Online Appendices

for

Building the city: from slums to a modern metropolis

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

Appendix 1: Theory

A1.1 Further details of the model:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$+ v_F(x,\tau_2)\int_{\tau_2}^{\tau_3} p(x,t)e^{-\rho t}dt - k\left(v_F(x,\tau_2)\right)e^{-\rho\tau_2}\ .$$

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

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

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

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

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

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

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

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

The ratios reported in the text are 1.64/13.7= at the centre and 1.64/6.7 at 10 km out.

Appendix 2: Data

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

A2.1 Data sources

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

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

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

A secondary set of maps that we use comes from the Center for Sustainable Urban Development (CSUD) at Columbia University. The CSUD maps land-use in 2003, including slums, based on a more detailed, copyrighted, land-use map created by the JICA and the Government of Kenya under the Japanese Government Technical Cooperation Program which was published and printed by the survey of Kenya 1000 in March 2005 (Williams, et al. 2014). In principle, polygons are categorized as slums

if they seemed to contain small mostly temporary buildings that are randomly distributed in high-density clusters. We use this set of slums to offer a descriptive comparison of how slums have changed on the extensive margin, but for our analysis we defer to a single definition based on IPE due to discrepancies in the definition of slum across the data sources. We also make use of the CSUD land-use map to identify areas that we remove from our formal classification. The areas that we chose to remove are listed in Appendix Table A2.4 and are areas in permanent public use.

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

Vacant land price listings: We also require data on land values in order to calibrate the model, for this we rely on property values that have been scraped from property24.co.ke over the period September 2014 to November 2015. This data source provides us with vacant land listings recording information on asking price and plot area and location, all of which are provided for in over 80% of the listings. Prices are listed as 1,000 Kenyan shillings per acre and we convert these to 2015 USD by 4047 m^2 to an acre, and 100 Kenyan shillings to a US dollar. The locations are descriptive and so we entered geo-coordinates by manually searching the addresses and location descriptions. These listings are only found in the formal sector.

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

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

A2.2 Measures of cover and volume

Our unit of analysis is 150x150m grid squares. For calculating cover within the grid square in a usage, each of these is broken into 2500 3x3m cells and use type classified by what is at the centroid of the 3m square in each period. There are three uses: vacant land, slum area and formal. For each 150x150 square we sum across the 2500 cells to get total use of each type. Most 150x150 squares are either all slum or all formal sector. However, there are about 12% which are mixed grid squares, for which we record the cover or volume of slum and formal separately.

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

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

A2.3 Overlaying buildings

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

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

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

is the relationship between a set of candidates in one period to a set of candidates in the other time period.

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

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

Shape Matching Algorithm: After the pre-processing of each cross-section is complete, we run our shape matching algorithm to establish links between buildings across time periods. For any given building we consider 5 possible scenarios; that it has a link to no building, that it has a link to one building (one to one match), that it has a link to multiple buildings (one to many), that it is part of a group of buildings that match to one building (many to one), or that it is a part of a group of buildings that matches to a group of buildings (many to many). We follow 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:

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

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 2003.

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

Results: We first break down matches by their mapping type. There are five types of manual link: redeveloped/infill/demolished (0), one to one match (1), one to many match (2), many to one match (3), and many to many match (4). For the algorithm we further split (0) into infill/demolished (-1) and redeveloped (0). Appendix Table A2.1 shows the correspondence between the two mappings by building (a) and roof area (b). We can see that most errors come from the one to one matches, however,

the many to many matches have the worst performance. Overall the diagonal values are quite high, which means not only are we matching buildings well, but also the algorithm is recognising the clumping of buildings as a human does (bear in mind that, for example, the one to one matches which we 'misclassify' as many to many will still be classified as match in the final data). Finally, we have perfect correspondence for demolition and in 2015 nearly perfect for infill.

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

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

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

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

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

In our first attempt we arbitrarily picked 50% as a cut off of the overlay ratio. Here we take a closer look at this choice. Using our manually classified links we can maximize the F1 score with respect to the cut off. In Appendix Fig. A2.1 we plot the F1 score weighted by roof area against cut-offs of the overlay ratio for the 2003 data. We find that the highest F1 score comes just below 50% suggesting our first estimate was not far off.

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

Table A2.1: Mapping Correspondence 2003

a) Weighted by Building

	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4
Manual=0	280	413	41	24	30	10
Manual=1	0	12	729	20	3	3
Manual=2	0	10	13	315	0	6
Manual=3	0	13	3	0	145	1
Manual=4	0	82	69	57	47	84

b) Weighted by Area (m^2)

	Algo=-1	Algo=0	Algo=1	Algo=2	Algo=3	Algo=4

Manual=0	12708	27194	3303	3295	1982	686
Manual=1	0	691	112262	6055	347	242
Manual=2	0	2343	1545	94617	0	515
Manual=3	0	347	202	0	15562	23
Manual=4	0	5308	5543	10978	4429	8950
Mapping definitions: -1 demolition or infill; 0 redevelopment; 1 one to one match; 2 one to many match; 3 many to one match; 4 many to many match.						

Table A2.2: Matching all areas 2003

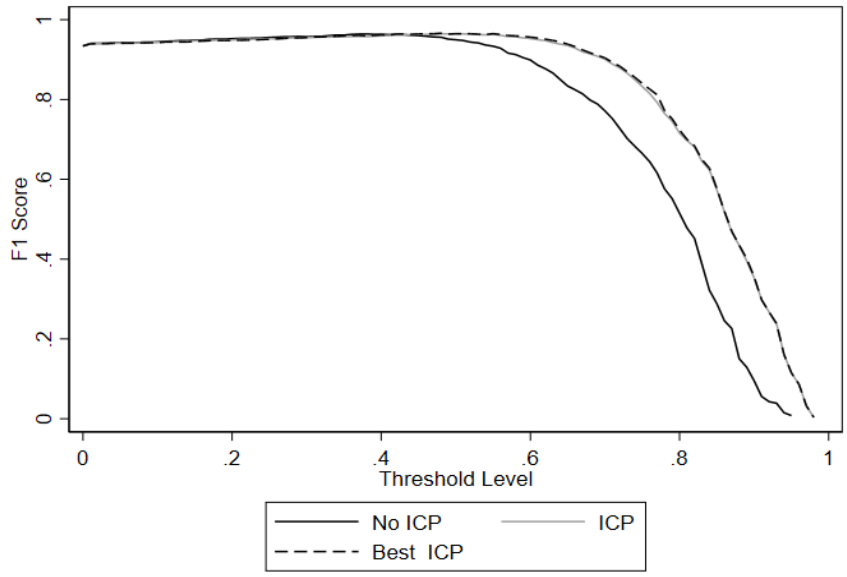
a) Weighted by Building

	Algo=0	Algo=1	Recall
Manual=0	413	105	0.8
Manual=1	117	1495	0.93
Precision	0.78	0.93	F=0.93

b) Weighted by Area (m^2)

	Algo=0	Algo=1	Recall
Manual=0	27194	9266	0.75
Manual=1	8689	261270	0.97
Precision	0.76	0.97	F=0.97

Figure A2.1: F1 Scores under ICP Routines



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

Table A2.3: List of public uses

Recreational a) Impala club, Kenya Harlequins, and Rugby Union of East Africa (0.14kmsq) b) Golf Course (0.9kmsq) c) Arboretum (0.25kmsq) d) Central park, Uhuru park, railway club, railway golf course (0.5kmsq) e) Nyayo stadium (0.1kmsq) f) City park, Simba Union, Premier Club (1.1kmsq) g) Barclays, Stima, KCB, Ruaraka, Utali clubs, and FOX drive in cinema (0.3kmsq) ***Undeveloped*** a) Makdara Railway Yard (1kmsq) b) John Michuki Memorial Park (0.1kmsq) ***Special use (includes poorly traced areas)*** a) State House b) Ministry of State for Defence c) Forces Memorial Hospital and Administration Police Camp	***Public utility*** a) Dandora dump (0.5kmsq) b) Sewage works (0.25kmsq) c) Kahawa Garrison ***Public use*** a) Communications Commission of Kenya (0.1kmsq) b) Langata Womens prison (0.2kmsq) 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

d) Langata Army Barracks e) Armed Forces f) Moi Airbase	***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

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

Monocentricity

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

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

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

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

Table A2.4: Monocentricy of population in developing and developed country cities		
	Developing country cities	Developed country cities

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

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

Real estate data and robustness

We use quarterly reports on the Nairobi housing market from HassConsult, a market leader in the Kenyan Real Estate sector.[1] These reports have been an important source of information for Kenyan real estate investors and homeowners for over ten years. The primary source of data is based on listings available in public sources (newspapers, magazines, social media, and online property portals). These listings are cleaned by HassConsult to remove properties and vacant land that are deemed atypical or duplicates, and then averaged in a way that ensures that price changes are not driven by composition effects.

The Hass Composite Sales and Lettings indices go back to 2000, and over its first 17 years used 163,000 cleaned listings after having dropped 15% due to duplications and 4% as outliers. Assuming that the collection rate has increased since the initial years, we can give a lower bound to the number of property listings used each year of 10,000. The Hass Land Composite Index goes back to 2007, and over its first ten years used 58,000 cleaned land listings. Assuming that the collection rate has increased since the initial years, we can give a lower bound to the number of land listings used each year of 5,800. The indices are broken down geographically into 18 Nairobi Suburbs[2] which is what we use and 14 Nairobi Satellite Towns[3]. We geolocate these neighbourhoods by a manual google maps search. Each neighbourhood had a distinct polygon which we were able to trace and georeference neighbourhood boundaries. We use the centroid of these boundaries for any spatial analysis of the HassConsult data.

We first looked for bubbles. Based on Glaeser, Huang, Ma and Shleifer (2017), bubbles are hard to define; but we looked at two common indicators for which there are data: overbuilding and price paths. The evidence suggests there is no overbuilding, and price increases are not high and see quite steady around the 2015 time period. First on sustained overbuilding, as noted in the Introduction, built volume increased over the 12 years by 59%, while population growth increased by 67%. Second, we investigated the rate of price increases using data from HassConsult. For looking at bubbles, researchers

[1] All original reports are available in the supplementary data and were last downloaded from the HassConsult report archive on May 5th, 2020: https://hassconsult.co.ke/real-estate/hass-index/16-report-archives
[2] The HassConsult suburbs are: Donholm, Eastleigh, Gigiri, Karen, Kileleshwa, Kilimani, Kitisuru, Langata, Lavington, Loresho, Muthaiga, Nyari Estate, Parklands, Ridgeways, Spring Valley, Upperhill and Westlands.
[3] The HassConsult satellite towns are: Athi River, Juja Town, Kiambu Town, Kitengela, Limuru, Mlolongo, Ngong Town, Ongata Rongai, Ruaka, Ruiru, Syokimau, Thika Town and Tigoni.

look at property value data, to see if rates of increases are very high or if a bubble has burst. For the full set of communities, prices rose 2.38 fold from 2007 to 2019 for an average annual increase of 7.5% (HassConsult, 2019). For years around our Q4 2015 data from the Q4 2014, Q4 2015 and Q4 2016 archived reports, the prior year's annual rate of property value changes are respectively 8.3, 9.6 and 7.6, so 2015 looks like a typical year (HassConsult; 2014, 2015a, and 2016). The official inflation rate varies a lot annually but averaged 7.4% p.a. from 2010 to 2019, with about a 6% rate around 2015. The property price changes are not indicative of any bubble; they seem steady and not high.

We also used the HassConsult reports on rental prices to look at real rates of increase in space-rents. One index uses the full sample of communities, where over time more properties in outlying lower quality housing areas are added. Rental prices for this sample of communities have risen 1.85 fold over 2007-2019, which gives a 5.3% average annual rental price increase (HassConsult, 2019). Another uses more traditional central communities without this bias.[4] House rental price increases for each of these communities range from 1.96 to 3.25 fold over the period 2007-2019, or 5.8% to 10.3% average annual rental price increases (HassConsult, 2019). Inflation was averaging 7-8% p.a., so the 0.94% annual rate of real space-rent increase that we use falls between these numbers.

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

A2.5 Other key regressions and figures:

Table A2.5: Full key regressions in Tables 1 and 2

	Ln land sales price (USD per m^2)	Ln Formal redeveloped height; quantile: 80th percentile	Ln Slum BVAR		Ln formal space-rent per m^3 vol. in $2015	Ln slum space-rent per m^3 vol. in $2015
	(1)	(2)	(3)		(4)	(5)
Distance to	-0.172***	-0.101***	-0.0948***	Distance to	-0.0855***	-0.00943

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

centre (km)	(0.0476)	(0.00521)	(0.0172)	centre (km)	(0.0310)	(0.0243)
SD Elevation (km)	-0.0114 (0.0531)	0.0222* (0.0116)	-0.0138 (0.0136)	Ln SD Elevation	0.00226 (0.144)	-0.138 (0.120)
Elevation (km)	0.00535*** (0.00178)	321.7 (197.9)	-1142.5*** (404.3)	Ln Elevation	2.145 (2.122)	6.294*** (1.204)
Lot size	-0.0302 (0.103)			No written tenancy	-0.428*** (0.133)	-0.320 (0.209)
Lot size^2	-0.00108 (0.00159)			No piped water	-0.335** (0.152)	-0.278** (0.109)
Coordinates estimated	-0.468 (0.372)			Ln # Floors	0.287** (0.108)	0.126 (0.126)
				# Bathrooms		
				One	-0.162 (0.144)	-0.0714 (0.125)
				Two+	0.00980 (0.220)	0.296** (0.132)
				Structure type		
				Single-story Shared facil.	0.0893 (0.175)	0.260** (0.116)
				Multi-storey	0.165 (0.200)	0.347** (0.133)
				Room in house		-0.603*** (0.169)
				Type of walls		
				Some mud	0.404* (0.215)	-0.839*** (0.211)
				Wood only	-0.370 (0.225)	-0.228 (0.234)
				Iron or tin	0.400** (0.196)	-0.375* (0.192)
Constant	-1.328 (3.234)	2.744*** (0.329)	3.163*** (0.700)	Constant	-12.46 (15.68)	-43.72*** (8.862)
Month F.E.s	Yes					
Observations	136	4621	958	Observations	361	439
R^2	0.292		0.104	R^2	0.244	0.391

Notes: Columns 1-3 are based on 2015 data for observations inside the 2003 extent of the city. Columns 4 and 5 are based on NORC data and restricted to observations inside the 2003 extent of the city. Standard errors in parentheses. Errors in col. 1 are clustered at the neighbourhood area from the on-line listing service; in col. 2 are not clustered; in col. 3 are clustered based on a 750m x 750m grid; and in cols. 4 and 5 at the census enumeration area. In column 2, for height regressions, we only include grids for which there is cover. For the BVAR regression, column 3, about 5% of

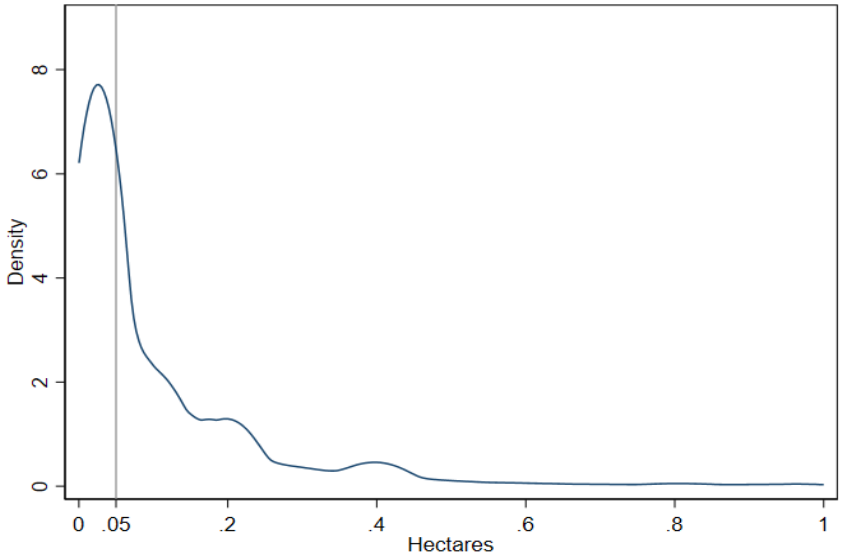
observations have no cover and hence volume (e.g. playing fields, overpasses, small parks and the like). A Tobit including these as 'censored' at 0 yields almost identical slope coefficients. Monthly fixed effects in col. 1 refer to the month of first listing. In predicting the intercept in Table 1 we assign the last 3 months a weight of 1/3 each and the rest 0. Significance levels for the null hypothesis that the true parameter is zero are denoted with asterisks: * p<0.10; **p<0.05; ***p<0.01.

Table A2.6: Robustness				
(1) Ln land price ($2015 per m2)	Standard	Adding distance to industrial	New centre definition	Census slum definition
	(1)	(2)	(3)	(4)
Distance to centre	-0.172	-0.187	-0.17	n/a
	(0.0476)	(0.0534)	(0.0474)	n/a
Distance to industrial centre		0.046		n/a
		(0.0974)		n/a
'Intercept' for typical item	7.254	7.308	7.242	n/a
	(0.277)	(0.279)	(0.275)	n/a
Observations	136	136	136	n/a
R^2	0.183	0.177	0.183	n/a
(2) Ln for. Redevelop. height; quantile: 80th percentile				
Distance to centre	-0.101	-0.107	-0.101	-0.09
	(0.0052)	(0.0095)	(0.0047)	(0.0049)
Distance to industrial centre		0.007		
		(0.0107)		
'Intercept' for typical item	3.315	3.368	3.298	3.225
	(0.036)	(0.072)	(0.032)	(0.033)
Observations	4621	4621	4621	4133
(3) Ln slum BVAR				
Distance to centre	-0.095	-0.082	-0.099	-0.094
	(0.0172)	(0.0242)	(0.0181)	(0.0205)
Distance industrial centre		-0.016		
		(0.0203)		
'Intercept' for typical item	1.275	1.162	1.286	1.12
	(0.118)	(0.199)	(0.120)	(0.152)
Observations	958	958	958	1590
R^2	0.101	0.101	0.105	0.087
(4) Ln formal space-rent m3 in $2015				
Distance to centre	-0.086	-0.048	-0.086	n/a
	(0.0310)	(0.0571)	(0.0307)	n/a
Distance to industrial centre		-0.051		n/a
		(0.0750)		n/a

'Intercept' for typical item	3.148	3.142	3.128	n/a
	(0.071)	(0.072)	(0.071)	n/a
Observations	361	361	361	n/a
R^2	0.216	0.216	0.217	n/a
(5) Ln slum space-rent m3 in $2015				
Distance to centre	-0.009	0.001	-0.014	n/a
	(0.0243)	(0.0416)	(0.0245)	n/a
Distance to industrial centre		-0.013		n/a
		(0.0450)		n/a
'Intercept' for typical item	1.886	1.859	1.886	n/a
	(0.058)	(0.087)	(0.058)	n/a
Observations	439	439	439	n/a
R^2	0.371	0.37	0.371	n/a

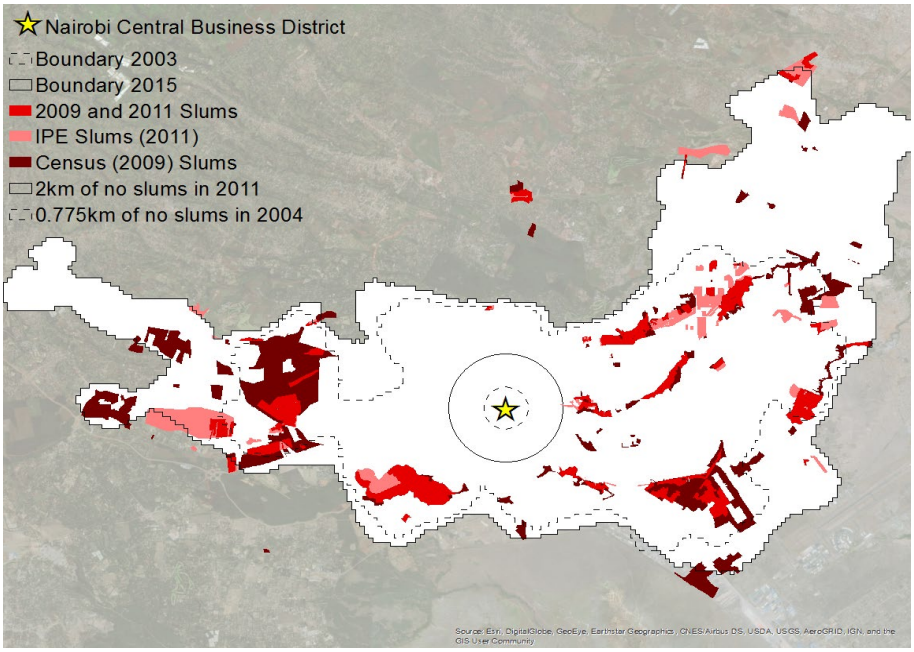
Notes: This table runs four robustness checks where each panel corresponds to a column in Table A2.5. Column 1 repeats the regressions from Table A2.5, column 2 adds a control for the distance to the industrial centre in kilometers, column 3 uses an alternative definition of the city centre based on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

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



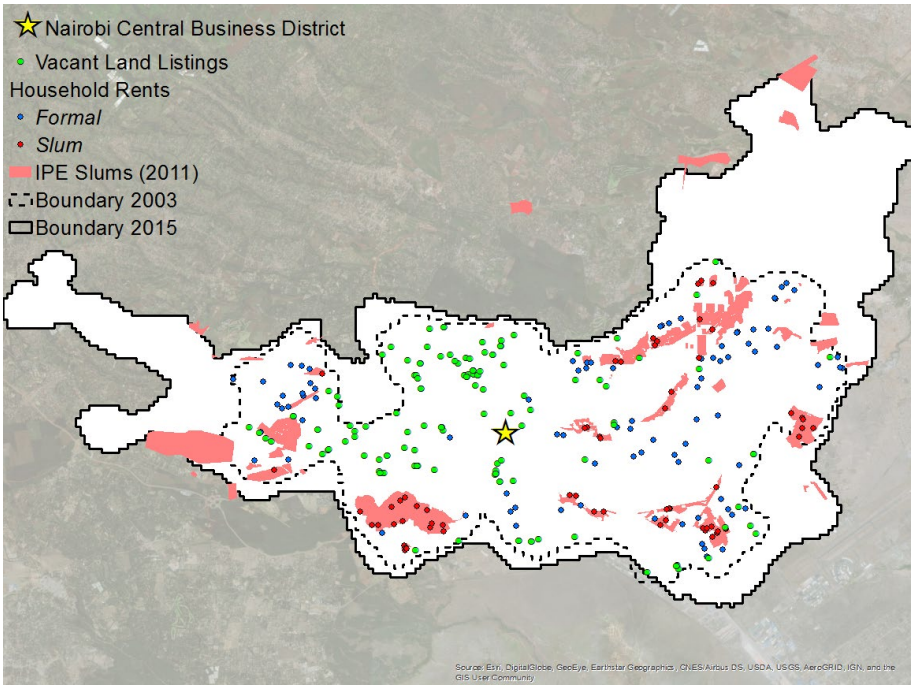
Notes: This figure shows the distribution of formal plot sizes from the cadastre. Plots above one hectare are truncated from the graph. The vertical line denotes 0.05 hectares. Of all recorded plots, 54% are smaller than 0.05 hectares, and modal density is at 0.018 hectares. Density is calculated using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

Figure A2.3: Census 2009 slums versus IPE slums



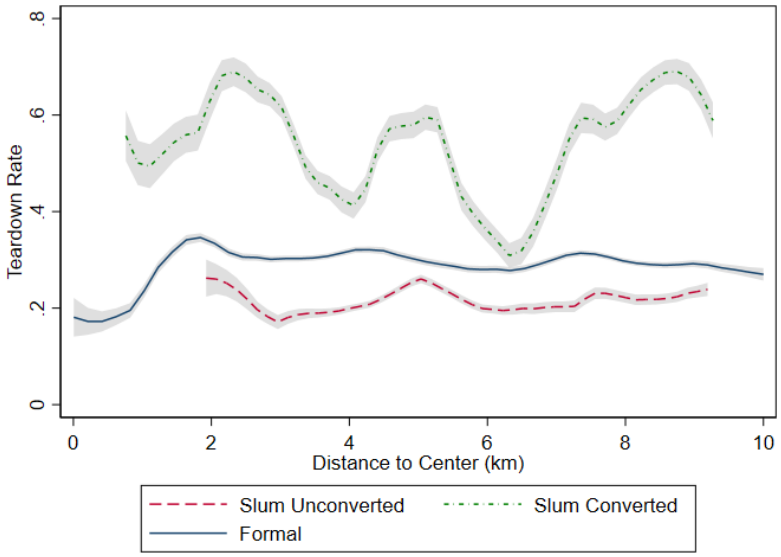
Notes: This figure maps two slum definitions; IPE from 2011 (light red), Census from 2009 (dark red), and the overlap of the two (red). The Kenyan National Cenusus Bureau defines all Enumeration areas as either planned or unplanned, even if the are unpopulated. Our 'Census' definition here maps all Enumeration areas that are unplanned and have at least 3,000 people per km^2.

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



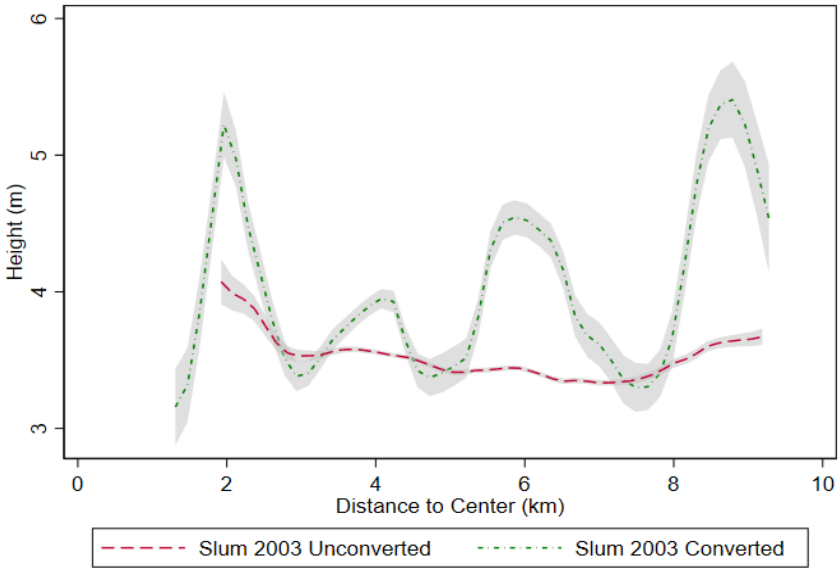
Notes: This figure maps formal (blue) and slum (red) household rents from the NORC survey and vacant land listings (green).

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



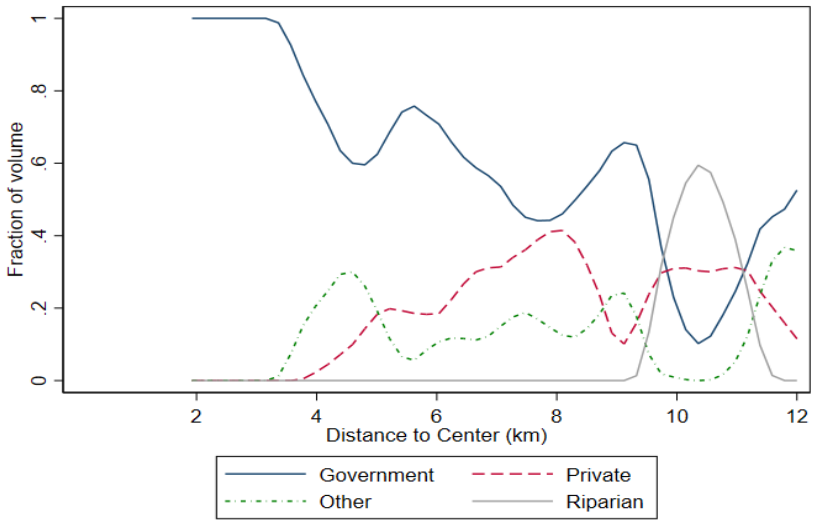
Notes: This figure shows teardown rates of 2003 buildings along distance to the city centre for the formal sector, and the 2003 slum sector. The 2003 slum sector is further broken down into unconverted slums (areas that remained classified as slum in 2011 by the IPE) and converted slums (areas that were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

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



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

Figure A2.7: Ownership of slums



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

Appendix 3: Land Quality.

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

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

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

Finally, we are concerned that results in our welfare analysis may be driven by land quality. That is there may be a correlation between government owned slum land (where formalisation costs are high) and low quality land (where construction costs are high for natural reasons). In particular, if central slum land was on worse land than neighbouring greenfield formal this could be partially responsible for the gap in land rents. We focus on infill land because, over the past period, it is the relevant comparison for what could have been slum redevelopment. In Table A3.3 we look at the sample of cells that are within 300m of a government slum boundary, restricted to either government slum cells or formal infilled land. We run a fixed effects regression so the analysis focuses on variation within arbitrary 300x300m blocks. Within these neighbourhoods we compare cells in and outside of slums at distance bins from the city centre for different land quality outcomes. Results show that, especially for slums inside of 6km, where we focus our welfare analysis, there is very little difference in land quality between slums and neighbouring formal land. Inside 6km slums are only found to be nearer to water in the bin from 4-5km, and in the bin 2-3km slums are actually less rugged and on higher local land in the 450x450m window. So for the 20 possible coefficients there is only one suggesting significantly lower quality in slums compared to neighbouring formal areas.

Table A3.1: Quality of infilled land

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

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

Table A3.2: Redeveloped height and land quality

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

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

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

	Elevation (m)	Water indicator 90x90m	S.D. elevation in 90x90m	S.D. elevation in 150x150m	Elevation difference to mean of 150x150m
Distance to Center=3-4	0.806*	-0.0866	0.468***	0.397***	0.145
	(0.474)	(0.0933)	(0.169)	(0.0899)	(0.498)
Distance to Center=4-5	0.487	-0.420**	0.439	-0.117	0.667
	(1.919)	(0.207)	(0.626)	(0.701)	(0.624)
Distance to Center=5-6	1.978	-0.249	0.458	0.504	-0.102
	(2.902)	(0.231)	(0.725)	(0.799)	(0.725)
Distance to Center=6-7	0.236	0.201	0.935	0.999	-0.0776
	(3.393)	(0.300)	(0.683)	(0.723)	(0.732)
Distance to Center=7-8	-2.433	0.0997	1.645**	2.403***	0.533
	(3.297)	(0.298)	(0.729)	(0.837)	(0.798)
Distance to Center=8-9	-1.291	0.0558	2.308**	3.038***	0.306
	(3.748)	(0.314)	(0.937)	(1.036)	(0.860)
Distance to Center=9-10	-2.200	0.0562	1.825**	2.298**	0.0277
	(3.747)	(0.310)	(0.844)	(1.025)	(0.907)
2-3 # Slum in 2015=1	-0.449	0.161	-0.0351	-0.261**	0.725*
	(1.353)	(0.149)	(0.129)	(0.116)	(0.372)
3-4 # Slum in 2015=1	-0.799	0.0844	0.0265	-0.125	0.0183
	(1.237)	(0.0828)	(0.301)	(0.296)	(0.217)
4-5 # Slum in 2015=1	-0.468	0.291***	-0.144	0.321	-0.157
	(0.603)	(0.0819)	(0.449)	(0.554)	(0.312)
5-6 # Slum in 2015=1	-0.561	0.0597	-0.0597	-0.183	0.315
	(1.434)	(0.0738)	(0.320)	(0.426)	(0.280)
6-7 # Slum in 2015=1	0.211	-0.194***	0.326**	0.408***	0.510***
	(1.120)	(0.0666)	(0.132)	(0.153)	(0.170)
7-8 # Slum in 2015=1	0.525	-0.0246	-0.103	-0.337	-0.0156
	(0.561)	(0.0720)	(0.207)	(0.405)	(0.181)
8-9 # Slum in 2015=1	-1.215	0.107	-0.696	-0.903	0.330
	(1.688)	(0.0729)	(0.589)	(0.593)	(0.351)

9-10 # Slum in 2015=1	-0.317	-0.0542	-0.157	-0.136	0.638
	(1.604)	(0.0452)	(0.286)	(0.369)	(0.399)
Observations	10167	10167	10167	10167	10167
R^2	0.995	0.416	0.369	0.584	0.081

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

Appendix References

Besl, P., and Mckay, N. 1992. “A Method for Registration of 3-D Shapes.” IEEE Transactions on Pattern Analysis and Machine Intelligence. 14(2), 239—256

Glaeser, E, W. Huang, Y. Ma, A. Shleifer. 2017. "A Real Estate Boom with Chinese Characteristics," Journal of Economic Perspectives, 31(1), pages 93-116

Google Earth V 7.3.0.3830. (February 2, 2002). Nairobi, Kenya. 1° 18’ 19.09”S, 36° 49’ 16.26”W, Eye alt 4.61km. DigitalGlobe 2017. http://www.earth.google.com [August 18, 2017].

Google Earth V 7.3.0.3830. (December 15, 2003). Nairobi, Kenya. 1° 18’ 19.09”S, 36° 49’ 16.26”W, Eye alt 4.61km. DigitalGlobe 2017. http://www.earth.google.com [August 18, 2017].

Google Earth V 7.3.0.3830. (February 14, 2004). Nairobi, Kenya. 1° 18’ 19.09”S, 36° 49’ 16.26”W, Eye alt 4.61km. DigitalGlobe 2017. http://www.earth.google.com [August 18, 2017].

Gulyani Sumila, Wendy Ayres, Ray Struyk and Clifford Zinnes.2012. Kenya State of the Cities Baseline Survey 2012-2013. Ref: KEN_2012_SOCBL_v01_M. World Bank & NORC

HassConsult, 2014. “House Price Index Quarter Four Report 2014”. Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

HassConsult, 2015a. “House Price Index Quarter Four Report 2015”. Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

HassConsult, 2015b. “Land Price Index Quarter Four Report 2015”. Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

HassConsult, 2016. “House Price Index Quarter Four Report 2016”. Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

HassConsult, 2019. “House Price Index Quarter Four Report 2019”. Nairobi, Kenya: HassConsult LTD. www.hassconsult.co.ke [May 5, 2020].

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

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

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

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