21/11/2017

Building the city: urban transition and institutional frictions

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

We model the building of cities and calibrate the model. We distinguish formal and slum construction. In contrast to slums, formal structures can be built tall, are durable, and non-malleable. With city growth, areas are initially developed informally, then formally, and then redeveloped periodically. Formalisation costs 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 calibrate the model. In older slums, even after paying off slumlords, formalisation nets real income gains amounting to $16-17,000 per slum household who currently pay rents of $600-700 annually.

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, 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 most parts of Sub-Saharan Africa the populations of many cities, especially primate ones, 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 2/3 of the non-governmental capital stock in countries 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 and rapid redeployment of nations' capital stocks. Some parts of cities are subject to rapid change and redevelopment; in some development takes the form of slums; and in some formal sector high rise buildings are adjacent to low level slums, suggesting inefficient land-use and distorted investment decisions.

This paper addresses these issues by doing three things. First, we develop a formal model of the built environment of a growing city with both formal sector and slum housing. Second, we develop a novel and extremely rich data set of the built fabric of Nairobi and use it to show how building volume varies across the city, and how this pattern has changed over a 12 year period. Third, we use this data to calibrate the model and thereby produce estimates of the real income cost of inefficient land-use, in particular the persistence of slums in the core areas of the city.

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. Investment decisions are therefore based on the 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 city building volume delivered by slums increases not by building taller, but by already high cover-to-area ratios increasing through time. New slums appear near the expanding city edge and within the city there is conversion of slums to formal development.

These predictions from the model accord well with Nairobi experience, except that conversion of slums to formal sector usage is limited and patchy. 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. We capture this in the model by the existence of place specific 'formalisation' costs that inhibit formalisation. We discuss what the costs may be and, following calibration of the model, bound their size and calculate the implied real income loss they cause.

The second element of our study is the Nairobi data set. The core is tracings of all buildings in the city from aerial photo images for 2003 and 2015 which gives 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 and map roads. 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.

The third element of the study combines these data sources to paint a rich picture of the development of the built fabric of Nairobi, and to calibrate our theoretical model and quantify inefficiencies in land use. With the calibrated model we calculate the dates when areas of the city would have first gone from informal to formal sector usage in the absence of formalisation costs, and when initial formal sector developments would have been torn down and replaced with taller buildings. We calculate the lost real income losses from delayed conversion of slum to formal sector buildings, especially in the urban core.

Our own findings and other sources give the following picture of Nairobi. Nairobi, while far from circular, appears to be 'monocentric' with building height, volume, and land prices all declining with distance from the centre. In Fig. 1 we show this for 2015 height with a 3-D map of the city for the 2003 built area of the city. The city centre with radial circles at 2 and 4km is where the highest buildings are (and is 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 grid square.[2] The city appears monocentric with tall but variable height at the centre and then diminishing. Slum areas in red are based on a 2011 slum mapping by IPE Global Limited as described later. Slums are generally low height as assumed in modelling. 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. Nairobi's sausage shape 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.

[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.

[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).

Nairobi's monocentricity is in line with other African cities such as Addis Abba, Dar es Salaam, and Kigali, which have land use intensity gradients that decline fairly smoothly with distance from the centre, without the flat portions or dual peaks seen in Paris, London, or Barcelona, let alone American cities like Houston or Atlanta (Lall, Henderson, and Venables 2017, p 21-23). For a sample of about 320 Africa cities, Baruah, Henderson and Peng (2017) find that land cover generally declines sharply with distance from the centre, again without multiple peaks. It seems many African cities have the monocentricity that many developed country cities started losing decades ago.

The city population was given as 3.1mn in the 2009 census, growing at 3.9% pa from 1999, while the greater metropolitan area has a population in excess of 6.5mn. Our estimates suggest that the built volume of the city increased at the same rate as population (3.9% pa), expanding by 59% between 2003 and 2015 for the 2015 built area of the city (as defined later). Within the 2003 built area growth was 48%, and at the extensive margin between the 2003 and 2015 built areas growth was 120%. While slum volume increased, its growth at 55% was slower than the formal sector at 60% and its share fell modestly.

To accomplish this increase in volume, the rates of destruction and creation in the core area of Nairobi have been enormous. For example, at just 3 km from the centre, 35% of buildings were torn down in the last 12 years, an astonishing number by developed country standards where 10% would be high. At 3km from the centre, demolition goes with redevelopment to much taller buildings. That is largely responsible for the 60% increase in volume of building volume at that distance over the 12 years.

This rapid change contrasts with the persistence of slum areas near the centre, most notoriously 'Kibera', the large slum area to the southwest of the centre and often referred to as Africa's largest slum. There are a number of obstacles to formalising slum lands. Formalisation means acquiring the private property rights essential for capital intensive formal sector development, to mitigate the chances of expropriation and allow property to be collateral for financing of construction. Formalisation costs arise from history and each African city and areas within cities have their own histories. Nairobi is a colonial origin city founded in 1899 by the British as a rail depot. The colonial government used segregation policies (Olima, 2001) to control the African population, containing them to informal settlements without title. In the decades after independence (1963) a series of 'reforms' resulted in 85% of land in Nairobi becoming privately owned (under charges of widespread corruption, Southall, 2005). However older slums remain mostly 'government owned' out to about 5 km from the centre, meaning they are run by slumlords operating illegally or quasi-legally, with the government taking a hands off approach since the 1970's. In much of these areas there has been little slum redevelopment. Based on an analysis of the literature and examples later in the paper, we argue that this persistence of older slums near the centre is largely due to formalisation costs that are a product of legal and institutional forces particular to the history of each slum (Bird et. al. 2017). Using our calibrated model we estimate that, in the older slums nearer the city centre, even after paying slumlords for the value of their land in perpetual slum use, formalisation would bring a gain of about $16,000-17000 per slum household, in a context where slum households spend on average about $500-700 pa on housing. We estimate a lower bound on formalisation costs which is very high, typically about the value of land held in slum use in perpetuity.

In a recent paper on Brazilian slums, reflecting other work on Latin America, Cavalcanti and da Mata (2017) model slum dwellers as '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. In Nairobi the share of slum land that is privately owned (titled) after 5km rises to 40%, which is one basis for our modelling of slums as more of a technology choice, regardless of land tenure. Also in Nairobi, similar to many Africa cities,[3] the vast majority of all residents and 90% of slum dwellers rent. But the issues are not fundamentally different from Latin America. 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. And 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.[4]

There are four novel aspects to the paper. First is the modelling. While Braid (2001) has a dynamic monocentric model with durable capital, no dynamic model deals with informality and formalisation costs. Second are the data. While there is work on the USA using demographic census data to examine building ages (Brueckner and Rosenthal 2009), no work we know of utilizes city-wide data on individual buildings, with demolition, redevelopment, and infill to detail the changes in the urban landscape. Third, by calibrating the model, we develop a methodology to calculate the welfare cost of institutional frictions. This highlights the role of policy for fast growing cities with major market land failures that deter investment, because of lack of transparency and weak institutions governing land markets. Finally, we deliver a new set of facts about development and redevelopment of the built environment for a major developing country city.

---

[3] Addis Abba, 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] 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).

The paper is organised as follows. The basic model and core theoretical results are set out in section 2. Section 3 presents data, looks at the urban cross-section and patterns of change, and provides estimates of key empirical relationships used to back out parameters needed to calibrate the model. Section 4 does the calibration. Section 5 develops the welfare measures used to characterize misallocation of land between older Nairobi slums and the formal sector. Section 6 discusses additional features of the city and Section 7 concludes.

2. Theory

In this section we 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 appendix we show how this can be endogenised in an open city equilibrium. Section 2.1 analyses building decisions associated with each of 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 shows how this path varies across points in the city, giving a complete description of both the cross-section of the city and its evolution through time. Section 2.4 addresses expectations and closing the model. We will note throughout critical relationships for later calibration.

2.1 Building technology and housing supply

There are two distinct building technologies, formal and informal, which supply building volume per unit land in different ways. The formal sector (*F*) can build tall, and the informal sector (*I*) can 'crowd', increasing cover, the proportion of land that is covered by building footprint. The volume of building delivered on a unit of land at a particular place, *x*, and time *t*, is the product of height and cover, $v_i(x,t) = h_i(x,t)c_i(x,t)$, $i = I, F$.

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. The informal 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 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 rent (and willingness to pay) for a unit of informal housing is the product of two elements; the rent of informal housing of unit quality at place x at date t, $p_I(x,t)$, and a quality or amenity factor, $a(v_I(x,t))$, 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

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

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

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

If informal house quality is iso-elastic in cover, $a(v_I(x,t)) = a_I v_I(x,t)^{(1-\alpha)/\alpha}$, $\alpha > 1$, then the first order condition is $p_I(x,t)a(v_I(x,t)) = \alpha\kappa_I$ and optimally chosen volume and maximised land-rent are respectively

$$v_I(x,t) = [a_I p_I(x,t)/\kappa_I\alpha]^{\frac{\alpha}{\alpha-1}}, \quad r_I(x,t) = \kappa_I(\alpha-1)[a_I p_I(x,t)/\kappa_I\alpha]^{\frac{\alpha}{\alpha-1}}. \tag{3}$$

The first order condition (with iso-elasticity) implies that informal sector house-rent adjusted for quality is constant throughout the city, $p_I(x,t)a(v_I(x,t)) = \alpha\kappa_I$. Essentially, and as we will see later in the data, increased crowding near the centre will offset the advantage of improved access to the city centre. The degree of diseconomy in crowding, α, is a key parameter and will be estimated based on the informal sector volume gradient, measuring how informal volume per unit area varies with house rent and associated location. Iso-elasticity also implies that construction costs and land-rent are respectively share $1/\alpha$ and $(1-1/\alpha)$ of revenue earned by informal sector housing, so

$$r_I(x,t) = [1 - 1/\alpha]p_I(x,t)a(v_I(x,t))v_I(x,t). \tag{4}$$

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, we assume that formal sector land cover is not a choice variable but is set exogenously at $c_F = 1$, and that volume is achieved by choice of height, $v_F(x,\tau_i) = h_F(x,\tau_i)$. Height is chosen at date of construction, denoted τ_i, and then 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

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

structures. Construction costs per unit land are one-off and sunk, and are an increasing and convex function of building volume on that land, $k(v(x,\tau_i))$, 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. Both sets of materials are not sufficiently load bearing to allow much in the way of height. In contrast, over 90% of formal sector housing is made of stone or some type of brick/block.[6]

We assume there is no amenity loss or gain from building tall so the house-rent of a unit of formal sector building volume, $p_F(x,t)$, is exogenous to the developer, and is place and time specific. 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 $k(v(x,\tau_i))$ sunk and volume fixed at the date of construction this is given by

$$R_F(x,\tau_i) = \int_{\tau_i}^{\tau_{i+1}} p_F(x,t) v_F(x,\tau_i) e^{-\rho(t-\tau_i)} dt - k(v_F(x,\tau_i)). \qquad (5)$$

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

$$\Phi(x,i) \equiv \int_{\tau_i}^{\tau_{i+1}} [p_F(x,t) / p_F(x,\tau_i)] e^{-\rho(t-\tau_i)} dt, \qquad (6)$$

so $R_F(x,\tau_i) = p_F(x,\tau_i)\Phi(x,i)v_F(x,\tau_i) - k(v_F(x,\tau_i))$. The integral $\Phi(x,i)$ is akin to the 'value-to-rent ratio' on a newly constructed property in the terminology of the real-estate literature (noting the time horizon in (6) is cut at τ_{i+1}).

The first order condition for choice of volume is

$$\partial R_F(x,\tau_i) / \partial v_F(x,\tau_i) = p_F(x,\tau_i)\Phi(x,i) - k'(v_F(x,\tau_i)) = 0 \ . \qquad (7)$$

[6] We note that some on-going studies focus on classifying slums by the use of corrugated iron for roofs. This would not work in Nairobi, where over 50% of formal sector residential buildings also have corrugated iron sheet roofs (88% in slums).

If the cost function is iso-elastic, $k(v_F) = \kappa_F v_F{}^{\gamma}$, $\gamma > 1$, then (5) – (7) imply that two key relationships we use later, the volume at construction and the maximised present value of land-rent, are[7]

$$v_F(x,\tau_i) = \left[\frac{p_F(x,\tau_i)\Phi(x,i)}{\kappa_F \gamma}\right]^{\frac{1}{\gamma-1}}, \qquad R_F(x,\tau_i) = \kappa_F(\gamma-1)\left[\frac{p_F(x,\tau_i)\Phi(x,i)}{\kappa_F \gamma}\right]^{\frac{\gamma}{\gamma-1}}. \qquad (8)$$

The diseconomy in building high, γ, is another key parameter which will be estimated based on how the height of newly constructed buildings varies with location as house rent varies.

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 then costs are fully covered by setting μ to satisfy $\mu p_F(x,\tau_i)\Phi(x,i)v_F(x,\tau_i) = k(v_F(x,\tau_i))$. With $k(v_F) = \kappa_F v_F{}^{\gamma}$ and (8) the amortization rate is $\mu = 1/\gamma$.[8] Flow land-rent, $r_F(x,t,\tau_i)$, defined as gross revenue net of amortized costs, is therefore

$$r_F(x,t,\tau_i) = [1-1/\gamma]p_F(x,t)v_F(x,\tau_i). \qquad (9a)$$

Notice that $r_F(x,t,\tau_i)$ depends on location, time, and date of construction, and that the present value of land-rent over the lifetime of a formal sector building can be written as[9]

$$R_F(x,\tau_i) = \Phi(x,i)r_F(x,\tau_i,\tau_i). \qquad (9b)$$

The flow land-rent net of amortization is fraction (1-1/γ) of revenue earned by land and structure together, while in the informal sector land-rent is fraction (1-1/α) of revenue (eqn. 4). We will use these relationships in calibration.

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 x that has not yet been developed is

[7] The iso-elastic form implies an elasticity of substitution between land and capital (i.e. construction cost) of unity. This is higher than many estimates in the literature, although at the centre of the range suggested in by Ahlfeldt and Mcmillen (2014).

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

[9] To see this, take the ratio of eqns. (8), $R_F(x,\tau_i)/v_F(x,\tau_i) = p_F(x,\tau_i)\Phi(x,i)(\gamma-1)/\gamma$ and use (9).

$$PV(x) = \int_0^{\tau_0} r_0 e^{-\rho t} dt + \int_{\tau_0}^{\tau_1} r_I(x,t) e^{-\rho t} dt + [R_F(x,\tau_1) - D(x)] e^{-\rho\tau_1} + \sum_{i=2} R_F(x,\tau_i) e^{-\rho\tau_i} . \quad (10)$$

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 calculated 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 one-time fixed cost $D(x)$ of converting to formality. [10] The final term in (10) gives the discounted value of land-rents earned over the lives of consecutive formal sector buildings, constructed at dates τ_2, τ_3... The land-rent terms in this expression depend on house-rents per unit volume $p_I(x,t)$ and $p_F(x,t)$ (eqns. (3) and (8)), which we assume to be monotonically increasing and exogenous in x and t.

The formalisation cost, $D(x)$, captures the fact that formal sector development requires reasonably well defined property rights, such as land titling or a formal leasehold system like in Hong Kong. Obstacles to obtaining these rights on some properties may be substantial, particularly in African countries where much land is held traditionally under possessory and communal rights. $D(x)$ includes the cost of obtaining formal title, which is highly variable even within a city depending on the history of the local area, as discussed later. As noted in the Introduction, substantial portions of slums especially nearer the city edge are on private (titled) land, so slums are not coincident with lack of land title. Slums nearer the city centre are classified as 'government owned' which later we interpret as being clouded by rights issues, political influence and corruption, which are costly to overturn.

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

$$\frac{\partial PV(x)}{\partial \tau_0} = e^{-\rho\tau_0}[r_0 - r_I(x,\tau_0)] = 0 . \quad (11)$$

The first formal development takes place at date τ_1 satisfying

$$\frac{\partial PV(x)}{\partial \tau_1} = e^{-\rho\tau_1}[r_I(x,\tau_1) - p_F(x,\tau_1)v_F(x,\tau_1) + \rho\{k(v_F(x,\tau_1)) + D(x)\}] = 0 . \quad (12)$$

(see appendix). A necessary condition for this to have an interior solution is that $0 < p_F(x,\tau_1)v_F(x,\tau_1) - \rho k(v_F(x,\tau_1)) = p_F(x,\tau_1)v_F(x,\tau_1)[\gamma - \rho\Phi(x,i)]/\gamma$, the second part of which

[10] For simplicity, we do not let this depend on time. The dependence on location is drawn out in section 2.4.

uses $k(v_F) = \kappa_F v_F^{\gamma}$ and (8). We assume henceforth that this inequality holds. In Section 5, we will use eqn. (12) as an inequality to estimate a lower bound on formalisation costs.

The first redevelopment of formal land is at date τ_2 satisfying

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

(see appendix). Generalising this for all redevelopments gives:

$$p_F(x,\tau_{i+1})\left[v_F(x,\tau_{i+1}) - v_F(x,\tau_i)\right] = \rho k(v_F(x,\tau_{i+1})), \quad \text{for } i \geq 1. \tag{13}$$

This condition says that demolition and reconstruction occur at the date at which the revenue gain from the change in volume equals the interest cost of the construction expenditure incurred. Similar intuition applies to eqn. (12).

Eqns. (11) – (13) implicitly define the dates at which sites are (re-)developed. Using our iso-elastic functional forms and the optimised values of $v(x,t)$ given by eqns. (3) and (8), the house-rents that trigger development and hence the dates of development are given by the following eqns. (11a) – (13a). The date at which site *x* becomes informally developed, τ_0, is implicitly defined by

$$p_I(x,\tau_0) = \frac{\kappa_I \alpha}{a_I}\left[\frac{r_0}{(\alpha-1)\kappa_I}\right]^{(1-1/\alpha)}. \tag{11a}$$

The right hand side of this expression is constant, and can be thought of as giving a trigger value; location *x* becomes informally developed on the date at which house-rent at *x* reaches this trigger level.

The date at which informal settlement becomes formalised, τ_1, is given by eqn. (12) which using (3) and (8) becomes

$$\kappa_I(\alpha-1)\left[\frac{a_I p_I(x,\tau_1)}{\kappa_I \alpha}\right]^{\frac{\alpha}{\alpha-1}} = \kappa_F\left(\frac{\gamma}{\Phi(x,1)} - \rho\right)\left[\frac{p_F(x,\tau_1)\Phi(x,1)}{\kappa_F \gamma}\right]^{\frac{\gamma}{\gamma-1}} - \rho D(x)\,. \tag{12a}$$

The dates at which successive formal redevelopments of *x* take place, τ_i, $i > 1$, become

$$\left[\frac{p_F(x,\tau_i)\Phi(x,i)}{p_F(x,\tau_{i+1})\Phi(x,i+1)}\right]^{\frac{1}{\gamma-1}} = \frac{\gamma - \rho\Phi(x,i+1)}{\gamma}\,. \tag{13a}$$

These three equations, (11a)-(13a) together with the definition of the value-to-rent ratio, $\Phi(x,i)$ in eqn. (6), 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, $\hat{p}_I, \hat{p}_F > 0$ yields analytical results. The full general equilibrium model that supports constant exponential price growth is discussed in section 2.6 and detailed in the Theory Appendix, but for the present 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

To draw out results we look first at successive redevelopments of formal areas of the city, and then turn to the city edge and informal development.

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

$$\Phi = \int_0^{\Delta\tau} e^{(\hat{p}_F - \rho)t} dt = \frac{1 - e^{(\hat{p}_F - \rho)\Delta\tau}}{\rho - \hat{p}_F}, \qquad \Delta\tau = \frac{(\gamma - 1)}{\hat{p}_F} \ln\left[\frac{\gamma}{\gamma - \rho\Phi}\right]. \tag{14}$$

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

$$\frac{v_F(x, \tau_{i+1})}{v_F(x, \tau_i)} = e^{\frac{\hat{p}_F \Delta\tau}{(\gamma - 1)}} = \frac{\gamma}{\gamma - \rho\Phi}. \tag{15}$$

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

The first part of this proposition comes from integrating eqn. (7), using it in (13a), and noting that there is a unique solution solving the two parts of (14) with constant Φ and $\Delta\tau$ over time and space. The second part follows by using this in the first order condition for volume, (8). The third comes from noting that (14) and (15) 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 time τ_i with a new building is given by

$$PV_F(x,\tau_i) = r_F(x,\tau_i,\tau_i)\Phi / \left[1 - e^{-(\rho - \hat{p}_F\gamma/(\gamma-1))\Delta\tau}\right] \qquad (10a)$$

This is derived from eqn. (9b), $R_F(x,\tau_i) = \Phi(x,i) r_F(x,\tau_i,\tau_i)$, which gives the present value of land-rent over the life of a single development; the denominator in (10a) extends this to infinitely repeated cycles of redevelopment. This relationship between current rent and capital value (reflecting future rent increases) will be used in Section 4 in backing out an estimate of $\hat{p}_F$, the rate of increase in housing rents.

What about the earlier stages of informal development? The first transition we assumed is from agriculture to informal settlement. For land at *x* this occurs at date τ_0 when $p_I(x,t)$, the quality un-adjusted informal sector house-rent, reaches the trigger value given by (11a).[11] The transition from informal to formal settlement is given by date τ_1 that solves (12a). There is a unique transition date satisfying the second order condition if the return to formal development is rising faster than the return to informal settlement (i.e. the right hand side of (12a) is increasing faster than the left). If $D = 0$, a necessary and sufficient condition for this is $\hat{p}_F\gamma/(\gamma-1) > \hat{p}_I\alpha/(\alpha-1)$. If $D > 0$, then this condition is sufficient but not necessary. We assume the condition to be satisfied, as it will be if house-rents (before being deflated for crowding) increase at the same rate and $\alpha > \gamma$, or there are sharper diseconomies in informal sector crowding than is formal sector building height. The condition $\alpha > \gamma$ means also that the share of land-rent in revenue is higher (and the share of construction costs lower) in informal development than in formal (eqns. (4), (9)). Our empirical work will show that the condition is readily satisfied in Nairobi.

Figure 2 pulls these stages together and illustrates the development path, using model parameters estimated and calibrated in Section 4. 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, i.e. fix *x* and look along a line sloping up and to the right with 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_I causes Meccano pieces to be rearranged and building cover to increase. Formal development takes place at τ_1 and, as

[11] A period of informal settlement exists only if the return to informal settlement at date τ_0 is greater than commencing formal settlement, $r_I(x,\tau_0) > p_F(x,\tau_0)v_F(x,\tau_0) - \rho\ k_F(v_F(x,\tau_0)) + D(x)$ (see eqn. (12)). If not, then initial development will be formal, with date τ_1 implicitly defined by $r_0 = p_F(x,\tau_1)v_F(x,\tau_1) - \rho\ k_F(v_F(x,\tau_1)) + D(x)$.

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 depends on *x*. Henceforth, *x* is interpreted as distance from the CBD, and house-rents decrease with distance. We assume exogenous rates of decrease at exponential rates θ_I, θ_F, and the appendix derives these from commuting costs. Exponential decline with respect to distance together with exponential growth through time mean house-rents are

$$p_I(x,t) = \bar{p}_I e^{\hat{p}_I t} e^{-\theta_I x}, \quad p_F(x,t) = \bar{p}_F e^{\hat{p}_F t} e^{-\theta_F x} \qquad (16)$$

Given these, the trigger house-rent for informal development in eqn. (11a) depends on both date and place according to

$$p_I(x,t) = \bar{p}_I e^{\hat{p}_I t} e^{-\theta_I x} = \frac{\kappa_I \alpha}{a_0} \left[\frac{r_0}{(\alpha - 1)\kappa_I} \right]^{(1-1/\alpha)}. \qquad (11b)$$

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. (12) and (12a) can be interpreted as giving the distance at which first formalisation occurs at date *t*, $x_1(t)$, and (13a) the distances between areas of redevelopment occurring at date *t*, $x_i(t)$, $i > 1$. Eqn. (12a) indicates there are two cases, one where formalization costs $D(\cdot) = 0$ everywhere (e.g., all land in a city is privately owned so there are no formalisation costs). In this case the city has a continuous and easily described evolution. Different phases of development (informal, first formal, second formal etc.) are in rings of constant width which are continuously moving outward as the city grows. In the second case $D(\cdot) \geq 0$ and takes different values throughout the city, ultimately leading to a hotchpotch. We look at each in turn.

2.3.2.1 No formalisation costs

Proposition 2 states how different stages of development (building types and heights) vary across the 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 $\rho > \hat{p}_I$, $\hat{p}_F > 0$ and declining with distance at constant rates $\theta_I, \theta_F > 0$, 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 $dx_0 / dt = \hat{p}_I / \theta_I$.

(ii) If $D(x)$ =0, the distance from the city centre to the edge of formal development increases through time according to $\dfrac{dx_1}{dt} = \dfrac{\hat{p}_F \gamma(\alpha - 1) - \hat{p}_I \alpha(\gamma - 1)}{\theta_F \gamma(\alpha - 1) - \theta_I \alpha(\gamma - 1)}$.

(iii) The distance between successive formal sector redevelopments, Δx, is constant,

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

See Appendix for proof

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 first formal sector development is pushing continuously into the inner edge of the informal sector ring. Part (iii) then implies that redevelopment in each formal sector ring is continuously expanding into the adjoining old formal sector development. These patterns are illustrated in Fig. 2, where we 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. The increase in volume is achieved by increasing land cover in the informal area and by greater height in formal areas closer to the centre. With $D(x)$ =0, the width of the ring of informal area, $x_1 - x_0$ is constant through time. Hence one can show that, even in a circular city, the share of urban land area that is informal falls with time and as the city gets larger.[12]

[12] Generally, the area of land occupied by the informal sector becomes narrower through time if house-rent growth is faster in the formal sector than informal (quality unadjusted) or the formal sector house-rent gradient is flatter than that of the informal sector, $\hat{p}_F / \theta_F > \hat{p}_I / \theta_I$ The general expression is

$$\frac{dx_1}{dt} - \frac{dx_0}{dt} = \frac{\gamma(\alpha - 1)\{\hat{p}_I \theta_F - \hat{p}_F \theta_I\}}{\theta_I \{\theta_F \gamma(\alpha - 1) - \theta_I \alpha(\gamma - 1)\}}.$$

2.3.2.2 Formalisation costs.

At any location in the city, x, the time of first formal sector development is given by (12a). A higher value of $D(x)$ means that higher formal sector house-rents are needed to trigger formalisation, so has the effect of postponing the first formal sector development. When the delayed development occurs, it will occur at a higher rent and hence volume from eqn. (8). But given that starting point, from Proposition 1 we know the time interval between each redevelopment is fixed and the percent volume increases at each redevelopment are the same.

We illustrate two cases. Fig. 3 illustrates a case there is an interval or ring of x within which $D(x)$ is positive, but $D(x) = 0$ elsewhere. As expected, this extends the period during which the area is occupied by informal settlement; during this period volume per unit area and crowding increase through time. A history of informality has a persistent legacy on the area as discussed. Formal development starts later at a higher price and volume, compared to its neighbours where $D(x) = 0$, and so therefore does subsequent redevelopment, as illustrated.

Fig. 4 illustrates a more vivid case in which $D(x)$ is set at random non-negative values by distance from the centre. All locations see volume increase with time, but initial and subsequent formal development take place at different dates and build to different heights. Gradients of volume for existing buildings may no longer be monotonically decreasing from the centre in such a city and there will be heterogeneity by ray from the city centre. Such patterns are the hotchpotch we see in the data. We note one result which will be important in the calibration and empirics. From eqn. (8), any *new* formal sector developments that occur at the same date t have less volume the further they are from the city centre,

$$\frac{d\log(v_F;\tau_i = t)}{dx} = \frac{-\theta_F}{\gamma - 1}, \qquad (17)$$

since they follow house-rents at the given date, which decline with distance from the centre. It is this relationship from eqn. (8) which will help us identify diseconomies of building high.

2.4 Expectations and closing the model

We have assumed, so far, that decisions are based on perfect foresight. What are the consequences of removing this assumption? We do not have an empirical counterpart to this, but it is an important issue. We did an example, here footnoted, to show that if developers underestimate the rate of future price increases, formal development based on pessimistic expectations has less volume and height at any time and location. With less investment, buildings become obsolete more rapidly, and the interval between redevelopments falls. This underbuilding and greater churning has a substantial welfare cost or loss in land values.[13]

To this point our analysis has posited given time paths for house-rent each type at each location. The model is completed in the Theory Appendix by specifying household behaviour, commuting and hence the demand for space at each place and time. This is constructed in a way consistent with the preceding analysis, offering a model of price and population growth, arising as city productivity grows faster than productivity outside. The city is open with free migration from outside, which then also defines welfare costs to be in terms of forgone land-rents.

As detailed in the Theory Appendix, we assume identical consumers, who have log linear preferences and commuting costs such that income net of commuting costs declines exponentially with distance from the centre. Our intention in using a representative consumer model is 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. We would need either a different functional specification to preferences or to add heterogeneous tastes for slum housing quality. In the context of Nairobi, we note that slums draw from all education groups and are not populated overwhelmingly by very low education individuals. 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% in slums versus 21% in the formal sector) and more who terminate education after primary (standard 8, with 23% in slums versus 13% in formal housing).

3. Empirical work on Nairobi

Our empirical work provides descriptive evidence on the built fabric of Nairobi and its evolution through time. We also use it to calibrate the model (section 4) and as the basis for our examination of the costs of land market imperfections (section 5). In this section we first add to the data description given in the Introduction. Then section 3.2 describes the 2015 urban cross-section and draws out key statistics that we use in calibration. Section 3.3 outlines the dynamics of the building volume changes that occurred in the period 2003-2015.

[13] The perfect foresight value-to-rent ratio on a newly constructed property is $\Phi(x,i)$, and for the parameters used in Figure 1, Φ =20.6 and the interval between redevelopments is $\Delta\tau$ = 90.4. What happens if developers have less positive expectations and build on the basis of a value-to-rent ratio of 15.5 (imposed at 75% of the perfect foresight value)? The transition from rural to informal settlement is unaffected by this, but formal development is based on these less optimistic expectations. As a consequence, developers build less volume and hence buildings become obsolete more rapidly and the interval between redevelopments drops to $\Delta\tau$ = 60.

3.1 Data and mapping

As noted in the Introduction and detailed in the Appendix, we develop a data set for Nairobi which captures characteristics of the built environment 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. These data give building footprints for 2003 and 2015. We also have LiDAR height data for 2015, but not for 2003. We infer 2003 heights, assuming that the height of unchanged buildings is the same as in 2015 (ignoring the possibility of adding floors to a structure). For demolished buildings, we assume 2003 height equals 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.

In addition to this data on the built fabric of the city we have price data. Asking prices for vacant land in 2015 are obtained by scraping from property24.co.ke, a website that advertises property for sale in Kenya.[14] Listings are only found for the formal sector. We also have data on rent and housing characteristics. These are derived from a georeferenced household level data set from the 2012 ‘Kenya: State of the Cities’ survey by the National Opinion Research Center (NORC). They record household-rent per square meter, many dwelling unit characteristics, and distinguish between formal sector and slum housing.

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. The IPE mapping used satellite imagery and topographic maps with the general idea that slums are “unplanned settlements” which have some aspects of low house quality, poor infrastructure, or insecure tenure. We refine these areas in two ways. IPE classifies slums with very tight boundaries cutting off vacant land adjacent to the slum (including roads or rivers dividing a slum) and even edge slum housing. The formal sector is a residual of everything else in the city. To offset the tight delineation of slum areas, 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. Second to focus on private sector development, we remove all grid squares entirely in permanent public uses listed in the Appendix amounting to 11% of land in the 2003 city boundary, and 25% at the centre (0-1 km, including parks and the Presidential palace). Note that neighbourhood schools and roads are not removed.

[14] 80% of the listings have information on prices and plot area. Sales are for formal sector land that is being redeveloped.

An alternative delineation of slums comes from a 2003 land use map prepared by the CSUD at Columbia University. As well as being older than the IPE mapping, 2004 satellite imagery (SPOT) indicates that it misses some emerging slums areas. It is clear that the effective definitions differ sufficiently across the two studies that they cannot be used to precisely distinguish slum creation and destruction per se.

Nevertheless, in Fig. 5 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 boundary based on built cover. For a (150m x150m) grid cell to be within 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. 5 shows the city in dashed outline in 2003 and in solid outline in 2015. We focus on the 2003 city, but at points discuss the larger 2015 city margin.

Slum areas as recorded in the two studies are marked green if in both studies, yellow if only in IPE (2011) and blue if only in CSUD (2003). There is an area near the CBD in which there were no recorded slums, marked by the dashed and solid rings for each period. The areas inside these rings expand considerably between dates, from a 0.775 km to a 2.0 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. However, there seems to be little overall slum removal (blue), the issue of concern in Section 5. We 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.

3.2 Characteristics of Nairobi's built environment and its evolution

We now use the data and our analytical structure to draw out the facts about Nairobi and derive the relationships used to calibrate the model. We graph and describe these relationships, and quantify them with regression analysis, looking in particular at how land prices, rents, building heights, cover, and volumes vary according to sector (formal or informal) and with distance from the centre. This quantification is used in calibration to reveal key parameters of the model such as α, γ, θ_F, and $\hat{p}_F$.

Panel A of Table 1 gives the regression information that is used in calibration. The actual numbers used are in bold. All regressions in the table contain controls on elevation and ruggedness; different columns have additional noted controls depending on the relationship being analysed. Panel A records gradients with respect to distance from the centre, and level intercepts projected to activity at the city centre. Panel B describes other relationships of interest and just reports gradients. Coefficients on all covariates in panel A are in Appendix Table 2.4, as are details of controls.

Table 1. Land price, House-rent, Height, CAR and BVAR gradients					
Panel A. Key gradients used in calibrations					
	(1)	(2)	(3)	(4)	(5)
	Ln land sales price ($2015 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	**-0.173** (0.0518)	-0.0618 (0.0304)	-0.00145 (0.0288)	**-0.101** (0.00494)	**-0.0989** (0.0103)
Intercept for typical item	**7.247** (0.301)	**3.140** (0.0704)	**1.892** (0.0612)	**3.313** (0.0339)	**1.318** (0.0681)
Controls, apart from ruggedness & elevation	yes	yes	yes	no	no
Observations	136	361	442	4589	983
R-squared	0.292	0.308	0.409		0.111

Panel B. Other gradients					
	(1)	(2)	(3)	(4)	(6)
	Ln slum height	Ln slum CAR	Ln formal CAR	Ln formal BVAR, all.	Ln formal height, all.
Distance to centre	0.00757 (0.00409)	-0.104 (0.00948)	0.0219 (0.00429)	-0.0510 (0.00534)	-0.0757 (0.00239)
Other controls	no	no	no	no	no
Observations	983	983	5394	5394	5394
R-squared	0.018	0.145	0.055	0.038	0.212

Note: Regressions are based on observations for the 2003 extent of the city. Standard errors in parentheses. The intercept for slum and formal sector house-rents is based on two regressions; for slums rents we set the distance coefficient to 0, as in the model (i.e. there is no distance variable in the regression); and for the formal sector we set it to θ solved for in the model of -0.072. The impact of these forced gradients on estimated house characteristics is minimal, given the within sector differences in gradients are so small.
As given in Table 2.4 of the Appendix, land sales price regressions also control for month of listing, whether a plot is geo-coded, and lot size and its square. Rent per unit volume regressions control for a variety of characteristics giving house layouts and utilities and wall, roof and floor materials.

3.2.1 Vacant land prices and house-rents:

We start with prices of land. Col. 1 of panel A of Table 1 gives the estimated gradient with respect to distance from the centre and the intercept of ln asking price of land (per square meter) in USA$, using the 2015 property24.co.ke data for the 2003 area of the city. An exponential form (as used in eqns. 8-10 and 11b) captures the relationship well, and the gradient is steep, with price declining by 17.3% 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 gradient will be used to infer how all rents and prices vary with distance to the centre. The intercept of the regression will be used to measure the level of land prices, giving the predicted value at the city centre of a *typical* listing based on median area, ruggedness and elevation, and the like.

Information on house-rents comes from the NORC data. Table 1 panel A cols. 2 and 3 report gradients and intercepts for formal and slum rents, in 2015 USA$ for cubic meters of volume.[15] Our main interest in cols. 2 and 3 is in the intercepts which give projected rents at the city centre.[16] These columns control for house characteristics in order to define rents per unit volume for *typical* formal and slum sector houses. However, for welfare analysis in Section 5, we want projected slum rents to capture quality differences due to differential crowding and other social aspects over space, so there are no controls on neighbourhood attributes in the regressions. Of course, the omitted crowding characteristics may be correlated with observed characteristics, potentially biasing all coefficients. An alternative would be to control for no house characteristics and base results on raw rent gradients. In Section 5 we will show that how that affects welfare analysis. In the current estimation, we note the tiny estimated gradient of -0.0015 (col. 3 of panel A) for *observed* slum rent inclusive of crowding quality differences, $p_I(x,t)a(v_I(x,t))$, is flat assumed in the model.

3.2.2 The built environment and building volume

We now turn to the total volume of building space provided across the city and the composition of this volume. Our data is on building cover and height, and these combine to give volume. The intensity of land-use at each *x* is captured in built volume per unit area (BVAR), defined in eqn. (18)

$$\begin{aligned} BVAR(x) &= \rho_I(x)BVAR_I(x) + (1-\rho_I(x))BVAR_F(x) \\ &= \rho_I(x)h_I(x)CAR_I(x) + (1-\rho_I(x))h_F(x)CAR_F(x) \end{aligned} \tag{18}$$

At each distance from the city centre, *x*, the share of area in slums is denoted $\rho_I(x)$ and the share in formal as 1 - $\rho_I(x)$**,** where each cell (3x3m) is classified as either informal/slum (*I*) or formal (*F*) on the basis of the adjusted IPE map. The cover to area ratio is now denoted *CAR*.

[15] 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$.

[16] We do not rely on these gradients to infer how house rents vary with distance from the centre (θ), but rather solve that in Section 4 from model relationships driven by the land price gradient in col. 1, which we consider more reliable. The estimated gradient in col. 2 for formal sector rents of -0.062 is close to the number (-0.072) we derive in Section 4 based on land prices, but the estimated gradient in col. 2 is sensitive to the exact hedonic specification. For example if we don't control for house characteristics the gradient is -0.103.

We use this to compute BVAR. Smoothed values for each sector are plotted in Fig. 6 for all buildings, redeveloped or not.[17] In the formal sector up to almost 2 km from the centre, BVAR is very high, averaging around 7 cubic metres of space per meter of ground area. At 2km and beyond, slums and the formal sector deliver essentially the same BVAR. This is consistent with the theory where we saw that cover in slums and height in formal areas could deliver similar housing volumes. 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. For the opposing views of whether formal sector height trumps slum coverage in providing volume of built space, this tells us that, in Nairobi, they do equally well on average, albeit at very different quality levels.

Total volume at each distance from the centre is based on BVAR and the amount of land overall and in each sector at that point. Smoothed total volume is illustrated on Fig. 7, peaking at almost 13.5 million cubic meters for points around 4 km from the centre, as the amount of potentially available land in any circumference increases up to about 4kms. Total volume then falls to average around 7-8 million, the decline partly reflecting the fact that Nairobi has little available land beyond 4km to the direct north and south. In terms of composition, there are no slums within 2km of the centre, and from about 4km outwards the slum share of volume nears 20%. Across the city as a whole slums amount to just 10% of total building volume, and at no distance up to 10km from the centre do slums occupy more than 20% of non-public land.

We now turn to the components of BVAR, the base (CAR) times the height (*h*). We do this by sector, since their treatment differs given differential durability of capital and the role of height and cover-to-area ratios.

3.2.3 Formal sector: height and cover-to-area ratio

Height information is illustrated in Fig. 8 which gives the mean and the 5th and 95th percentiles for the height of formal sector buildings regardless of age or status, as well as that for slums, at a series of distances from the centre. The formal sector exhibits large variability; especially near the centre there are office towers, historical buildings, parking lots and housing. Variability also comes from the differential timing of redevelopment, based on the specific history of the property's path to formalization (as in Fig. 4). Overall, it is clear that Nairobi has height; buildings from 0-1 km of the centre average (at the 3m x 3m pixel) 10 stories (at about 3m, a storey) and in Fig. 8, 5% of these pixels are over 16 stories. This

[17] 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, with default settings. A graph on heterogeneity of BVAR by sector looks much like the one for height with lots of heterogeneity in the formal sector and much less heterogeneity in slums.

despite the view sometimes expressed that height in African cities is constrained, e.g. by unreliability of power for elevators.

Fig. 9 gives the smoothed average height of redeveloped and unchanged buildings in the formal and informal sectors. In the formal sector, from 1.5 to 5 km, redeveloped buildings average twice the height of unchanged buildings. This is building taller with redevelopment, consistent with the model and will drive the large volume changes between 2003 and 2015 that we analyse below.

Regression analysis provides estimates of the gradient of heights with respect to distance from the centre. For all formal sector buildings (redeveloped or not), this gradient is -0.0757 (Table 1.B col. 6), indicating a 7.6% decline with each km distance. For calibration of the formal sector we need to know the gradient for heights at date of development (eqn. 17), so use the height of redeveloped buildings (Table 1. A, col. 4), which will be steeper. An issue is that redeveloped buildings were not built instantaneously in 2015 but over a 12 year period, where we expect those built in 2015 at higher land prices to be generally considerably taller than those built in 2003. To account for this, we use quantile regressions for higher percentiles. This mainly affects intercepts. Gradient coefficients on redeveloped buildings are pretty stable, starting with -0.0956 for OLS and rising to a high of -0.11, for various cuts between the 70th to 92nd percentiles. We settled on the 80th percentile, balancing out wanting more recent hence taller buildings, against getting extreme outliers at any time.[18] For the 80th quantile the gradient coefficient in Table 1.A col. 4 is -0.101, so redeveloped formal building heights decline by 10.1% per km from the centre.

The second dimension of the built fabric is the cover to area ratio, CAR. Note that CAR is not the usual floor-to-area ratio; the area involved is all land (except for large public uses), including roads, sidewalks, public schools, small parks and the like. This in part accounts for the huge variability to CAR for 2015 as plotted in Fig. 10. In the formal sector, average CAR is typically 25% throughout the city. The gradient is almost flat (as assumed in the model), rising by 2% per km (col. 3 in panel B of Table 1). In calibration, the gradient on redeveloped height as in the model is used to give the BVAR gradient for new buildings, relying on the approximately flat CAR in the formal sector. This choice is also forced because we cannot accurately assign land areas around building footprints into parts attributable to neighbouring redeveloped versus existing buildings to differentiate BVAR of old versus redeveloped buildings. [19]

3.2.4 Slums: Cover-to-area ratio and height

---

[18] We also wanted growth in building heights from calibrated parameters to be in line with growth predicted from various intercepts for different percentiles.

[19] For the record we note the gradient for formal stock BVAR in Panel B, col. 4 (-0.051) is different than the height gradient (-0.101) for redeveloped buildings just as the gradient for stock height was different.

Slums in some sense are easier to analyse. Because of the presumption of malleability we do not need to distinguish older versus redeveloped buildings. In Fig. 8, not only are slum heights low but generally they have little variability especially in older slums nearer the city centre. The estimated height gradient for slums is flat (Table 1.B col. 1). Consistent with this, in Fig. 9, heights of unchanged and redeveloped slum buildings are almost the same. In short, slum BVAR varies because of variation in CAR. This is consistent with assumptions in the model.

For slums, CAR near the city centre is very high at 50-60%, recalling the 25% number for the formal sector. The high CAR means that slums have little green/open space around houses, with attendant loss of amenity.[20] Slums have a distinct CAR gradient, where in Panel B col. 2 we see that slum CAR declines by 10.4% per km from the centre. Given fixed height, CAR changes define BVAR changes. For calibration of slums, we estimate the BVAR gradient and intercept directly, as reported in Table 1, Panel A, col.5. The slope of the slum BVAR gradient, -0.0989, is and should be virtually that same as the CAR slope (-0.104). We use the BVAR rather than CAR gradient in calibration simply because it directly gives a projected BVAR intercept at the city centre, as reported in the table.

3.3 Development dynamics and volume change.

How does built volume change between 2003 and 2015? As noted in the Introduction, total volume in non-public use within the 2015 boundary increases by 59%, similar to population growth. That growth is split between the 48% growth within the intensive margin of the 2003 boundary in Fig. 5 and the 120% growth at the extensive margin between the 2003 and 2015 boundaries. Slum growth is modestly smaller (55%) than the formal (60%) with slums accounting for only 9.2% of the increase in total volume. A declining relative role of slums is consistent with the model. The slum share of population at about 29% in the 2009 Census is just under the 1999 Census number.

Volume change is broken out by sector and distance in Fig. 11. There are 45-70% increases in total volume from 2-8 km, highest around 4-5 km. There is less change in the first km from the centre which is locked in by historical buildings and roads, and by sky-scrapers already built over the last 35 years (see Section 5 on predictions of redevelopment dates). Within 9 km, percent increases in the formal sector generally exceed those in the slums, showing the increasing relative role of the formal sector in the main part of the city. However, slum changes dominate at the city edge, as the model predicts. Note since slums have such a small weight in total area, total changes generally mimic those in the formal sector.

[20] 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%.

Our detailed analysis of the aerial photographic images enables us to distinguish how change occurs: redevelopment, in-fill, or greenfield development. For redevelopment, we note that from 1 to 4km about 35% of buildings are torn down and 50% of these are redeveloped. 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, reflecting its rapid redevelopment fuelled by growth. Redeveloped buildings have a significant increase in average footprint size of 100% at 3 km, rising to 200% by 6km, because of their increase in height. In reality, scale economies and construction efficiency require more footprint as height increases to allow for elevators, staircases, and reinforced construction.

Fig. 12 decomposes the *formal* sector percentage volume changes into the relevant elements. To highlight changes in the core of the city, we cut the *x*-axis at 8km (and give numbers at 10 km in the fn. to the figure). We see that the increase in building volume nearer the city centre is dominated by redevelopment. The net increase in volume just at 3km out is over 35% due to redevelopment, with about another 20% from infill.[21] Further from the centre where there is less initial development, infill starts to escalate. There is also a residual: demolition (without yet redevelopment) involves small coverage and volume changes throughout. There are two notes on this. First between the net of demolition, in-fill, and larger footprints for redeveloped buildings, CAR in the formal sector does increase over time, a factor not captured in the model but one which makes sense given increasing opportunity cost of land. Second for the slum sector all increases in volume, given fixed low level height, are driven by CAR increases that exactly mimic the volume increases in Fig. 11.

4. Calibration

We now pull together the theory and empirics, calibrating the model to Nairobi. For ease of exposition we divide the calibration process into three parts.

Gradients and elasticities: The theory implies that land-rents and building volumes decline with distance from the centre; and Table 1, Panel A presents relevant estimates from which we can calculate key parameters of the model. In the informal sector the spatial gradient of volume is, from eqn. (3) with (6), $\frac{dv_I}{dx}\frac{1}{v_I} = \frac{-\theta_I \alpha}{(\alpha - 1)}$ (which also equals $\frac{dr_I}{dx}\frac{1}{r_I}$), and the

[21] The components do not add up exactly to the total because some buildings classified as unchanged exhibit small changes in drawing footprints between the years.

numerical value of this comes from the slum BVAR gradient from Table 1.A col.5. In the formal sector building volume is fixed within each stage of development and discretely lower at development stages further from the centre. The volume gradient between points of redevelopment in the hotchpotch cross section is given by $\frac{dv_F}{dx}\frac{1}{v_F} = \frac{-\theta_F}{(\gamma-1)}$ from eqn. (17). In this sector CAR is fixed, so the gradient is given empirically by that for heights of newly redeveloped formal buildings from the quantile regression in Table 1.A col. 4. Finally the slope of formal sector land-rents, $\frac{dr_F}{dx}\frac{1}{r_F} = \frac{-\theta_F\gamma}{(\gamma-1)}$, from (9b) with (16), is that of the land price gradient from eqn. (10b), which slope we have in Table 1.A column 1.

This information is summarised in Table 2. With the further assumption that the spatial gradient of quality adjusted house-rents is the same in both sectors, $\theta_I = \theta_F = \theta$. We calibrate the model parameters $\{\alpha, \gamma, \theta\}$ from this information. The calibrated values are given in the bottom row of the table.

The calibrated value of $\gamma = 1.71$ implies that the share of land-rent in formal sector revenue is 42%, while $\alpha = 3.68$ implies a corresponding share in the informal sector of 73% (eqns. (9a) and (4) respectively). The formal sector share, while high, is in line with recent developed country data (Ahlfeldt and McMillen 2014, Duranton and Puga 2015). 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. Recall, the model predicts that the formal sector is closer to the city centre if $\alpha > \gamma$, which is the case.

Table 2. Gradients and parameters

Gradients:	Informal:	Formal:
Volume (m^3 per land m^2)	$\frac{dv_I}{dx}\frac{1}{v_I} = \frac{-\theta_I\alpha}{\alpha-1} = -0.0989$	$\frac{dv_F}{dx}\frac{1}{v_F} = \frac{-\theta_F}{(\gamma-1)} = -0.101$
Land-rent (per land m^2)	$\frac{dr_I}{dx}\frac{1}{r_I} = \frac{-\theta_I\alpha}{\alpha-1}$	$\frac{dPV(\tau \geq \tau_1)}{dx}\frac{1}{PV} = \frac{dr_F}{dx}\frac{1}{r_F} = \frac{-\theta_F\gamma}{\gamma-1} = -0.173$
Calibrate: $\alpha = 3.677$, $\gamma = 1.713$, $\theta = 0.072$		

Growth, present values and time periods: Calibration of the dynamics requires that we know the discount rate and the rate of increase in the price of housing. We set the real interest rate at $\rho = 0.057$, which is the average of the World Bank real interest rate numbers for Kenya for the 14 years post 2002.[22] To get the rate of increase in housing rents, we relate formal sector land sales prices to current land-rents to back-out the implied rates of increase in housing- and land-rents. From eqn. (10a) we know the present value of land at location *x* redeveloped at time τ_i is given by $PV_F(x,\tau_i) = r_F(x,\tau_i,\tau_i)\Phi / \left[1 - e^{-(\rho - \hat{p}_F \gamma/(\gamma-1))\Delta\tau}\right]$. Noting that $\Delta\tau$ and Φ are time and spatially invariant, so too is $PV_F(x,\tau_i)/r_F(x,\tau_i,\tau_i)$. The relevant numbers are based on intercepts reported in cols. 1, 2 and 4 in Table 1A. [23] Other variables in this equation are Φ, $\hat{p}_F$ and $\Delta\tau$. Two more equations are given by the value-to-rent ratio $\Phi = \left[1 - e^{(\hat{p}_F - \rho)\Delta\tau}\right] / (\rho - \hat{p}_F)$ and building life $\Delta\tau = ((\gamma-1)/\hat{p}_F)\ln(\gamma/(\gamma - \rho\Phi))$, from eqns. (14). We solve the 3 equations in 3 unknowns and derive

$$\hat{p}_I = \hat{p}_F = 0.0091,\ \Phi = 20.62,\ \Delta\tau = 90.42 \text{ years.} \qquad (19)$$

Real house-rents are appreciating at about 0.9% annually, the value-to-rent ratio is 20.6 and the length of life of a building is about 90 years.

Levels: Remaining parameters are the levels of house-rents, construction costs in each sector and the informal sector amenity adjustment, $p_I(x,t) = p_F(x,t)$, κ_F, κ_I and a_I. Table 3 gives the relevant equations of the model for each sector, and intercept terms in Table 1.A give values at *x* = 0 and *t* = 2015. Informal sector house-rents are from col. 3 of Table 1.A, which given *α* above yields κ_I. Formal sector house-rents come from col. 2 of the table, and informal and formal sector volumes respectively from cols. 4 and 5. The parameters implied by these equations are given in the bottom row of the table.

Table 3. Levels and parameters: all evaluated at *x* = 0, *t* = τ_i = 2015, in $US		
	Informal: Eqns 1-4	Formal: Eqns 8, 16
House-rent per BVAR	$p_I(x,t)a_I v_I(x,t)^{(1-\alpha)/\alpha} = \alpha\kappa_I = 6.63$	$p_F(x,t) = 23.10$
BVAR: Volume (per unit land area)	$v_I(x,t) = \left[p_I(x,t)a_I/\kappa_I\alpha\right]^{\frac{\alpha}{\alpha-1}} = 3.74$	$v_F(x,\tau_i) = \left[\dfrac{p_F(x,\tau_i)\Phi}{\kappa_F\gamma}\right]^{\frac{1}{\gamma-1}} = 6.78$
Calibrate:	$p_I(x,t) = p_F(x,t) = 23.10;\ a_I = 0.75,\ \kappa_I = 1.80,\ \kappa_F = 71.02$	

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

[23] The predicted values are $PV_F(0,2015) = \$1403$ and $r_F(0,2015) = \$24.80$. The former comes directly from Table 1.A.1. The latter uses eqn. (9), $r_F(x,t,\tau_i) = [1 - 1/\gamma]\, p_F(x,t) v_F(x,\tau_i)$, with house-rent data from Table 1.A.2 with $p_F(0,2015) = \$23.1$ given ln $p_F(0,2015) = 3.14$, and BVAR from Table 1.A.4 giving $v_F(0,2015) = 6.784$ obtained by multiplying the intercept for redeveloped height (80th percentile) by the median CAR of 0.247 for grid squares in the city.

We note that for later welfare analysis we will want formal and slum land-rents throughout the city in 2015, not just at the centre. For slums, from eqs. (4) and (16)

$$r_I(x,t) = Ae^{-x\theta\alpha/(\alpha-1)}, \quad A = [1-1/\alpha]p_I(0,t)a(v_I(0,t))v_I(0,t). \qquad (4a)$$

For the formal sector from (9a) and (16)

$$r_F(x,t,\tau_i) = Be^{-x\theta\gamma/(\gamma-1)}, \quad B = [1-1/\gamma]p_F(0,t)v_F(0,t=\tau_i) \qquad (9c)$$

where $p_F(0,t)$ is the intercept of the formal house-rent gradient and $v_F(0,t=\tau_i)$ is the implied BVAR intercept (see fn. 23).

5. Slum redevelopment, lack thereof, and welfare costs

Slum redevelopment into formal sector usage appears very limited in the 12 years of our data on Nairobi. As noted earlier we cannot precisely map slum area formalisation over time, but Fig. 5 indicates little change. Moreover when we look at changes over time within areas defined as slums in 2011, while there is considerable redevelopment, the heights of unchanged and redeveloped slum buildings are similar. The vast majority (70%) of both are 3m high. Further, there is little change at the upper end of the distributions. For example, 2.9% of unchanged slum buildings are over 9m high and only 3.4% of redeveloped slum buildings exceed 9m. These small changes are also clear from visual inspection of the high resolution SPOT satellite images we possess for 2004 and 2015. In Section 5.1, we discuss the reasons for persistence and in section 5.2 we use the calibrated model to calculate the loss in land values associated with inefficient land-use.

5.1 Formalisation costs

Earlier, we argued that persistence of slums nearer the city centre is due to formalisation costs. While most land is privately held in Nairobi, remaining lands that are not privatised, especially central government lands, present thorny problems. These lands are not managed by the government, but by slumlords who operate quasi-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 73%. Moreover, slumlords have particular features which are problematic to formalisation costs. In Kibera, 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 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 don't own the land and their presence is essentially illegal. They would simply lose profitable businesses. Having well connected bureaucrats and political figures opposed to formalisation presents a political problem.

This problem for 'government owned' slums is magnified when there are historical private claims to the land. Kibera gives a nice example. The 1000 acres in Kibera was awarded to Nubian soldiers for service in 1912 by the British colonial government. 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 not occupied by Nubians and their descendants was settled by others and had titles illegally allocated by local chiefs and bureaucrats. The moral claim of the Nubian descendants to at least the land they occupy is well recognized in the society but the unwillingness to grant them title is yet another road block to redevelopment (Etherton 1971; Joireman and Vanderpoel, 2011).[24] The literature has related stories for other major slums in Nairobi.[25]

5.2 The loss of land value due to slum persistence.

In our open city model welfare costs of inefficient land use are given by land-rents foregone. To assess the cost of delayed formalisation we now use the calibrated model to calculate the present value of land-rents earned by land in different uses, looking in particular at rents foregone due to persistence of slums near the city centre. This measures the real income loss due to inefficient land-use, although it does not capture further 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.

Our calculation is based on the present value of land-rents under alternative uses, informal and formal. For a unit of land at place x we calculate the present value at date s of conversion from informality to formality at date z. We denote this present value $PV(x, s, z)$ and, from eqn. (10) with D's set at zero, it is given by

---

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

[25] 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.

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

$$\begin{aligned} &= r_I(x,s)\frac{1-e^{-(\rho-\hat{p}_I\alpha/(\alpha-1))(z-s)}}{\rho-\hat{p}_I\alpha/(\alpha-1)} + r_F(x,s,s)\Phi\frac{e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))(z-s)}}{1-e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))\Delta\tau}} \\ &= Ae^{-\theta\frac{\alpha}{\alpha-1}x}\frac{1-e^{-(\rho-\hat{p}_I\alpha/(\alpha-1))(z-s)}}{\rho-\hat{p}_I\alpha/(\alpha-1)} + Be^{-\theta\frac{\gamma}{\gamma-1}x}\Phi\frac{e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))(z-s)}}{1-e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))\Delta\tau}} \end{aligned} \quad . \qquad (20)$$

The second term in the second line comes from eqn. (10a) 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. The third line substitutes in (4a) and (9c). With (20) we can solve for land values at different dates of conversion to formal sector use. We also note that, for $D=0$, eqn. (20) implies an efficient date $z = z^*$ for each x, when land should have been developed from slum to first formal use, in the absence of formalisation costs. That comes from the FOC of (21) with respect to z.[26] One can also differentiate that FOC to show that z^* rises as x rises.

Table 4 gives relevant numbers for s = 2015 for rings 0-1 up to 5-6km which cover the core area of the city on which we focus. The first row gives $PV(x,s,z)$, where slum lands are converted to formal sector at date z = 2015, and the second gives $PV(x,s,\infty)$, when land is held in slum use in perpetuity. In all cases, there are large gains to redeveloping slum lands over holding them in slum use indefinitely, a situation we will focus on below. In the third row, we show 2015 values for land which conversion is delayed by 90 years to 2105. Delay is very costly compared to conversion today. These present value numbers as we will see are robust to specific parameter assumptions we have made, as they are based on formal and slum house rent differentials in the data and the derived rate of house rent growth looking forward from 2015.

Rows 6 and 7 give the amount of slum land and households at stake. There are no slums for 0-1km and virtually no lands nor any assigned people for 1-2km, noting assignment for people is based on a separate 2009 Census slum mapping which is generally highly correlated with our 2011 IPE mapping.[27]

---

[26] $Ae^{-\theta\frac{\alpha}{\alpha-1}x}\frac{e^{-(\rho-\hat{p}_I\alpha/(\alpha-1))(z-s)}}{\rho-\hat{p}_I\alpha/(\alpha-1)}\left(\rho-\hat{p}_I\alpha/(\alpha-1)\right) - Be^{-\theta\frac{\gamma}{\gamma-1}x}\Phi\frac{e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))(z-s)}}{1-e^{-(\rho-\hat{p}_F\gamma/(\gamma-1))\Delta\tau}}\left(\rho-\hat{p}_F\gamma/(\gamma-1)\right) = 0$

[27] We don't use the Census mapping because it is earlier and also tends to map large tracts of undeveloped land as slums.

Table 4. The cost of delayed formalisation Present values at s = 2015 in \$2015 per m^2.						
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)	1288	1083	911	766	645	542
PV (*x*, 2015, $z = \infty$)	386	350	317	287	260	236
PV (*x*, 2015, *z* = 2105)	434	390	350	314	283	254
Efficient formalisation year, *z**	1912	1920	1928	1936	1944	1952
Efficient 1st redevelopment year (*z** + 90)	2002	2010	2018	2026	2034	2042
Sq. m slum land, 2011	0	2718	263430	1129311	2263428	1946034
No. slum households, 2009	0	0	2920	29070	45810	33100
PV at efficient *z** in \$2015	10393	6735	4391	2886	1919	1294
D (lower bound)	490	393	313	247	194	150

The rest of the table is more speculative and also sensitive to specific assumptions, since it must make heroic assumptions such as perfect foresight and a rate of house rent appreciation and construction technology that are unchanged since Nairobi was founded in 1899. Given that, it is interesting to see how well the model does. In row 4, we list the efficient *z** year at each distance, when properties would have moved from informal to formal usage absent formalisation costs; and in the following row the year of first redevelopment. For 0-1 km the first formal sector developments would have been in the early 20th century when the city started and at 4-5 km out after World War II. First redevelopment at 0-1 km would have started around 2002. For lands that were formalised on time, we see the model predicts first redevelopment on the area 2-4 km from the centre in the 2010-2026 time period. Above we saw there was massive volume and height changes from redevelopment in this area during 2003-2015. We suspect our calibrated building length of life is high at 90 years especially given that the city centre had massive redevelopment 30 years ago, 20 years before 2002. We also note that if we look beyond 6km, calculations would imply that slums should only persist beyond 15-20 km of the city centre where current development is much sparser.

The second to last row gives the value of land formalised at its efficient date and compounded to 2015. Note the huge numbers and the implied opportunity costs of lost rents if lands had been held in slums at the centre, rather than doing formal sector development 100 years ago. The last row gives a lower bound estimate of *D* of formalisation costs per sq. meter, from eqn. (12) 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.

We use the numbers in the table to give estimates of the gains today of converting slum usage near the city centre. We give two examples from the table which include much of the Kibera area. 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 \$766-\$287=\$479 per m^2. There are 1.13mn sq. meters of slum land in that distance band, of which we estimate an additional 10% would in roads and public schools, so about 1mn sq. meters are available for redevelopment. There is thus an aggregate gain from converting at 2015 compared to perpetual delay of about \$479mn. For a perspective, suppose slumlords were compensated for conversion by \$287 per m^2, as if they had the right to hang on forever. That leaves the remaining surplus of \$479 per m^2. For the 29,000 households affected, the gain is about \$16,500 per household. This is a very large sum, for households paying on average under \$700 a year in house-rents. At 4-5 km, the same type of calculation gives a surplus of about \$17,000 per household, noting the drop in population density between rings 3-4 and 4-5km.

How sensitive are these results to assumptions made? Changing some assumptions makes little difference to the welfare calculations. For example raising the discount rate from 0.0575 to 0.07 lowers $\Delta\tau$ from 90 to 78 and raises and z^* for 0-1 km from 1912 to 1933. However, the gap in present values at different x's changes by less than 1%. Using the 75% quantile rather than the 80th for formal sector redeveloped height lowers $\Delta\tau$ to 84 and raises z^* to 1929 but it only lowers the surplus by 3.5%. As shown in Appendix Tables A2.5 and 2.6, lowering slum rents at the centre by one standard error lowers the date at 0-1 km of first development to 1900 and first redevelopment to 1990. Not surprisingly it does raise the surplus and welfare gains to converting slum lands today at 3-4 or 5-6 km by about 7%. However, lowering both slum and formal sector rents by one standard error each, while lowering $\Delta\tau$ from 90 to 68 and raising z^* at 0-1 km from 1912 to 1963, has a minimal impact on surpluses and welfare lowering them by 2%.

Changing the whole hedonic rent estimation does matter. Suppose rather than keying off of a typical house to infer differential slum-formal sector rents at the city centre, we go to a framework with no house characteristics so rents reflect all endogenous aspects of house and neighbourhood characteristics.[28] In Appendix Table 2.7 Table, that has a modest impact on timings ($\Delta\tau = 91.7$ and z* at 0-1 km is 1921), but it reduces the welfare gain at 3-4 and 5-6 km by about 27%. That does still leave a \$12, 250 surplus per household.

Apart from detailed choice of assumptions, what are the biases in our estimates? First are sources of 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

[28] Now for intercepts, house rent regressions reflect only median elevation and ruggedness with, in regressions, the slum gradient set to 0 and the formal sector gradient to the Table 2 derived slope of -0.072.

higher values nearer the city centre. On the other side, we have ignored moving costs for slum residents. Some would move anyway. But for those forced to move there would be losses 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. Other Considerations

The model only examines residential housing capital and, in the data, we lump all built space together. We do not know building usage per se and in slums many buildings may have a dual residential-production purpose. We do have general land use maps. Much land at the centre is classified as commercial use. Industry is in the eastern half of the city, with older industrial areas starting to the immediate south-east of the city centre and then stretching out. Other large industrial areas are to the north-west away from the centre. All these are far from Kibera in the western part of the city. We also can calculate the ratio of volume of built space to population in the formal and slum sectors. In the centre of the city, volume to population is very high given the intensity of commercial use. Volume to population in the formal sector falls with distance out to 2km; it then is similar to that in slums (which is flat throughout) until 4km, before rising again. Between 4 and 8 km, volume to population is high perhaps because of industrial uses.

6. 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, by steps in the formal sector as redevelopment involves 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 calibrated 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.

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.

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

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

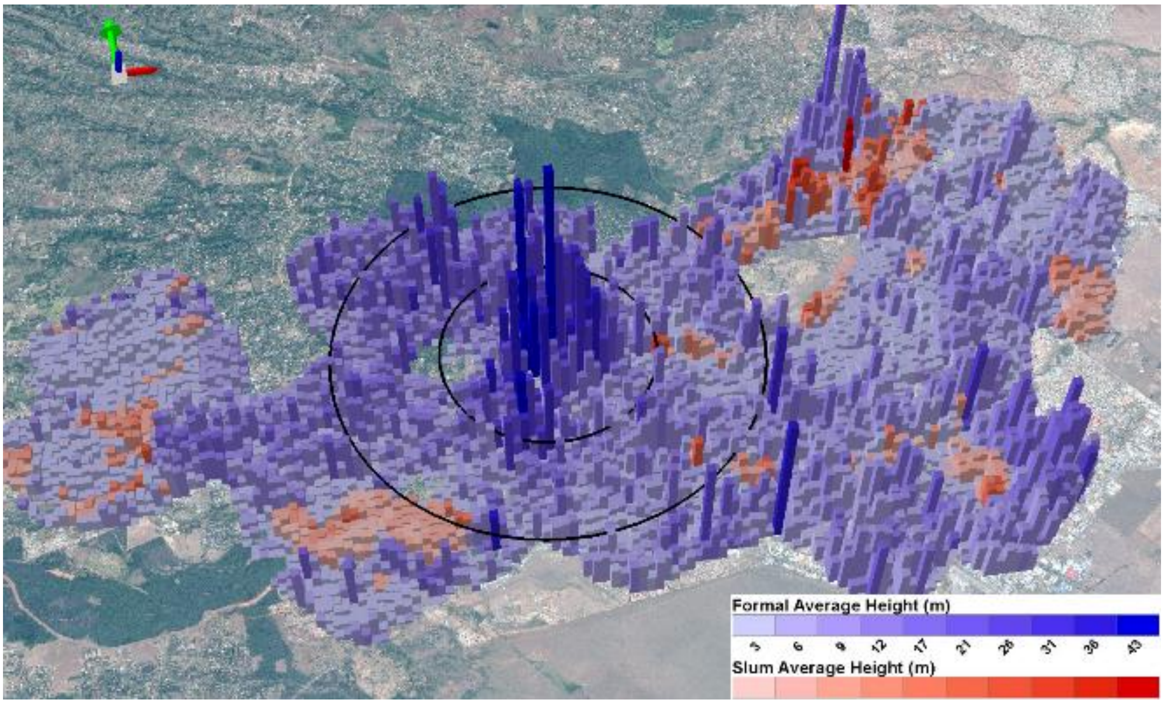


Figure 2: Urban development with perfect foresight

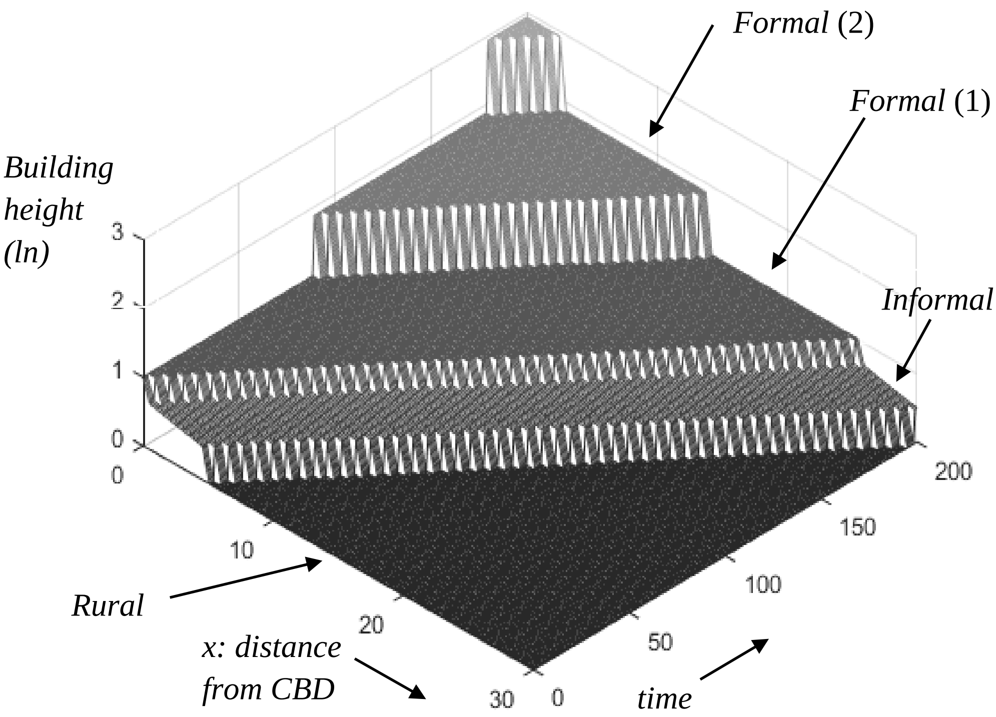


Figure 3: Formalisation costs

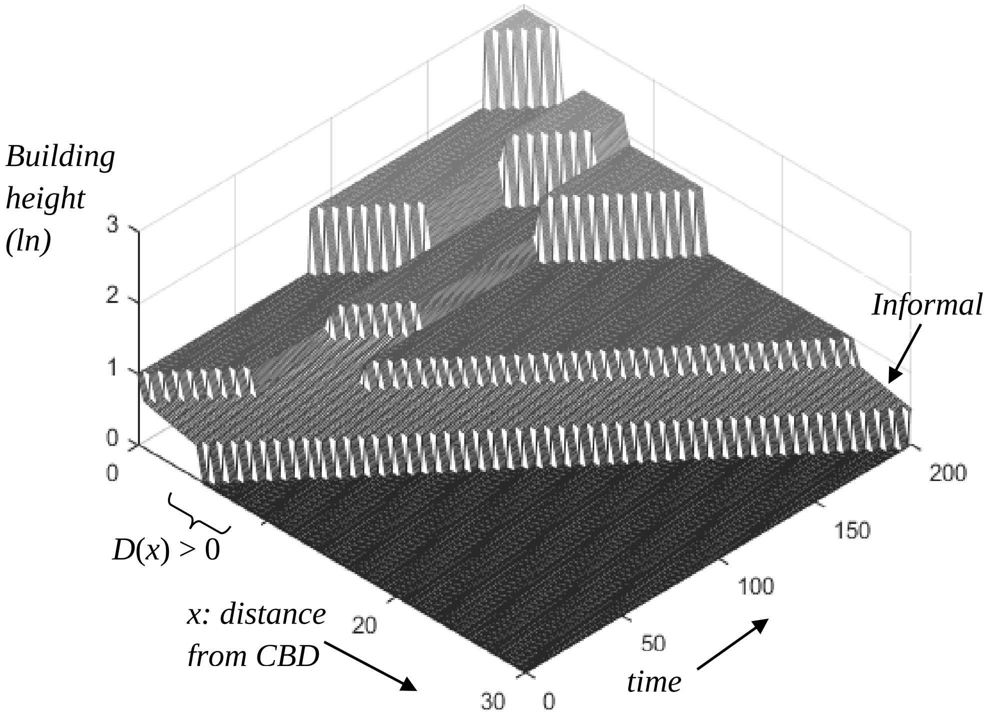


Figure 4: The hotchpotch: random variation in formalisation costs

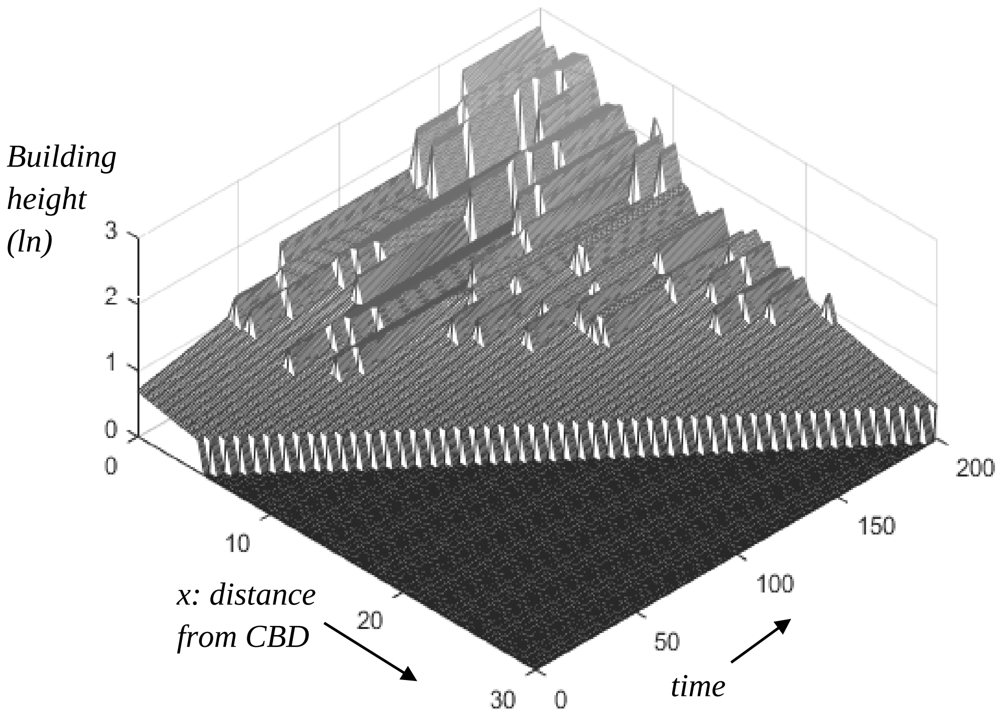


Figure 5. City shape and slums

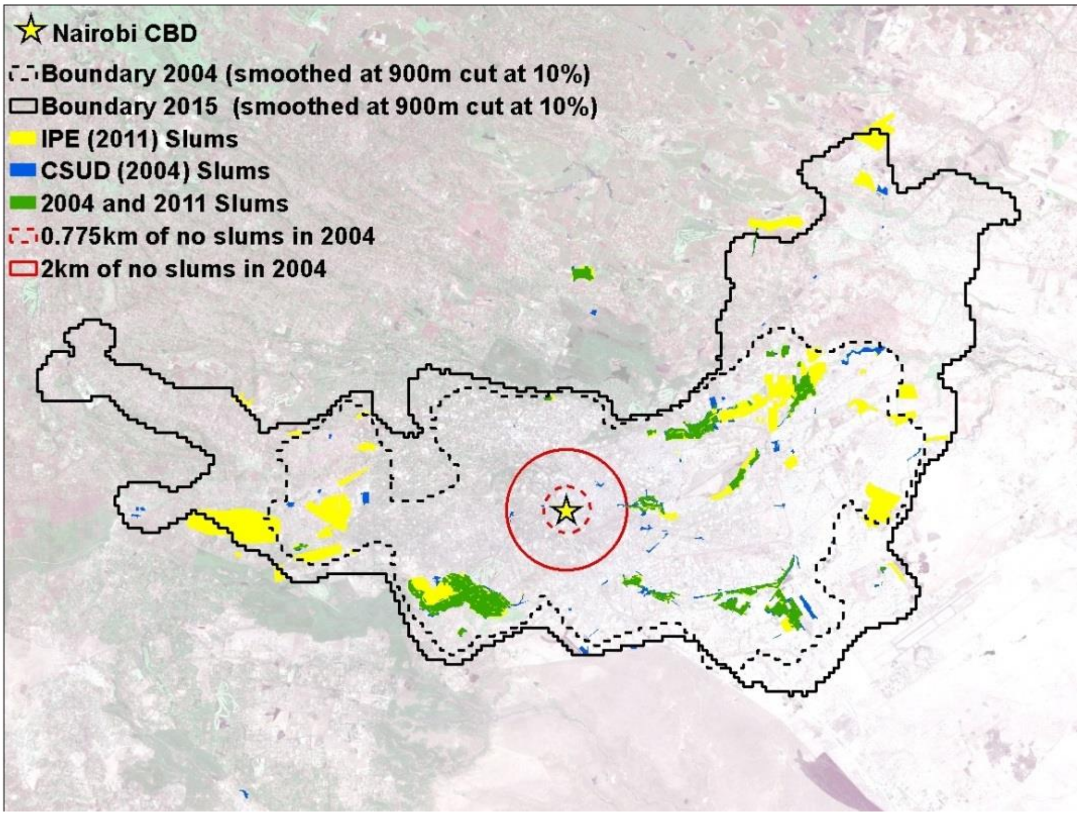


Figure 6. Built volume per unit area (BVAR)

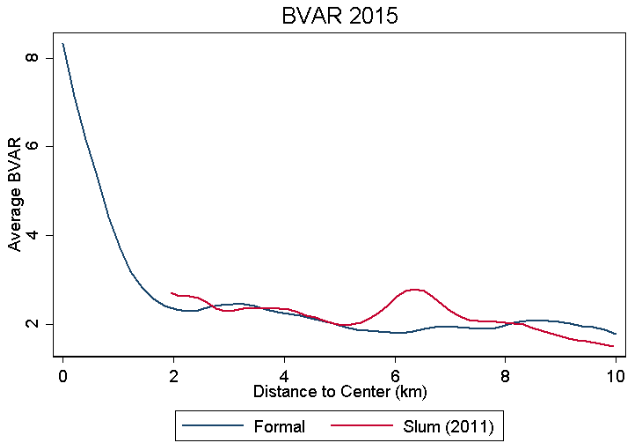


Figure 7. Total volume by distance and sector

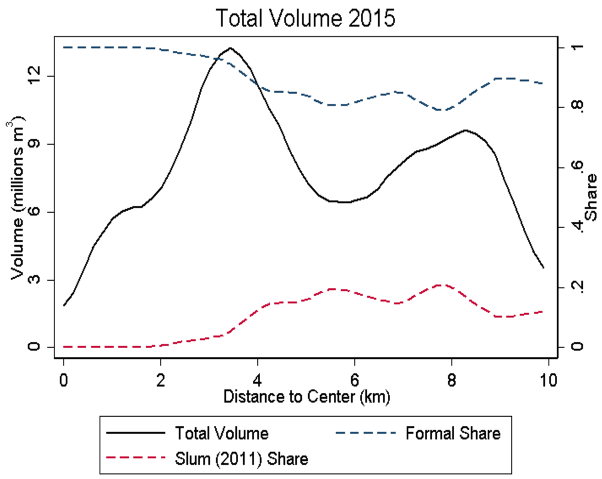


Figure 8. Building height and variability

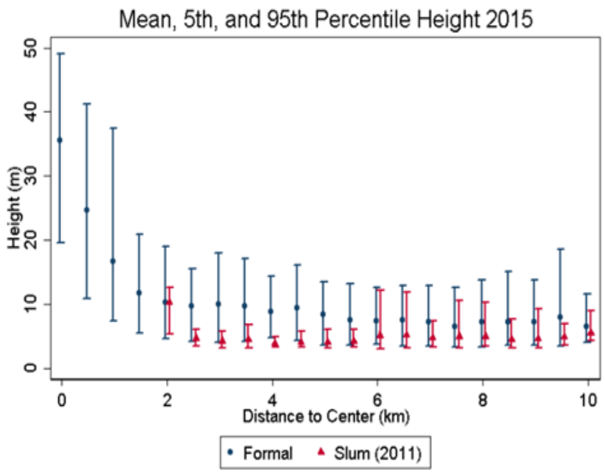


Figure 9. Building height by development & sector

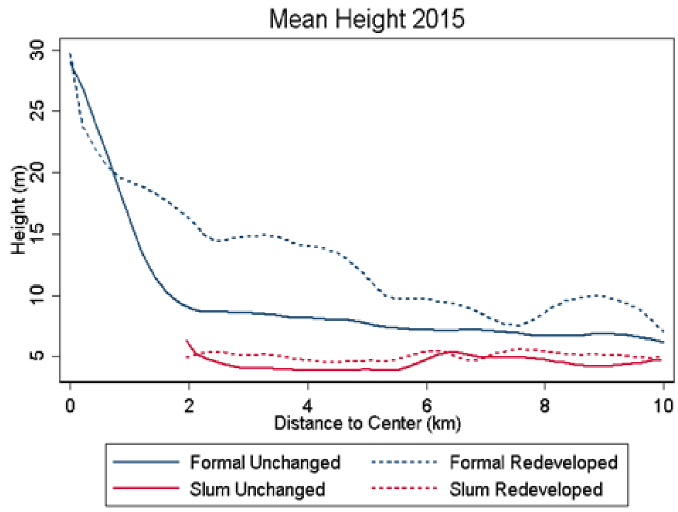


Figure 10. Cover to area ratio (CAR)

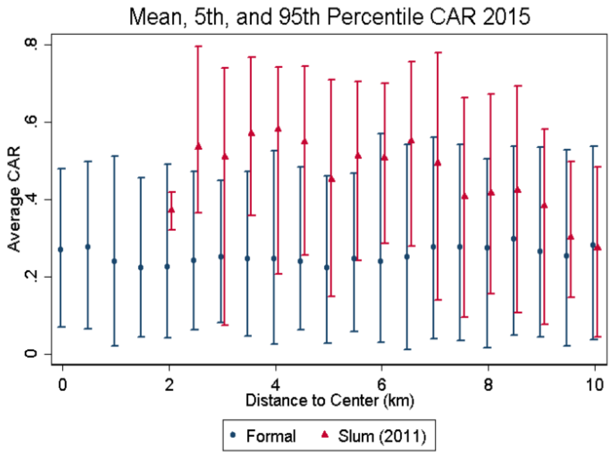


Figure 11. Growth in total volume and by sector

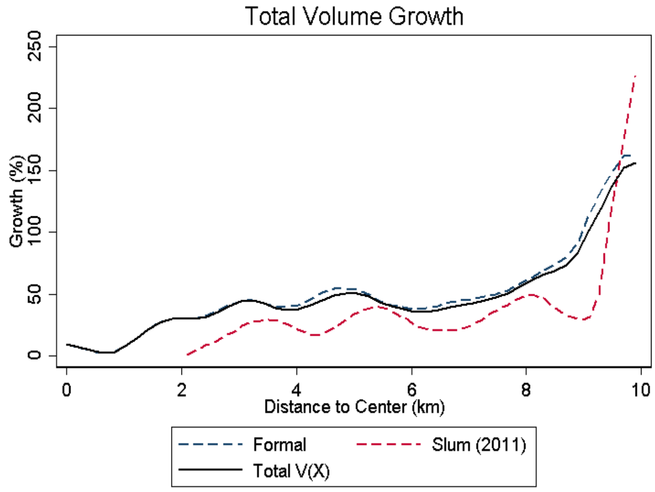
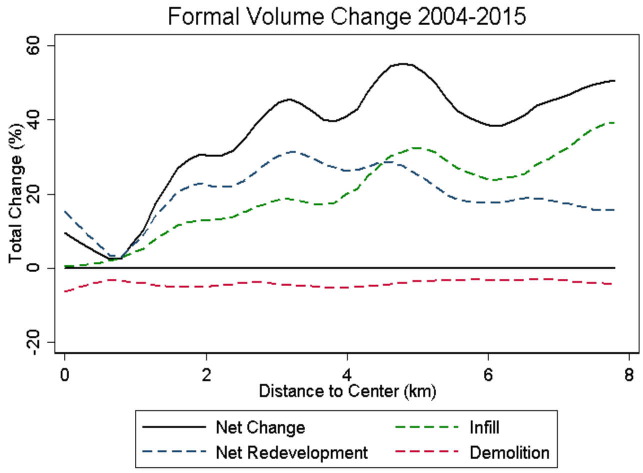


Figure 12. Changes in volume: decomposition



At 10km the total volume increase is about 125%. Infill and net redevelopment approximate 120 and 20% each.

On-Line Appendices

Appendix 1: Theory.

Derivation of equation (12) & (13): Derivation uses

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

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

and the fact that volume is optimised.

Proof of Proposition 2

Part (i) is simply the total differential of (11b) with respect to *x* and *t*. Part (ii) comes from differentiation of (12a) which implicitly defines the place at which first formal development occurs at date *t*, together with price equations (16), and the fact that Φ is independent of *x* and *t*, as given in proposition 1. Part (iii) of the proposition follows from eqns. (13a), noting that the price ratio on the left-hand side of (13a) now compares prices at different *x* and the same *t*, $p_F(x_i(t),t)\,/\,p_F(x_{i+1}(t),t) = e^{-\theta_F\Delta x}$, where $\Delta x \equiv x_i(t) - x_{i+1}(t)$, i.e. the distance between places undergoing successive redevelopments.[29] As expected, comparison of (17) with the equation in part (iii) of proposition 2 gives $\Delta x/\Delta\tau = \hat{p}_F/\theta_F$, this indicating how the prices that trigger redevelopment relate across space and across time.

Parameters for figures: Parameter values in figures 1 – 3 are: $\theta_F = \theta_I$ =0.072, γ = 1.71, α = 3.68. The discount rate is ρ = 0.057, $\hat{p}_I = \hat{p}_F = 0.0091$. Construction costs are $\kappa_I = 1.8$, $\kappa_F = 71.0$, and $a_I = 0.75$. These parameters come from Nairobi estimates as discussed in the text, units on the horizontal axis 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, $p(0,0) = 7.7$. In Fig 1 formalisation cost *D* = 0, and endogenous variables take values Φ = 20.6 and Δτ = 90.4. In Fig 2a, *D* is increased to 50 in distance band $x\ \varepsilon$ [4,8] and in 2b *D* takes positive standard normally distributed random variables times scale factor 10.

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 costs $w(t)T(x)$, 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,

[29] Whereas in proposition 1, the price ratio in equation (13a) was evaluated at given *x*, so $p_F(x,\tau_i)\,/\,p_F(x,\tau_{i+1}) = e^{-\hat{p}_F\Delta\tau}$.

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

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

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

Free choice of location type means that, at any occupied location, utility takes a common city wide utility level, $\bar{u}$, this giving bid-price

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

There are two types of housing, formal and informal ($i = F, I$), and we allow preference parameters, commuting costs, amenity and prices to be type specific. Formal housing offers amenity equal to unity, and we write the bid-price as

$$p_F(x,t) = \{w(t)T_F(x)B_F/\bar{u}\}^{1/\beta_F}.$$

Informal housing offers amenity $a(x, t)$, (depending on crowdedness, as in the text). We write the bid-price as

$$q(x,t) = p_F(x,t)a(x,t) = a(x,t)\{w(t)T_I(x)B_I/\bar{u}\}^{1/\beta_I}.$$

That is, we define $p_F(x,t) \equiv \{w(t)T_I(x)B_I/\bar{u}\}^{1/\beta_I}$ and express the bid-price as the product, $p_F(x,t)a(x,t)$, of amenity and an amenity free ('unit quality') price.

Constant exponential growth of the price of space is achieved by assuming that urban 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 rates $\hat{T}_I, \hat{T}_F$, so $p_i(x,t) = \left(\overline{w}e^{gt-\hat{T}_i x}/\bar{u}\right)^{1/\beta_i}$, $i = I, F$. This gives prices rising through time at constant rates $\hat{p}_I = g/\beta_I$, $\hat{p}_F = g/\beta_F$, and declining with distance, $\theta_I = -\hat{T}_I/\beta_I$, $\theta_F = -\hat{T}_F/\beta_F$.

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.

Appendix 2: Data

This section has three 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 2004 and 2015 depiction of building polygons.

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.

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 50 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 50 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 2004 we have no height data. To infer 2004 heights, we use what we think is an upper bound on height: the height of unchanged buildings, where we presume demolished buildings between 2004 and 2015 are likely to be of lower height than those which survive. To assign a height to a 3x3m square in 2004 in formal sector usage, we take the average height in 2015 of all buildings that were there in 2004 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 2004 and 2015? For demolition, at the 3x3m level the square is defined as demolition if its centroid is covered by a 2004 building which has been replaced by open space. Demolished coverage is lost 2004 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 2004 buildings; a 3x3m square is infill if its centroid is covered by such a building on 2015 where there was no building in 2004. 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 2004 buildings. So for each 150m150m meter square we have for redeveloped buildings, we have total coverage in 2004 measured at the 3x3m level (centroid covered by the old 2004 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 2004 so coverage change is likely to be an upper bound. Net volume change again assigns heights in 2004 to the 3x3m coverage based on neighbourhood averages for unchanged buildings and uses 2015 height information on the new buildings.

Overlaying Buildings

We match buildings across time by overlaying 2015 and 2004 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 2004 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 2004, including the Moi Air Base, and Nairobi State House. We drop all buildings in these areas for both 2004 and 2015, amounting to roughly 1,500 buildings from the 2015 data, and 100 buildings from 2004. Next we deal with overlapping shapes, an issue arising in the 2015 data, although not that for 2004.

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 2004 data does not. In order to avoid complications of censoring in the 2004 data, we simply drop all shapes that have an area of less than $1m^2$. We drop 2 small buildings in 2004, and 462 small buildings in 2015.

Another issue is that buildings are often defined as contiguous shapes in 2004, 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:

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

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 2004.

Accuracy Assessment

In order to assess the performance of the polygon matching algorithm we manually classified links between 2004 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 2004 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.

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

$$R = \frac{True\ Positive}{Positive\ Condition},$$

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

The confusion matrix in table 2 is done across all sampled buildings in 2004 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 2004 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.

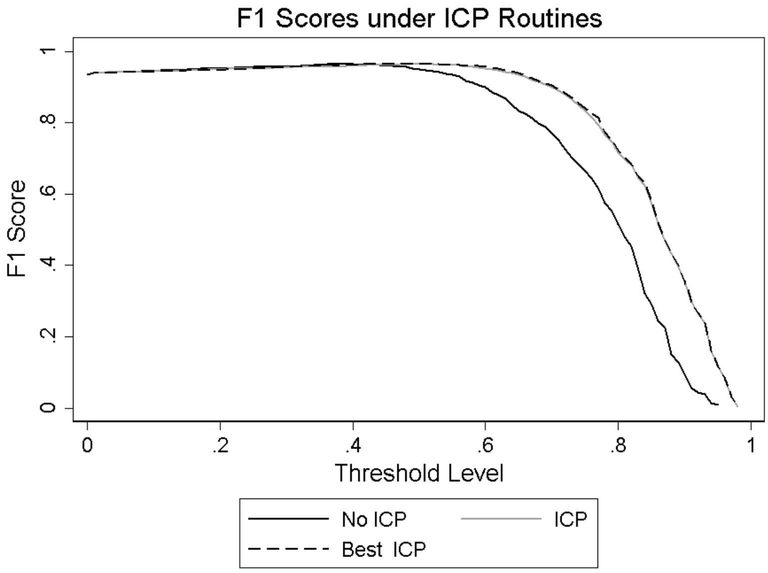
Appendix Table 2.1 – Mapping Correspondence 2004

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 2004

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

Appendix 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

Appendix Table 2.4: Full regressions for table 1						
	(1) Ln land sales price (USD per sq m.)		(2) Ln formal house-rent per cube vol. in $2015	(3) Ln slum house-rent per cube vol. in $2015	(6) Ln Formal redeveloped height; quantile: 80th percentile	(10) Ln Slum BVAR
Distance to centre (km)	-0.173***		-0.0618**	-0.00145	-0.101***	-0.0989***
	(0.0518)		(0.0304)	(0.0288)	(0.00494)	(0.0103)
SD Elevation (km)	-0.0239		0.0375	-0.0369	0.0207*	-0.0301**
	(0.102)		(0.0463)	(0.0491)	(0.0111)	(0.0121)
Elevation (km)	5.373***		1.834*	3.516***	314.8*	-1224.7***
	(1.801)		(1.087)	(0.635)	(186.3)	(233.3)
Lot size	-0.0306	No written tenancy	-0.324**	-0.270		
	(0.105)		(0.141)	(0.195)		
Lot size X Lot size	-0.00110	No piped water	-0.350**	-0.282**		
	(0.00158)		(0.136)	(0.108)		
Coordinates estimated	-0.460	Ln # Floors	0.280**	0.145		
	(0.374)		(0.120)	(0.135)		
		# Bathrooms				
Month=1	-0.310	One	-0.289*	-0.0138		
	(0.566)		(0.149)	(0.103)		
Month=2	-0.215	Two+	-0.224	0.196		
	(0.507)		(0.194)	(0.121)		
		Structure type				
Month=3	-0.547	Single-story Shared facil.	-0.0323	0.278**		
	(0.456)		(0.193)	(0.117)		
Month=4	-0.690	Multi-storey private bath	0.0463	-0.208		
	(0.531)		(0.243)	(0.203)		
Month=5	-0.862	Multi-storey shared bath	0.192	0.391***		
	(0.762)		(0.212)	(0.127)		
Month=6	-0.174	Shared house		-0.394**		
	(0.561)			(0.185)		
Month=7	-1.324*	Room in		-0.614***		

	(0.697)	house		(0.186)		
Month=8	-1.040**	Shack		-1.009***		
	(0.502)			(0.187)		
		Type of walls				
Month=9	-0.350	Brick/Block	0.411***	0.115		
	(0.518)		(0.145)	(0.264)		
Month=10	-0.619	Mud/Wood	0.705***	-0.696**		
	(0.563)		(0.215)	(0.304)		
Month=11	-0.271	Wood only	-0.0728	-0.150		
	(0.513)		(0.257)	(0.275)		
Month=12	-0.672	Corrugated	0.546***	-0.256		
	(0.489)	iron sheet	(0.193)	(0.256)		
		Mud/Cement		-0.701**		
				(0.335)		
		Tin	0.832***	-0.00643		
			(0.193)	(0.328)		
		Type of floor				
		Tiles	0.776***	0.770***		
			(0.229)	(0.289)		
		Cement	-0.0484	0.137		
			(0.0966)	(0.112)		
Constant	-1.365		-0.0751	-3.364***	2.750***	3.380***
	(3.233)		(1.781)	(1.125)	(0.310)	(0.397)
Observations	136		361	442	4589	983
R-squared	0.292		0.308	0.409	-	0.111

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

Appendix table 2.5: Welfare gains dropping slum rent intercept by one standard error						
Date of formalisation, *z*	0-1 km	1-2 km	2-3 km	3-4 km	4-5km	5-6km
PV (*z* = 2015)	1288	1083	911	766	645	542
PV (*z* = ∞)	344	312	282	256	232	210
PV (*z* = 2105)	393	352	316	284	255	229
Optimal *z* (year)	1900	1908	1916	1923	1931	1939
Opt redevelopment year (z+90)	1990	1998	2006	2013	2021	2049
D (low bound)	524	423	340	272	216	170

Appendix table 2.6: Welfare gains dropping slum and formal rent intercepts by one standard error						
Date of formalisation, *z*	0-1 km	1-2 km	2-3 km	3-4 km	4-5km	5-6km
PV (*z* = 2015)	1288	1083	911	766	645	542
PV (*z* = ∞)	398	360	326	296	268	243
PV (*z* = 2083)	612	538	474	419	370	327
Optimal *z* (year)	1963	1968	1973	1979	1984	1989
Opt redevelopment year (z+68)	2031	2036	2041	2047	2052	2057
D (low bound)	286	223	172	130	97	70

Appendix table 2.7: Welfare gains, using raw rent, or non-hedonic approach						
Date of formalisation, *z*	0-1 km	1-2 km	2-3 km	3-4 km	4-5km	5-6km
PV (*z* = 2015)	1288	1083	911	766	645	542
PV (*z* = ∞)	428	388	352	318	288	261
PV (*z* = 2105	471	423	381	343	308	278
Optimal *z* (year)	1921	1929	1937	1945	1953	1961
Opt redevelopment year (z+92)	2013	2021	2029	2037	2045	2053
D (low bound)	465	370	291	227	175	132

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