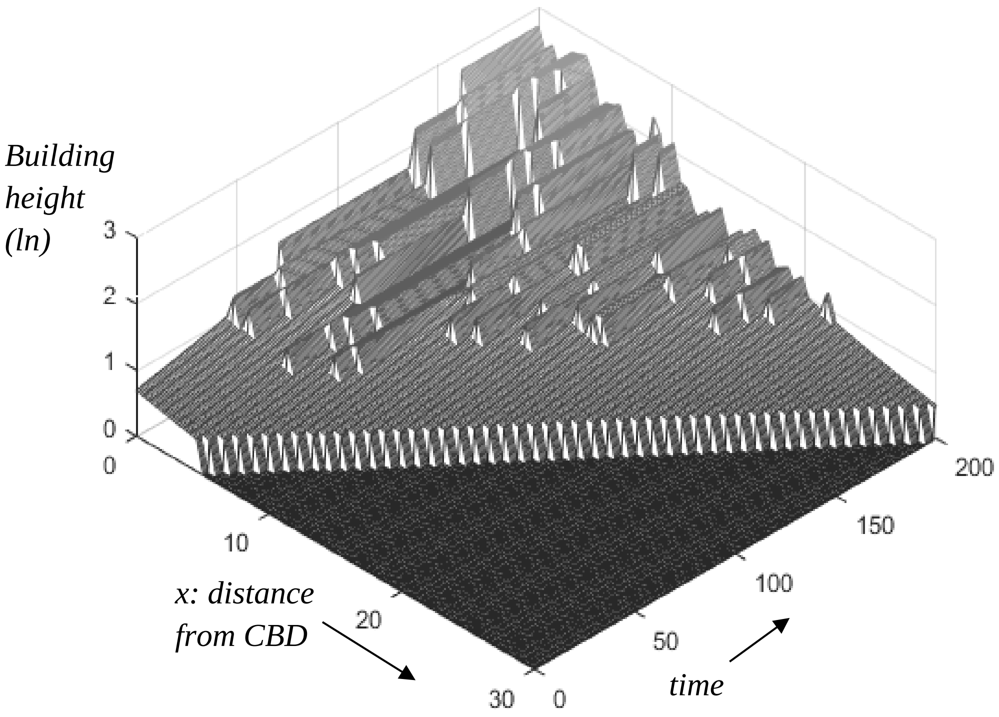


Figure 4. City shape and slums

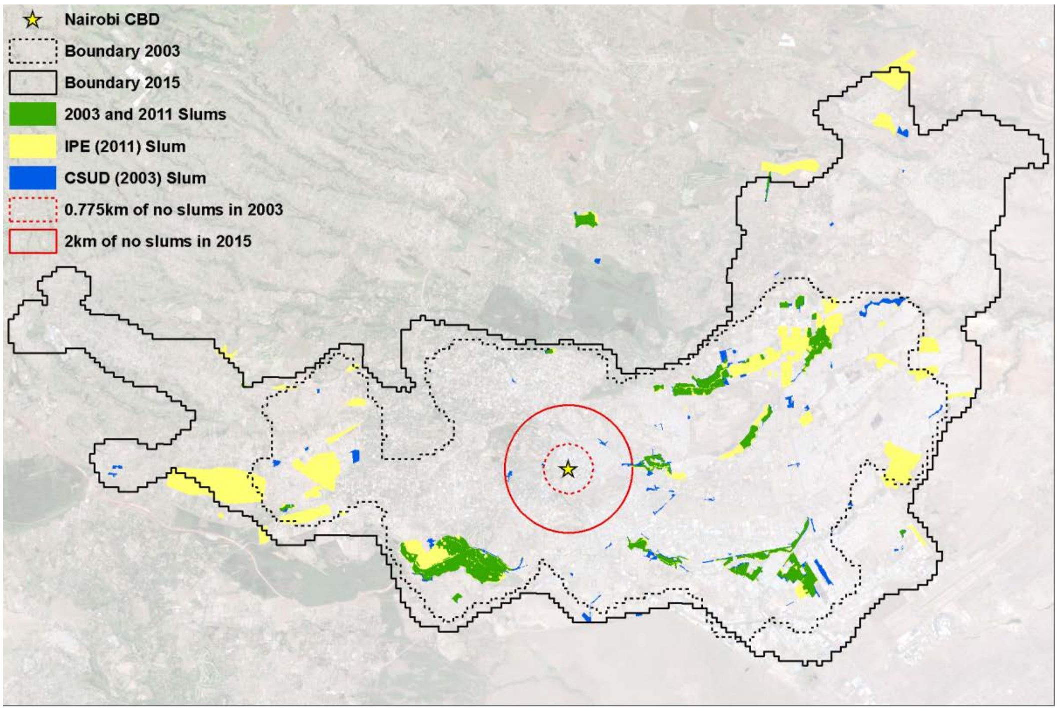


Figure 5. Built volume per unit area (BVAR):

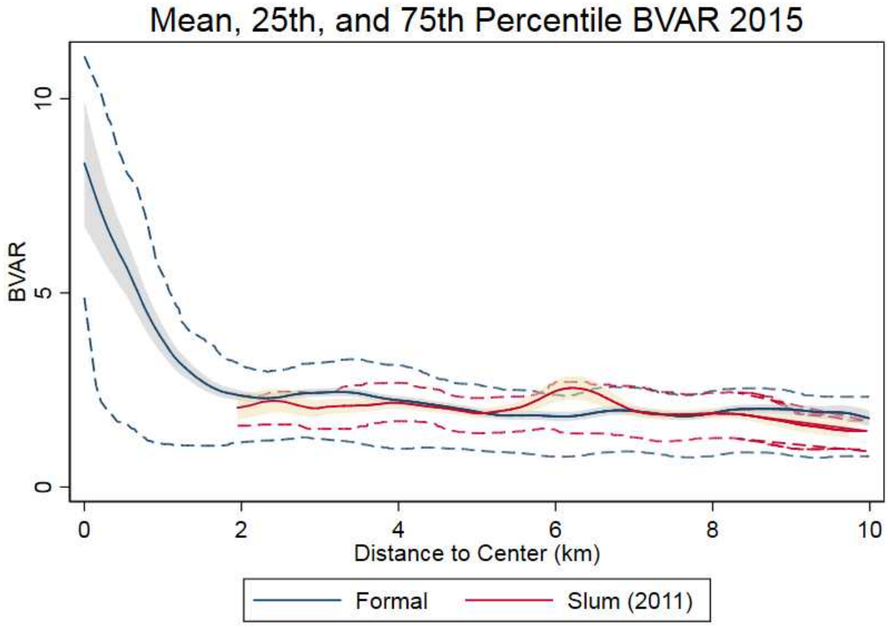


Figure 6. Cover to area ratio and height

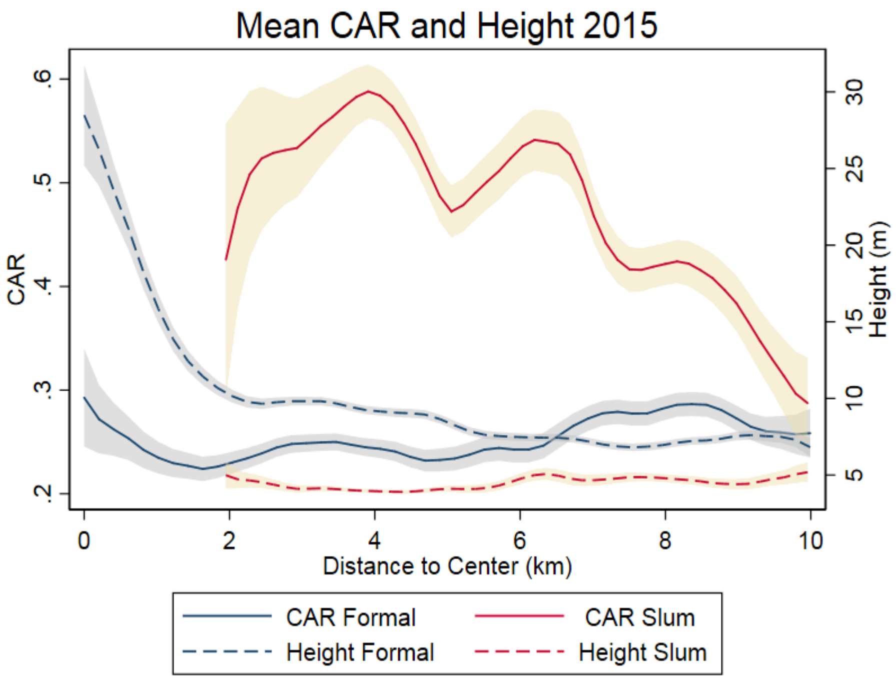


Figure 7. Formal sector changes in volume: Decomposition (2003 city out to 8kms.)

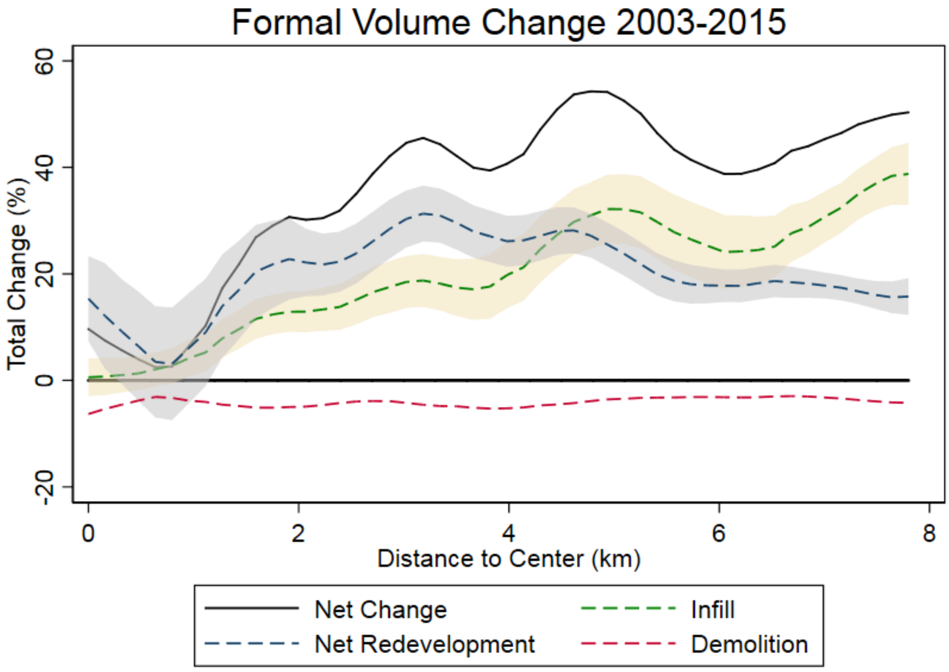


Figure 8. Mean height: Unchanged and redeveloped buildings

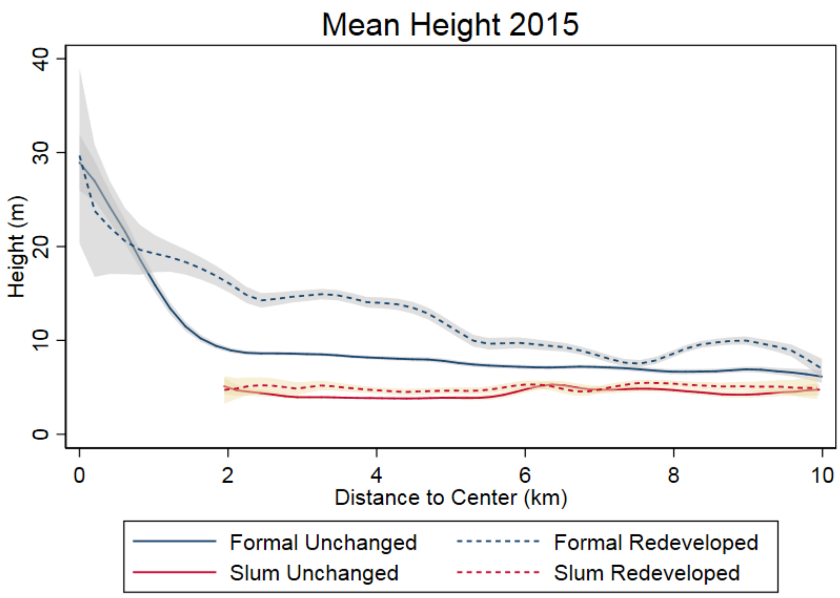


Figure 9. Slum volume change and decomposition out to 8kms.

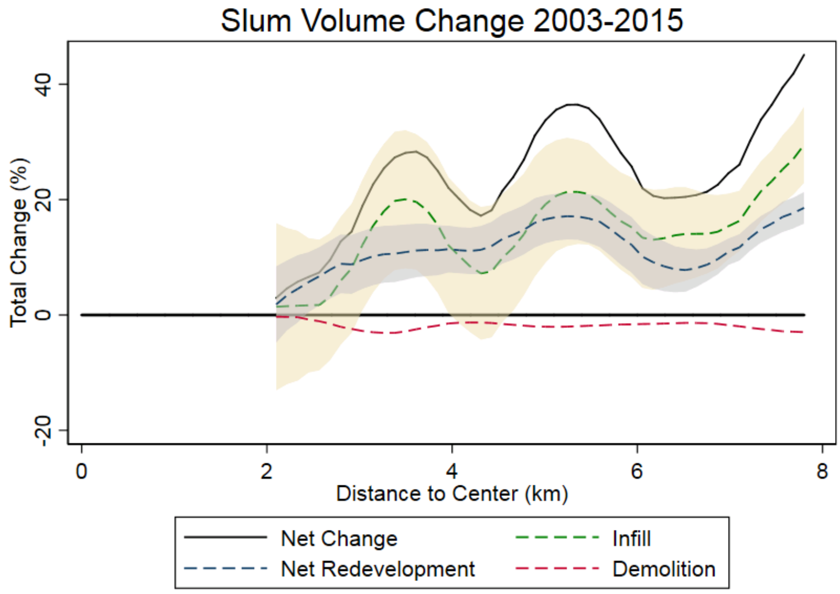
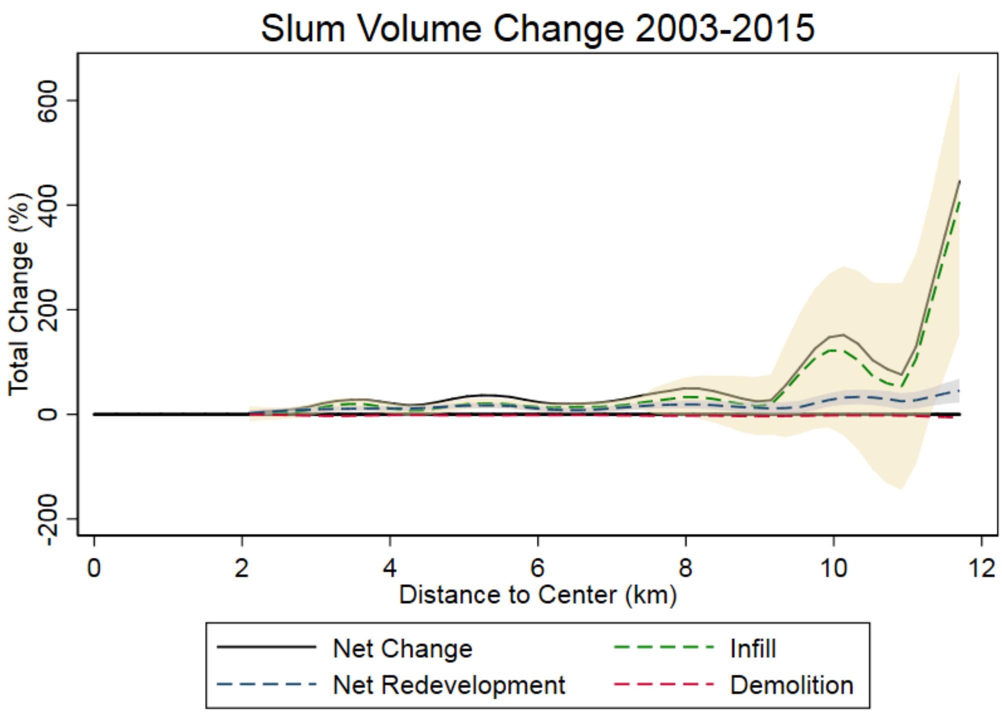


Figure 10. Total rates of volume change by distance and sector: (2015 city out to 12kms.)



On-Line Appendices

1. Theory Appendix:

1.1 Derivation of equations (13) & (14)

This derivation uses

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

$$= -p(x,t)v_F(x,\tau_i) + \rho\left[R_F(x,\tau_i) + k\left(v_F(x,\tau_i)\right)\right]$$

From eqn. (6), noting that volume is optimized.

1.2 Proof of Proposition 2

Part (i) is simply the total differential of (12a) and (18) with respect to *x* and *t*. Part (ii) comes from differentiation of (13a) with (5), (10) and (18); 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. (14a), noting that the house rent ratio in (14a) now compares prices at different *x* and the same *t*, $p(x_i(t),t)\,/\,p(x_{i+1}(t),t) = e^{-\theta\Delta x}$, where $\Delta x \equiv x_i(t) - x_{i+1}(t)$, i.e. the distance between places undergoing successive redevelopments.[32] As expected, comparison of (19) with the equation in part (iii) of proposition 2 gives $\Delta x / \Delta\tau = \hat{p} / \theta$, this indicating how the prices that trigger redevelopment relate across space and across time.

1.3 Parameters for figures:

Parameter values in Figs 2 and 3 are those derived from the Nairobi data as reported in Table 1, Appendix table 1.1, and section 3.3.4. Units on the horizontal axes can be interpreted as years and kilometres. The starting dates on the figure come from setting r_0 = 3 and starting price at date 0, *p*(0,0) = 23. In Fig 3, *D* takes positive parts of a normally distributed random variable with mean 0 and standard deviation = 150.

1.4 Closing the model:

The full open city equilibrium model underpinning the model of the text is as follows:

Households: At date *t* a representative urban household living at distance *x* from the CBD receives income net of commuting cost *w*(*t*)*T*(*x*), where *w*(*t*) is the wage at date *t* (the same for all households), and *T*(*x*) is the fraction remaining after commuting costs. Each household has Cobb-Douglas preferences between housing and other goods,

[32] Whereas in proposition 1, the price ratio in equation (14a) was evaluated at given *x*, so $p(x,\tau_i)\,/\,p(x,\tau_{i+1}) = e^{-\hat{p}\Delta\tau}$.

$$U(x,t) = \{a(x,t)s(x,t)\}^{\beta}\{w(t)T(x) - q(x,t)s(x,t)\}^{1-\beta}$$

where $s(x,t)$ is volume of housing, $q(x,t)$ is price, and $a(x,t)$ is housing amenity. The corresponding indirect utility function is

$$u(x,t) = \{a(x,t)/q(x,t)\}^{\beta} w(t)T(x)B, \quad B \equiv \beta^{\beta}(1-\beta)^{1-\beta}.$$

There are two types of housing, formal and informal ($i = F, I$). Amenity and price differ between the two types. Free choice of location and of housing type means that utility takes a common level, $\bar{u}$ for both types and across all occupied city locations. For formal sector housing, we set $a(x,t) = 1$ and we denote the bid-price $p(x,t)$, it taking the form

$$p(x,t) = q_F(x,t) = \{w(t)T(x)B/\bar{u}\}^{1/\beta}.$$

For informal sector housing $a(x,t)$ depends on crowding, as in eqn. 3. The bid-price is

$$q_I(x,t) = a(x,t)\{w(t)T(x)B/\bar{u}\}^{1/\beta} = a(v_I(x,t))p(x,t)$$

The bid-price for informal housing is therefore the product of the formal sector (unit quality) price and the amenity factor.

Constant exponential growth of the price of space is achieved by assuming that urban productivity and wages relative to outside utility (held constant at $\bar{u}$) grow at constant rate g. Similarly, constant exponential decline with respect to distance is achieved by the share of income net of commuting declining with distance at rate $\hat{T}$, so $p(x,t) = \left(\bar{w}e^{gt-\hat{T}x}/\bar{u}\right)^{1/\beta_i}$. This gives the unit quality price rising through time at constant rate $\hat{p} = g/\beta$, and declining with distance at rate, $\theta = -\hat{T}/\beta$.

Labour and population: To complete the model, we note that 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

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

The oldest formal development has been redeveloped the most times (which, at date t, we denote $i$max(t)). Notice that this expression assumes that the city is linear (or a set of rays), not a disc; modification to capture the latter is straightforward.

The final element is to close the model, either by setting $\bar{u}$ exogenously with $L(t)$ endogenous (open city), or with $L(t)$ exogenous and determining the equilibrium city wide level of utility (closed city). The analysis in the body of the paper follows the open city route, with exogenous growth of urban wages relative to outside utility driving housing price growth.

1.5 Other parameters.

Other than parameters reported in the text, we have levels of house-rents, construction costs in each sector and the informal sector amenity adjustment, $p(x,t), \kappa_I, \kappa_F, a_I$. Table 1 gives the relevant equations of the model for each sector, and intercept terms in Table 1 give values at x = 0 and t = 2015. Informal sector house-rents are from Table 3 which, given α, yields κ_I. Formal sector house-rents come from Table 3 and informal and formal sector volumes from Table 1. The parameters implied by these equations are given in the bottom row of the table.

Table 1.1 Levels and parameters: all evaluated at x = 0, $t = \tau_i$ = 2015, in $US		
	Informal: Eqns 1-4	Formal: Eqns 9, 18
House-rent per BVAR	$p(x,t)a\left(v_I(x,t)\right) = \alpha\kappa_I$ = 6.55	$p(x,t)$ = 23.19
BVAR: Volume per m^2	$v_I(x,t) = \left[a_I p(x,t) / \kappa_I \alpha\right]^{\frac{\alpha}{\alpha-1}}$ = 3.58	$v_F(x,\tau_i) = \left[\frac{p(x,\tau_i)\Phi(x,i)}{\kappa_F \gamma}\right]^{\frac{1}{\gamma-1}}$ = 6.80
Solutions: $p(x,t)$ = 23.19 , $a_I = 0.73$, $\kappa_I = 1.65$, $\kappa_F = 73$		

2. Data Appendix

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.

2.1 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 which 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) which was again produced by Ramani Geosystems. Without direct measurements of heights in 2003, we interpolate them 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 which contained 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 which 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 3.3 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. 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. 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.

2.2 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 which has been replaced by 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 which do now 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 150x150sqaure 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.

2.3 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 and 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 cu-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:

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

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 $r_{AB} = \frac{Area(A \cup B)}{Area(A)}$ 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:

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

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:

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

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:

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

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:

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

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 CBD. 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 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 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 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.

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

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

The confusion matrix in table 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 figure 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 figure 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.

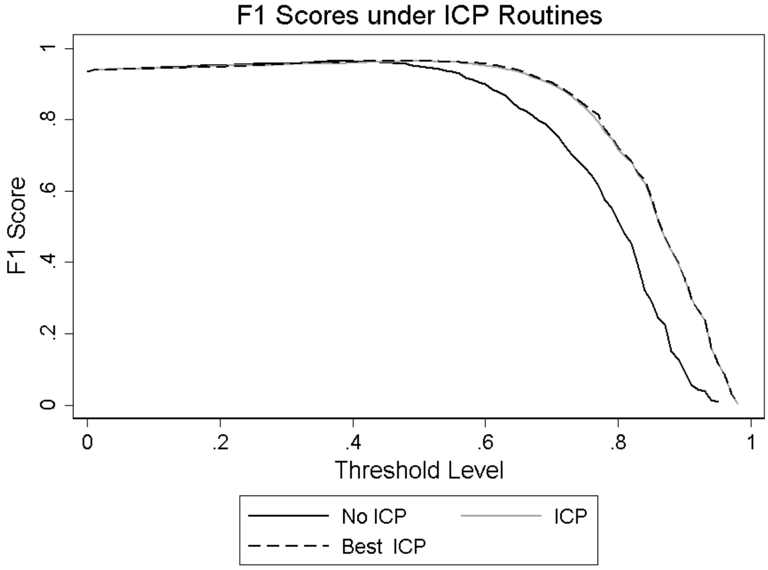
Appendix Table 2.1 – Mapping Correspondence 2003

a) Weighted by Building						
	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4
Manual=0	280	433	41	16	11	20
Manual=1	0	25	712	10	1	25
Manual=2	0	29	21	266	0	20
Manual=3	0	18	6	0	137	1
Manual=4	0	65	52	24	63	135
b) Weighted by Area (sq-m)						
	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4
Manual=0	12708	28187	4913	2780	943	1043
Manual=1	0	908	112762	4180	279	1775
Manual=2	0	3575	2328	89472	0	2819
Manual=3	0	910	1053	0	14148	23
Manual=4	0	5317	5528	4795	4464	14262
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.						

Appendix Table 2.2 – Matching all areas 2003

a) Weighted by Building			
	Algo=0	Algo=1	Recall
Manual=0	433	88	0.83
Manual=1	137	1473	0.91
Precision	0.76	0.94	F=0.93
b) Weighted by Area (sq-m)			
	Algo=0	Algo=1	Recall
Manual=0	28187	9679	0.74
Manual=1	10710	257888	0.96
Precision	0.72	0.96	F=0.96

Figure 2.1



Appendix Table 2.3: List of public uses	
Recreational a) Impala club, Kenya Harlequins, and Rugby Union of East Africa (0.14kmq) b) Golf Course (0.9kmq) c) Arboretum (0.25kmq) d) Central park, Uhuru park, railway club, railway golf course (0.5kmq) e) Nyayo stadium (0.1kmq) f) City park, Simba Union, Premier Club (1.1kmq) g) Barclays, Stima, KCB, Ruaraka, Utali clubs, and FOX drive in cinema (0.3kmq) Undeveloped a) Makdara Railway Yard (1kmq) b) John Michuki Memorial Park (0.1kmq) Special use -- Includes poorly traced areas a) State House b) Ministry of State for Defence c) Forces Memorial Hospital and Administration Police Camp d) Langata Army Barracks e) Armed Forces f) Moi Airbase	Public utility a) Dandora dump (0.5kmq) b) Sewage works (0.25kmq) g) Kahawa Garrison Public use a) Communications Commission of Kenya (0.1kmq) b) Langata Womens prison (0.2kmq) c) Nairobi and Kenyatta hospitals, Milimani Police Station, Civil Service club d) Mbagathi hospital, Kenya Medical Research Institute, Monalisa funeral home e) National museums of Kenya f) Kenya convention centre and railway museum g) Industrial area prison h) Mathari mental hospital, Mathare police station, traffic police, Kenya police, Ruaraka complex, and National youth service i) Jamahuri show ground Educational (not primary and secondary schools) 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

2.4 Other Tables: Rent regressions and welfare

Table 2.4: Full key regressions in Tables 1 and 3

	Ln land sales price (USD per sq m.)		Ln formal house-rent per cube vol. in $2015	Ln slum house-rent per cube vol. in $2015	Ln Formal redeveloped height; quantile: 80th percentile	Ln Slum BVAR
Distance to centre (km)	-0.172*** (0.0476)		-0.0557* (0.0308)	-0.00509 (0.0288)	-0.101*** (0.00521)	-0.0948*** (0.0104)
SD Elevation (km)	-0.0114 (0.0531)		0.00405 (0.0248)	-0.00813 (0.0227)	0.0222* (0.0116)	-0.0138 (0.0107)
Elevation (km)	0.00535*** (0.00178)		0.00199 (0.00120)	0.00355*** (0.000668)	321.7 (197.9)	-1142.5*** (236.8)
Lot size	-0.0302 (0.103)	No written tenancy	-0.316** (0.142)	-0.315 (0.218)		
Lot size X Lot size	-0.00108 (0.00159)	No piped water	-0.339** (0.139)	-0.274** (0.108)		
Coordinates estimated	-0.468 (0.372)	Ln # Floors	0.279** (0.117)	0.155 (0.138)		
		# Bathrooms				
Month=1	-0.306 (0.533)	One	-0.309** (0.144)	-0.0220 (0.104)		
Month=2	-0.269 (0.519)	Two+	-0.245 (0.178)	0.189 (0.120)		
		Structure type				
Month=3	-0.583 (0.482)	Single-story Shared facil.	-0.00640 (0.191)	0.279** (0.115)		
Month=4	-0.662 (0.481)	Multi-storey private bath	0.0634 (0.240)	-0.256 (0.217)		
Month=5	-0.862 (0.744)	Multi-storey shared bath	0.190 (0.209)	-0.0220 (0.104)		
Month=6	-0.195 (0.570)	Shared house		-0.382** (0.188)		
Month=7	-1.315* (0.703)	Room in house		-0.616*** (0.187)		

Month=8	-1.056**	Shack		-1.013***		
	(0.503)			(0.201)		
		Type of walls				
Month=9	-0.351	Brick/Block	0.433***	0.136		
	(0.502)		(0.147)	(0.262)		
Month=10	-0.617	Mud/Wood	0.637***	-0.689**		
	(0.526)		(0.190)	(0.307)		
Month=11	-0.280	Wood only	0.0847	-0.120		
	(0.519)		(0.176)	(0.277)		
Month=12	-0.677	Corrugated iron sheet	0.552***	-0.244		
	(0.481)		(0.204)	(0.251)		
		Mud and Cement	0.778***	-0.710**		
			(0.173)	(0.343)		
		Tin	0.777***	0.0115		
			(0.231)	(0.316)		
		Type of floor				
		Tiles	-0.0355	0.782***		
			(0.0953)	(0.295)		
		Cement	0.433***	0.146		
			(0.147)	(0.114)		
Constant	-1.328		-0.325	-3.427***	2.744***	3.163***
	(3.234)		(2.007)	(1.147)	(0.329)	(0.404)
Observations	136		361	442	4621	958
R-squared	0.292		0.307	0.406	-	0.104

Standard errors in parentheses
Sample is restricted to 2003 boundary
* p<0.10, ** p<0.5, *** p<0.01

Figure 2.2. Ratio of height of redeveloped to about to torn down buildings

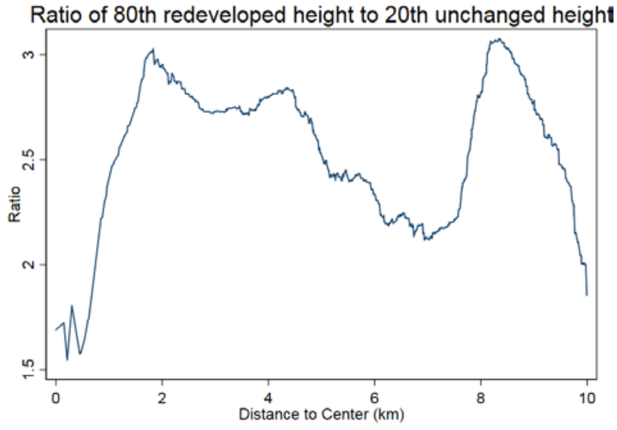


Figure 2.3. Teardown converted vs non-converted slums

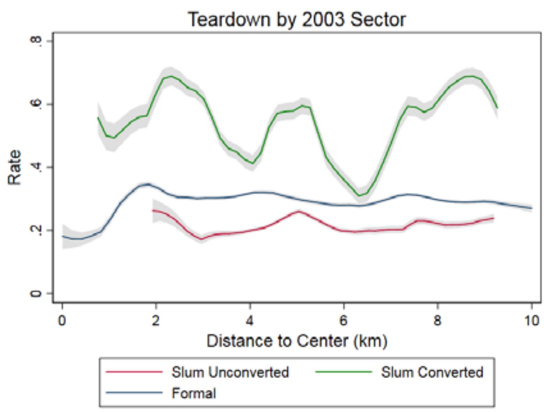
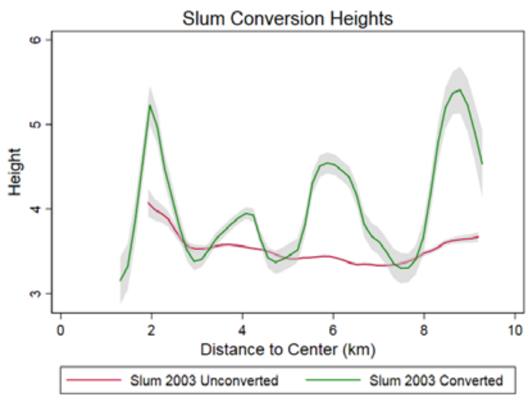


Figure 2.4. Heights in converted vs non-converted slums



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 landuse 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 infilled greenfield 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 why greenfield development occurs in areas near to the city centre, which according to the model should have been developed before those further out. In reality there is heterogeneity in land quality and therefore the costs to formal development, so despite higher demand for land nearer the centre development may occur further out where building is cheaper. In appendix Table 3.1 we show that greenfield development occurs on lower quality land, but this differential dissipates moving away from the centre. Near the centre, greenfield development tends to occur on land with higher elevation, nearer to water, and on land that is more rugged. While elevation may not necessarily make development more costly, greenfield also tends to be on land that is lower than other cells in its 450x450m window also dissipating over distance, however the differential gradient is insignificant and so only suggestive. Together this suggests that greenfield 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 3.2, we look at redeveloped heights and whether they vary with land quality. If during the initial formal development, a fixed 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. We find, oddly, that the standard deviation of elevation in the small scale (90x90m) is associated with higher redeveloped buildings.

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 comparatively worse land than neighbouring greenfield formal this could be partially responsible for the gap in land rents. We focus on greenfield land because, over the past period, it is the relevant comparison for what could have been slum redevelopment. In Table 3.3 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 greenfield 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 3.1: Quality of infilled land

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

Standard errors in parentheses are clustered by 2250x2250m arbitrary blocks
Sample is restricted to formal land inside ~~2004~~2003 boundary and 10km
* p<0.10; **p<0.05; ***p<0.01

Table 3.2: Redeveloped height and land quality

Distance to Center (km)	-0.0660***
	(0.00796)
Ln Elevation (m)	-0.000137
	(0.000265)
Water in 90x90m	-0.00499
	(0.0547)
S.D. elevation in 90x90m	0.0266***
	(0.00682)

S.D. elevation in 450x450m	0.00862 (0.00697)
Difference in elevation to mean of 450x450m	0.00410 (0.00288)
Constant	2.331*** (0.439)
Observations	33701
R-squared	0.053

Standard errors in parentheses are clustered by 2250x2250m arbitrary blocks
Sample is restricted to formal land inside ~~2004~~2003 boundary and 10km
* p<0.10; **p<0.05; ***p<0.01

Table 3.3 Comparison of slum land on one side of border to formal sector greenfield land on the other

	Elevation (m)	Water in 90x90m	S.D. elevation in 90x90m	S.D. elevation in 150x150m	Difference in elevation 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	0.0562	1.825**	2.298**	0.0277

	(3.747)	(0.310)	(0.844)	(1.025)	(0.907)
2-3 # Slum in 2015=1	-0.449	0.161	-0.0351	-0.261**	0.725*
	(1.353)	(0.149)	(0.129)	(0.116)	(0.372)
3-4 # Slum in 2015=1	-0.799	0.0844	0.0265	-0.125	0.0183
	(1.237)	(0.0828)	(0.301)	(0.296)	(0.217)
4-5 # Slum in 2015=1	-0.468	0.291***	-0.144	0.321	-0.157
	(0.603)	(0.0819)	(0.449)	(0.554)	(0.312)
5-6 # Slum in 2015=1	-0.561	0.0597	-0.0597	-0.183	0.315
	(1.434)	(0.0738)	(0.320)	(0.426)	(0.280)
6-7 # Slum in 2015=1	0.211	-0.194***	0.326**	0.408***	0.510***
	(1.120)	(0.0666)	(0.132)	(0.153)	(0.170)
7-8 # Slum in 2015=1	0.525	-0.0246	-0.103	-0.337	-0.0156
	(0.561)	(0.0720)	(0.207)	(0.405)	(0.181)
8-9 # Slum in 2015=1	-1.215	0.107	-0.696	-0.903	0.330
	(1.688)	(0.0729)	(0.589)	(0.593)	(0.351)
9-10 # Slum in 2015=1	-0.317	-0.0542	-0.157	-0.136	0.638
	(1.604)	(0.0452)	(0.286)	(0.369)	(0.399)
Observations	10167	10167	10167	10167	10167
R-squared	0.995	0.416	0.369	0.584	0.081

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

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].

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