Evaluating Urban Planning: Evidence from Dar es Salaam[*]

Vernon Henderson (*LSE*)   Francisco Libano-Monteiro (*LSE*)

Martina Manara (*Sheffield*)   Guy Michaels (*LSE*)   Tanner Regan (*GWU*)

October 18, 2023

INCOMPLETE DRAFT

Abstract

Urban informality, which is prevalent in Africa's rapidly growing cities, can reduce private investments, lower tax bases, and exacerbate urban disamenities. A key policy tool to address this problem is greenfield urban planning where governments purchase cheap agricultural land on the urban fringe and partition it into planned, surveyed, and titled de novo plots, which people can purchase and build houses on. Yet, there is very little systematic evidence on the effects of de novo planning choices, such as the size and configuration of residential and non-residential plots. This paper studies the consequences of the planned layout of Tanzania's "20,000 plot" project, which provided over 36,000 residential plots in 12 project areas on the fringes of Dar es Salaam in the early 2000s. To study this project, we use new data from questionnaires and satellite imagery from circa 2020 and combine within-neighborhood analysis and spatial regression discontinuity designs. We find that small plots, which command higher land values and are built more intensively, are under-provided; smaller plots owners value homogeneity in plot size, grid structures of layout of blocks are valued as is access to major paved roads; and public service provision lags the plans. About half the plots are still unbuilt; yet the areas nevertheless attract highly educated owners. These findings suggest that while the project led to large overall gains in land value, significant improvements to planning may be possible.

KEYWORDS: Urban Planning, Economic Development, Africa.

JEL CLASSIFICATION: R58, R31, O18, R14, O21

1 Introduction

Urban planning plays important roles in shaping and regulating land use in developed country cities, where typically about half the land is in public use and, in the other half, private use intensity is regulated by zoning.[1] In developing country cities, large areas are unplanned or governed by ineffective regulations, resulting in informal construction and poor quality housing, which can reduce private investments, lower tax bases, and exacerbate urban disamenities (Scruggs, 2015; UN-Habitat, 2013). Large and rapidly growing cities in Africa, in particular, face a proliferation of slums (e.g., Castells-Quintana, 2017; Henderson, Regan and Venables, 2021). Given these problems, projects that offer effective planning for portions of a metropolitan area are a key option. Such projects involve 'de novo' urban planning, whereby greenfield agricultural land on urban fringes is purchased and partitioned into formal surveyed and titled plots with roads and perhaps some utilities and services. People then buy plots and build homes on them.

This de novo approach was pursued in many developing countries by the World Bank as part of its "Sites and Services" agenda during the 1970s and 1980s, and some African governments have implemented similar strategies more recently (e.g., Lamson-Hall et al., 2019; Choi et al., 2020; MLHHSD, 2018; United Republic of Tanzania, 2021). Evidence indicates that the de novo approach in Tanzania was cost effective and promoted higher quality housing in the long run than in neighboring laissez-faire informal developments or slums that were upgraded ex-post (Michaels et al., 2021). About 40-50 years after the de novo projects' completion, there is a strong demand to live in such planned areas, which command significant price premia. However, we know little about the *economic* consequences of specific choices made by urban planners in de novo areas, including the allocation of land for private uses (e.g., the sizes and layout of residential and commercial plots) and for public uses such as roads, public buildings, and open spaces. This general gap in our knowledge is glaring, since planning decisions have shaped cities for millennia and the discipline of urban planning is taught in hundreds of universities worldwide (Symonds, 2023).

In economics, debates on the respective roles of planners and markets in determining the allocation of land are longstanding. In seminal contributions, Smith (1759) critiqued the "man of system" organizing lives as "pieces upon a chess board", and Jacobs (1962) criticized the strict urban planning of Le Corbusier and Robert Moses. But economists have also long recognized the importance of planning in accounting for externalities (Davis and Whinston, 1962, 1964) and allowing space for public goods, such as roads (Solow and Vickrey, 1971; Dixit, 1973)[2]. Recent work (Bertaud, 2018; Duranton, 2017) emphasizes the

---

[1]For detailed picture of the USA and Canada in the mid-20th century see American Planning Association (1950).

[2]In developed countries, roads alone can take up as much as 20-30% of the urban space (American Planning

2

challenge of balancing market-based development, which reflects people's preferences and information, against planning, which defines the "rules of the game" (e.g., property rights) and accounts for public goods, externalities, and distributional issues. While urban planners and economists could learn from each other about how to improve city design, such mutual learning is as yet limited. The stakes are high, as cities concentrate a large and growing share of the world's population and play an outsized role in the global economy (Fujita et al., 1999; Glaeser, 2012; Moretti, 2012).

This paper asks: how does the layout of de novo neighborhoods affect housing outcomes? We consider choices that planners make on the sizes of residential plots, their layout relative to each other, and the allocation of land non-residential uses including roads and public and commercial uses. The outcomes we study include bare land values measured directly using market transactions; time of development; housing investments measured using satellite imagery; and educational attainment of households, reflecting residential sorting, measured primarily using questionnaires we develop. We examine these outcomes using within-neighborhood variation, including both OLS and spatial regression discontinuity design, in the context of the "20,000 Plots" project, which the Tanzanian government implemented on the fringes of Dar es Salaam from around 2000-2010. For brevity, we often refer to the "20,000 Plot Project" as the "20k Project" or simply "20k". This project delivered around 36,000 residential de novo plots (almost twice the number initially intended) in 12 project areas on the fringes of Dar es Salaam, taking up a total of 75 square km.

We find that the government's de novo investments - which cost just under $1 (USD 2021) per square meter of residential plot - were rapidly recouped by the purchasers' payments. Bare land values then increased sharply in real terms, and are now double those in nearby informal areas, although there are large variations in value gains across neighborhoods and within them. We estimate that the elasticity of price per sqm with respect to plot size is approximately -0.5, so land in larger plots is much less valuable. We also find that larger plots have relatively more open space, attract better educated owners, and have overall lower population densities. These results are consistent with an over-provision of large plots, probably due to implementation of colonial-era planning rules, which have largely persisted until the present day.

We also find evidence that smaller plots are more likely to be built on and are built on more intensively when they are located next to plots of similar size, suggesting a role for the planned layout of the neighborhoods and the potential segmentation of areas into blocks of homogeneous plot sizes. However, surprisingly, that desired homogeneity is not apparent for larger plots. Layout also matters in other dimensions: rectangularity of residential blocks

Association, 1950), but in developing countries this figure is typically lower (e.g., Bertaud (2018), Figure 5.11)

and grid-type alignment of blocks are valued.

Finally, we find that land values and construction patterns reflect valuation of natural amenities (elevation) and disamenities (ruggedness as well as proximity to rivers, streams, water and wetlands, in a city prone to flooding). Proximity to pre-existing, large paved main roads is highly valued. However most planned amenities are not. The likely explanation for this is that most planned amenities, other than roads, have not yet been provided, more than a decade after the project was completed and its plots were sold.

Besides the research on planning noted above, our paper is related to several other strands of literature. First is the literature on the colonial origins of African institutions and their impact on outcomes today (Acemoglu et al., 2001; Baruah et al., 2021), which we contribute to by studying the impact of planning regulations, which originated from British colonial rules. There is also a descriptive literature on land use policies in Tanzania (Kironde, 2006, 2015; Tiba et al., 2005; Mwiga, 2011; Msangi, 2011), to which we contribute by evaluating the 20k project quantitatively.

There is recent work on land-use regulation differences at boundaries (Turner et al., 2014; Kulka, 2019; Chiumenti et al., 2022; Shertzer et al., 2018). Our contribution is to study specific variation in plot sizes and configurations (rather than bundled variation in regulations), and to examine de novo projects, which we study in a developing country. Also related is a descriptive literature documenting the prevalence of large minimum plot sizes in Africa (e.g., Konadu-Agyemang, 2001; Ahmad et al., 2002; Gulyani and Connors, 2002; Collier and Venables, 2014; Tipple, 2015), and discussions of minimum density in US (Glaeser et al., 2005; Gottlieb, 2018). Our contribution is to study the effects of these large minimum plot sizes.

A related but separate strand of the literature studies the costs of overly segmented small plots near large city centers (Harari and Wong, 2017; Yamasaki et al., 2023), but we focus on plots in suburban settings, where large tracts of land are not scarce. There are also descriptive papers showing that unit prices fall with plot size in suburban areas in developed countries (e.g., Asabere and Colwell, 1984; Brownstone and De Vany, 1991; Colwell and Sirmans, 1993; Kolbe et al., 2021; Combes et al., 2021). We provide tighter evidence on the effect of plot size starting from greenfield development.

There is also literature studying the valuation of local amenities (Asabere, 1990; Gibbons et al., 2013), some of which emphasizes sorting (Epple and Sieg, 1999; Bayer et al., 2007; Diamond, 2016; Song, 2021). We study planned amenities of different types and shed light on sorting that follows planning. Finally and tangentially, there is work on the value of planning combined with property rights protection, especially historically or in developing countries (De Soto, 1989; Libecap and Lueck, 2011; Fuller and Romer, 2014; Angel, 2012). We

contribute by looking inside formal areas and studying the consequences of specific planning decisions.

The remainder of our paper is organized as follows. Section 2 discusses the institutional background and the economic issues; Section 3 discusses the data; Section 4 presents the research design and our empirical findings; and Section 5 concludes.

2   Background

2.1   A brief history of urban planning

People have been planning towns and cities for millennia. Mohenjo Daro, in the Indus Valley (c. 2500-1900 BCE) had orthogonal main features and semi-orthogonal neighborhoods (Lawler, 2008; Smith, 2007). Cities in Mesopotamia, Assyria, and Egypt also had orthogonal features (Paden, 2001). Paden (2001) discusses how Ancient Greek cities initially developed organically around their acropolis, but in the fifth century BCE Hippodamus was credited with designing his hometown, Miletus, and the port of Athens, Piraeus, with gridded layouts. Miletus (Panel A of Appendix Figure A.1) had grids of two sizes, one agora (public space) near its harbor for commercial use, and another closer to the center, probably for wider civic uses, and its public buildings included a theater, a gymnasium, and a stadium. Gridded cities spread through the ancient world through the empire of Alexander the Great and later the Roman Empire. Over two millennia later, Howard (1902) set out de novo plans for "garden cities" (Panel B of Appendix Figure A.1), which influenced subsequent suburban planning in many countries (Hall and Tewdwr-Jones, 2019). Nowadays, exclusionary zoning is common in many cities (e.g., in the US), and there are hundreds of graduate programs for urban planning around the world (Symonds, 2023). Despite this rich history, systematic economic evaluations of urban planning are scant, especially for de novo planning (Bertaud, 2018).

2.2   Urban planning in Dar es Salaam

In Dar es Salaam, a grid was planned near the historic core under German rule in the late nineteenth century (MLHHSD, 2018). Nevertheless, under German and later British colonial rule, different planning and building standards developed for different parts of the city: the European parts had strict planning standards with large plot sizes; the Asian parts had lower standards but were still planned; and the African parts were unplanned (Kironde, 1994). After Tanzania's independence in 1961, Dar es Salaam's urban population grew from less than 280 thousand in 1967 to nearly 8 million today - an almost thirty-fold increase. The formal standards of planning were retained in theory, sometimes with

new justification (Kironde, 1994), and a series of masterplans were developed (MLHHSD, 2018). In practice, however, most of the city was comprised of informal settlements and, even in formal settlements, plots and building footprints were generally in violation of zoning regulations.

From the 1970s, some de novo planned neighborhoods were developed, notably through collaboration between the Tanzanian government and the World Bank, as part of the latter's Sites and Services projects (World Bank, 1974, 1977, 1984, 1987). Such projects purchased cheap greenfield agricultural land on (what was then) the city fringes and laid out planned, surveyed, and titled plots with a modest bundle of services - mainly unpaved roads. Similar projects were developed in Indonesia, Vietnam, Myanmar, Uganda, Kenya, Nigeria, Ethiopia, Egypt, and India (Bolton, 2020), as well as Latin America (Grimes, 1976). The World Bank retreated from this agenda in the late 1980s due to criticism that the projects had poor repayment rates and did not serve the poor (Mayo and Gross, 1987; Buckley and Kalarickal, 2006). As noted above, recent evidence, however, shows that the de novo approach resulted in better housing quality and price premia, compared to other neighborhoods (Michaels et al., 2021). Similar de novo approaches are still seen by some African governments as cost-effective, and have been explored not only in Tanzania, but also in Rwanda and Ethiopia (Lamson-Hall et al., 2019; Choi et al., 2020; United Republic of Tanzania, 2021).

2.3 The 20,000 Plots Project

The focus of our study is the "20,000 Plots" project, which was initiated by the Tanzanian government in response to perceived unmet demand for formal de novo plots in late 1990s. This project, which was implemented from 2000-2010, and as noted above delivered around 36,000 residential plots in a 75 square km area. The residential plots took up about half the total project area (around 38 square km) and were formally surveyed and titled. The project also provided around 1,500 non-residential plots (spanning around 12 square km), which were designated for public and commercial uses. The remainder of the project area (around 25 square km) was taken up by roads and shoulders, most of which were unpaved, and by land deemed hazardous (e.g., in or right next to streams or water bodies) which was left empty. [3]

Figure 1 contains maps of the project areas, which were mostly near the fringes of Dar es Salaam. Like the sites and service projects in Dar es Salaam from 40-50 years ago, the anticipation is that eventually the city will expand outward and the 20k projects will no longer be on the fringe of the metropolitan area. The maps show the pre-existing main paved roads and the boundary of the Dar es Salaam metropolitan area. The government set

---

[3]We discuss the sources and procedures we use to map the 20k areas in the Appendix.

a fixed price per square meter within each of the project areas to cover the project costs; most of the variation in prices across project areas seen in the first price map likely stems from the higher price of land near the coast (Mwiga, 2011).[4]

Of the concerns that halted the World Bank's Sites and Services projects, the 20k project adequately addressed the first - cost recoupment. The total cost of the project was around $30 million in USD 2021, or $0.8 per sqm of residential plot.[5]. The initial phase of the project was financed by an internal government loan from the Ministry of Finance, which had to be repaid quickly. This constrained the planning and sale process, but the plots were sold, and the entire cost recouped (Tiba et al., 2005).

But the second limitation of Sites and Services, that they did not cater to the poor, was not addressed. While the mean price per square meter of residential plot in the 20k project was lower by about an order of magnitude than in Tanzania's Sites and Services project (Michaels et al., 2021), the speed at which the plots had to be sold entailed a procedure that made them inaccessible to most and also resulted, reportedly, in a disproportionate share going to government officials. [6]

In addition to the rushed acquisition procedure, and the difficulty of accessing credit, another impediment to the purchase of de novo plots by poorer Tanzanians was their large size (and resulting high cost). As discussed above, large minimum plot sizes are common in former British colonies, especially in Africa (e.g., Konadu-Agyemang, 2001; Ahmad et al., 2002; Gulyani and Connors, 2002; Collier and Venables, 2014; Tipple, 2015). In Tanzania, large minimum plot sizes were retained long after independence, with different justifications (Kironde, 2006). When the 20k project was implemented, formal plot sizes in Tanzania ranged from 400-4,000 square meters.[7]

2.4 Economic issues of de novo planning

The large sizes of formal plots raise important economic aspects of urban planning. First, governments may want to increase land values to generate revenues and increase their tax

---

[4]We include in our analysis of the 20k project one area, Malindi, which was developed from 1998 and later integrated into the project, although we do not have the initial government-set price for this area. The 20k project also provided a few thousand additional plots in other cities in Tanzania, but we only know precise locations of plots in Dar es Salaam.

[5]Part of the improvement in this process compared to earlier "Sites and Services" projects was due to the use of Global Positioning System (GPS) technology, which made surveying quicker and cheaper (Tiba et al., 2005).

[6]Prospective buyers had to collect application forms from municipalities or the Ministry of Lands, fill them in, and submit them to municipal land office. Priority was then given to those who: (i) had owned land in this specific area; (ii) could pay for plot type they wanted to purchase; and (iii) met gender and disability criteria. Successful applicants had to collect an acceptance form and start making the payment within 14 days. Finally, failure to complete the payment and finalize the transaction within 60 days resulted in reallocation of the plot to another potential buyer. Given the limited access to credit, this process made purchasing the plots difficult, especially for the poor.

[7]The minimum has since been reduced to 300 sqm (MLHHSD, 2018).

base (e.g., Besley and Persson, 2014), but there is no evidence on the effects of de novo plot size distribution on land values.[8] Second, formal plots are scarce, so fallow plots are wasteful, a concern which has been raised by the Tanzanian Minister for Lands, Housing and Human Settlements Development (Jamal, 2018). But there is no systematic evaluation of the effects of plot sizes and other planning decisions (e.g., layouts and amenities) on the extensive margin of homebuilding. Third, intensive margin development may vary by plot size, affecting the density of population and economic activity, with implications for agglomeration benefits (e.g., Henderson, Nigmatulina and Kriticos, 2021). Fourth, neighbors' plots may affect an owner's plot price (e.g., Turner et al., 2014), so the configuration of plots may matter, but again we have no clear evidence on the externalities from plots of different sizes. Finally, plot sizes may affect affordability, and hence who is able to purchase and benefit from the plots. Overall, despite the potential importance of de novo planning decisions on the sizes and configurations of residential plots, we know surprisingly little about their implications.

Another important aspect of de novo planning is the allocation of land for roads and other public and commercial uses. While restrictive zoning is common, and there is some evidence on its impact in developed economies (e.g., Shertzer et al., 2018), we are not aware of any evaluation of the impact of de novo planning on the provision of services and their valuation.

Finally, it is important to understand the respective roles of governments and markets in providing de novo projects. In Tanzania, de novo provision has until recent years been dominated by the government, but in the last decade several firms have attempted to enter this market. Understanding which aspects of provision work well and which ones do not is something that would help improve both private and public initiatives..

3 Data

3.1 Data sources

This paper uses several different data sources, including project maps, high-resolution satellite imagery, and interviews, questionnaires, and enumerations that we conducted, as discussed below and in further detail in the Appendix. We obtained maps covering all the project areas. These include town planning drawings and survey maps, which cover eleven of the twelve project areas, as well as cadaster data that cover the remaining area. A detailed description of these data sources is outlined in Appendix Section A.1.

---

[8]There are papers looking at correlations between plot size and land values, mostly in developed countries (e.g., Asabere and Colwell, 1984; Brownstone and De Vany, 1991; Colwell and Sirmans, 1993; Kolbe et al., 2021; Combes et al., 2021)

We also obtained color satellite images with a resolution of approximately 0.5 meters. Spatially, these images cover the project areas and a buffer of 500 meters around them. Temporally, the images span the years 2019-2021. We paid a Nairobi-based firm, Ramani Geosystems, to digitize information from these images, including the footprints of buildings and fences around plots. We also use the images to trace modern roads and classify their surfaces and widths. The satellite data and its derivatives are detailed in Appendix A.2.

We use other sources to characterize the underlying locational fundamentals. These include a digital elevation model (United States Geological Survey, 2000), which we use to calculate elevation and ruggedness, and Open Street Map (OpenStreetMap contributors, 2017), which we use to determine the locations of rivers or streams and water or wetland. While these likely mostly reflect 'first-nature" differences across locations, it is possible that they were partly altered in the process of preparing the project areas and constructing housing (e.g., if some areas were flattened). These data are detailed in Appendix section A.3.

We also collected additional data on the current state of the 20k projects. First, we held interviews with (i) local experts and (ii) leaders of 34 mitaa (local administrative areas), whose jurisdictions span almost all the 20k plot areas and adjacent non-20k areas. Second, we administered questionnaires to (i) local real estate agents ('madalali' in Swahili, singular 'dalali'), who provided us with sales prices for individual plots in 20k areas and nearby non-20k areas and (ii) residents in over 3,200 households within 20k areas. Finally, we conducted enumerations of (i) the 20k non-residential plots and (ii) the public transport access points. We explain the data gathering procedures, including the sampling frame, the interview details and protocol, and the questions in Appendix section A.4.

3.2 Plots and land uses in 20k areas and outside them

The project plans give us a detailed view of the plots and the uses the planners had intended them for. [9] We classify plots as residential when they are not designated for non-residential use and have an area of no more than 4,000 square meters (which was the formal maximum size at the time of the 20k project). The remaining plots we define as non-residential, which includes both private and commercial uses, as described in the appendix.

Figure 2 offers a concrete example of our data for part of a relatively well-off area in the northern fringe of Dar es Salaam, Mbweni Mpiji. Panel A shows the project plan, with residential plots of different sizes grouped in residential "city blocks", which we call insulae, which are typically separated by roads (not shown on this version of the plan).[10] The plan

---

[9]The appendix describes how we combined different project maps to determine the planned uses.

[10]We use the term insulae (singular - insula) to describe sets of contiguous (planned) plots, following the common

also shows non-residential insulae with a variety of intended uses. Finally the figure gives an example of a super-insula, a collection, or neighborhood of insula of similar size plots, as defined later. Panel B shows an image of the same area, illustrating that housing units mostly conform to the planned plot outlines, although a minority of the residential plots shown in the image are unbuilt. The share of built non-residential plots in this image, is, however, considerably lower. The areas between insulae are largely taken up by unpaved roads, as planned.

Whereas Figure 2 shows variation within a 20k area, Figure 3 contrasts a 20k area with an area just outside it, in this case in a poorer area in southern Dar es Salaam - Tuangoma. Panel A of Figure 3 shows the area as it was in June 2001, when it was still agricultural and largely empty. Overlaid on the same image are the boundary of the planned area (in red) and the plot boundaries within it (in white). Panel B shows the same area and plan roughly 20 years later, in 2021. Within the planned area, buildings are large and regularly spaced out, with roads between the insulae, again conforming to the plan. In contrast, outside the planned area, the informal looks very different: buildings are less uniform in size and typically smaller; some are bunched together irregularly, and many seem inaccessible via roads. This visual illustration highlights some of the consequences of de novo planning.

3.3 Dataset construction

To construct our main dataset, we consider as our units of analysis small square parcels of land ("gridcells"). These parcels could differ in their first-nature locational fundamentals - for example, some may be more elevated or closer to streams. The square parcels may also be treated differently by the planners, who may designate them as part of residential plots of different sizes or with different proximity to planned amenities. Our empirical methodology, which we discuss in the next section, focuses on disentangling the effect of different treatments of gridcells by the planners.

Concretely, we define as our units of analysis 20 x 20-meter square gridcells, which correspond to the size of the minimum formal plot (400 square meters). We identify each gridcell with its centroid and relate it to the plot and the insula in which this centroid falls. We focus our analysis on the approximately 95,000 gridcells whose centroids fall inside residential plots. Figure 4 shows the size distribution of the plots associated with all our gridcells, almost all of which are in the official size range of formal plots (400-4000 square meter). This figure also shows the official minimum thresholds for small (400 sq meters), medium (800 sq meters), and large (1600 sq meters) plots. As the figure shows, many plots are large

---

usage in Roman residential terminology (Storey, 2004), and avoid the term "blocks", which in Tanzania refers to a number of adjacent insulae.

even by developed country standards, exceeding 1000 sq meters (approximately 1/4 acre), despite the low levels of income in Tanzania. [11]

One of our main outcomes of interest is the real price of plots that were unbuilt ("bare land") when they were sold, based on the questionnaires we administered to the real estate agents and residents. Bare land prices are available for 998 residential plots (669 from the real estate agent questionnaire and the remainder from the resident questionnaire).

Other key outcomes include measures of housing investment taken from the satellite imagery: the share of a gridcell (restricted to residential plots within the gridcell's insula) that is built; an indicator for a gridcell's plot having any building whose footprint is at least 30 meters and whose centroid falls within that plot; the extext of investment measured by the footprint of development (for the 3 largest buildings) if the plot is built upon for the plot that the cell is in, the footprint size of the largest building in the cell's plot; and an indicator for multiple buildings in the cell's plot. The simple dynamics model presented next suggests that a plot is more likely to be built upon (developed by a certain date) if it has better amenities than a neighboring plot. It may also have more investment because it sells to a higher income consumer.

Other variables that are important in our analysis include measures of proximity to amenities and neighborhood fixed effects discussed below.

4 Modeling the development of 20,000 plot areas

In this section we model the development of 20k areas, which are greenfield areas outside the city of Dar es Salaam. People living in the city are offered plots of size l for sale in time 0 (e.g., in 2005) by private owners. These initial owners at time zero were the lucky/connected ones who got under-priced plots from the government before time zero (e.g., 2000) and are selling to people who will actually move to these areas at some endogenous time, τ in the future. At τ, these purchasers will leave the city for the 20k area and invest a one-time irreversible k in housing capital on their plot.

The set of people in Dar es Salaam who would potentially want to leave the city to live in a "generic" 20k area have incomes distributed between $[\underline{w}, \bar{w}]$. In general plots in 20k areas are in limited supply, and only some people in the interval $[\underline{w}, \bar{w}]$ will live there. In equilibrium it will be those with higher incomes, for example those in the interval $[w_m, \bar{w}]$, where w_m is the minimum income person in the 20k area and $\underline{w} < w_m$. If the supply of 20k plots increases then the income interval covered will extend at the lower end. Those in the interval $[w_m, \bar{w}]$ face fixed supplies of plots of different sizes. We will solve an example with

---

[11]Very few plots are smaller than 400 square meters, and we are unsure as to why such small plots were delineated.

2 plot sizes, but will explain the direct generalization to a large number of different plot sizes. In the initial simple equilibrium we solve for, bigger plots will go to higher income people. Plots come with an amenity bundle B; initially all B's will be the same but then we will introduce differentiation in B's.

4.1 The consumer's optimization problem

Why do some people leave the city and move to the 20k area? In calibration, equilibrium unit prices of land in 20k areas will be cheaper than the city. But the key element in terms of timing is that we specify an amenity level in the city at time 0 of A which deteriorates at the rate θ; while, in 20k areas, the amenity level B does not change with time. $A > B$, so initially people stay in the city, but the deterioration in living conditions in the city (e.g., congestion, pollution) ultimately drives some people from the city. Of course one could think of this more broadly in contexts of, say, overlapping generation models, where people want to retire at some point to 20k areas and have family compounds there. One can also add escalating housing prices in the center city. We will discuss what happens if θ's differ across people of the same income.

The optimization problem of each person who moves is

$$\max_{h_1, z_1, k, z_2, \tau} \int_0^\tau [\varphi ln h_1 + \beta ln z_1 + A e^{-\theta t}] e^{-\rho t} dt + \int_\tau^\infty [\varphi ln(l^\alpha k^{1-\alpha}) + \beta ln z_2 + B] e^{-\rho t} dt$$
$$+ \omega \left(\int_0^\infty w e^{-\delta t} dt - \int_0^\tau (p h_1 + z_1) e^{-\delta t} dt - \int_\tau^\infty z_2 e^{-\delta t} dt - rk e^{-\delta \tau} - R(0) \right), \quad (1)$$

where h_1 and z_1 are housing and all other goods consumed while in the city, z_2 is all other goods consumed upon moving to the 20k area, k is the amount of housing capital invested at the time of move and τ is the date of move. l is the given (for the moment) plot for sale at price $R(0)$ in time 0. r is the purchase cost of capital; z is the numeraire, and p is the rental price of housing in the city (or the opportunity cost of holding a unit of housing in the city). For income for exposition purposes, we specific a constant wage, w. Note, we assume a perfect capital market so that people can smooth consumption. People who move at some point in the future, if they have a time invariant income stream, do save up while in the city and then invest k in housing structures when they move. However, with a perfect capital market all that matters is W, the present value of earnings over the lifetime. When we talk about high and low wage people, we really mean and high and low wealth people. Finally ρ is the personal discount rate and δ is the interest rate. We equate ρ and δ, which will imply that z consumption is constant over the lifetime, rather than rising or falling. The perfect

capital market assumption and equating of ρ and δ are simplifications that don't affect the generality of the principles we develop.

The first order conditions for this problem are in Appendix B from which we obtain expressions for ω, τ, and k. Differentiating (5) and (6) in the Appendix, by inspection one can show the items below. In this "comparative statics" exercise, we must hold w fixed, so sorting is not accounted for. That comes later when we do the full equilibrium. Here for later empirical application, we focus on interpreting regression results. So for example, what is the impact of amenity differentials across plots of the same size on the timing of development?

- Impact of amenities on τ: Holding l fixed, $\partial\tau/\partial B < 0$. However in any equilibrium, plots with better B's sell at a higher price. Nevertheless, we can show $\left.\frac{d\tau}{dB}\right|_{\bar{w},\bar{l},dR>0} < 0$. Better amenity lots are developed sooner.

- Impact of amenities on k: In (6) in the Appendix, B effects come indirectly through the impact of B on R and τ. Inspection shows that if higher B plots cost more (higher R) and are developed sooner (item above), the effect appears ambiguous. In the equilibrium example developed later, $dk/dB < 0$, as the higher R crowds out investment under the budget constraint.

- Impact of plot size on τ: In (5) in the Appendix, $\partial\tau/\partial l < 0$. By inspection, given that bigger plots will sell for more reinforces the negative effect, that bigger plots are developed sooner.

- Impact of plot size on k. As with amenities in (6) in the Appendix, plot sizes effects come through effects of plot size on the price paid and on timing of development. Again effects appear ambiguous, although in the equilibrium example developed later, $dk/dl < 0$ (for the same income) .

- In the full equilibrium, sorting is crucial and can lead to different observed relationships than the comparative static results above.

Although not observed in regressions, we note that, for the same price, people with higher θ's whose conditions deteriorate more quickly in the center city develop sooner (i.e., $\partial\tau/\partial\theta < 0$). Generally in equilibrium as θ varies, the price R for same l will not vary, having been "set" by a "marginal" consumer as defined below.

4.2   An equilibrium

We start with a simple example of an equilibrium with plots of 2 sizes, all with the same amenities, in order to illustrate the principles of a solution before moving to more complex situations. There are N_2 large plots of size l_2 and N_1 small plots of size l_1 for a total of N plots. People buying in 20k areas, who are thus in the interval $[w_m, \bar{w}]$, split themselves across plots, with the large plots going to the highest income people. Thus, people in the interval $[w(N_2), \bar{w}]$ have large plots and those in $[w_m, w(N_2))$ have small plots, so that the person with income $w(N_2)$ is at the margin between a big and a small plot. For the same supply of people, $w(N_2)$ will increase [decrease] as the relative supply of large plots in total N plots falls [rises].

There are two margins defining an equilibrium. The technical equations are given in Appendix B and here we describe them. The first margin concerns the lowest income person in 20k areas with income w_m. They determine the price of small plots such that this marginal consumer is indifferent between staying in the city forever versus moving to the 20k area at their optimal τ. The price of small plots, R_1 and the time of move are determined by two equations in these two unknowns, the one for τ which is (5) in the Appendix and the one equating utility from staying versus leaving the city for this lowest income 20k consumer, which defines R_1 and is (7) in the Appendix. All other consumers of small plots are intra-marginal and pay this same price R_1 in a competitive equilibrium. Those in the interval $(w_m, w(N_2))$ and also those consuming big plots in equilibrium get a surplus at the prices they pay relative to staying in the city forever.

The second margin is given by the person with income $w(N_2)$ who is the lowest income person on a large plot. That person pays a price R_2 which leaves them indifferent between being on a large plot versus being on a small plot and paying R_1. That condition is given by (8) in the Appendix and it, along with the equations for the optimal times of move if a person is on a small versus large plot, determine the equilibrium price of large plots and times of move for the marginal consumer with income $w(N_2)$. This person is better off than being in the city but is indifferent between a big and small plot. All consumers in the interval $(w(N_2), \bar{w}]$ are intra-marginal and get a surplus from being on a large as opposed to small plot, or being in the city.

Figure 5 illustrates an equilibrium. To calculate such an equilibrium we have to calibrate the model. We make the housing consumption share parameter, φ, and the land share parameter in housing production, α, both 0.3 based on the literature; and we make z's share parameter β to be 0.7. We set the real interest rate δ to be 0.06 from (Henderson, Regan and Venables, 2021), based on Kenyan data. We set the purchase price of capital r to be 16.7 so the rental rate on capital is $16.7 * 0.06 = 1$. We set the rental price of a unit of

housing in the city to be 2.56 which from the cost function implies a rental price of a unit of land of 3 in the city; the suburbs will be cheaper in equilibrium, in our first example about 0.4. Again this is consistent with urban land rent gradients (e.g. , (Henderson, Regan and Venables, 2021)). We set a small plot size to be l_1 = 800 and a large to be l_2 = 1600 in line with the data on plot sizes in square meters presented above. In the function $Ae^{-\delta\tau}$, we need a θ and we assume conditions in the city deteriorate at the rate 0.01. We arbitrarily set B = 0.691 and calibrate an A consistent with our data to be 1.048. [12] That means that, at time 0, center city amenities at 1.048 are higher than those in the 20k area at 0.691.

Finally we need to define the income margins. We assume the supply of plots versus consumers for the 20k area is such that the lowest income person is w_m = 4000, the highest income $\bar{w}$ = 21000. The choice of the income of the person indifferent between a small and a large plot, $w(N_2)$ reflects the relative supply of large versus small plots. We start with $w(N_2)$ = 7000, which in our example will reflect (as in the empirics) an excess supply of large plots such that the price per square meter will be lower on large versus small plots. We will then raise $w(N_2)$ which implicitly means lowering the relative supply of large plots until the price per sq meter is the same on the two plot sizes.

In Figure 5(a), we plot the present value of utility of people on small versus large plots by income, net of the utility they would receive if they stayed in the center city forever. We show that beyond w_m = 4000, people gain from being in the 20k area. People with income from 4000 up to 7000 gain more by buying a small plot than large plot. At $w(N_2)$, there is a "single crossing point" and beyond 7000, higher income gain more by being on a large plot than small. The marginal entrant to 20k area determines the small plot price of 5039 and the marginal person at $w(N_2)$ = 7000 yields a price of 8635 for large plot.

In parts (b) and (c) of Figure 5, we also show the equilibrium τ's and k's. In terms of τ, in Figure 5(c) , at w_m = 4000 people leave the center city at 9.5 years. As income rises on small plots, people leave later up to 15.1 years at w = 7000. At the switch to large plots at $w(N_2)$ = 7000, τ drops from 15.1 to 8.3 (so in fact some large plots are developed before any small plots). Then again τ rises with income maxing out in the example at about 19 years. In short there is not distinct relationship between τ and plot sizes. Some big plots are developed earlier and some later than small plots, with considerable overlap.

In terms of k, in Figure 5(b), investment rises almost linearly with income, but is little related directly to plot size per se. In fact the k on a small plot for w = 7000 is slightly larger than k on a large plot for w = 7000. There is substitution in production of land for capital; but the driving force is that the greater expenditure on land reduces money available for all

---

[12]From survey data discussed below, we focus on a typical consumer of a small plot with income 7000 USA dollars who would pay 5895 for that plot and will build at time τ = 15 years. Calibration then involves solving FOC 1-6 in the Appendix for the A, which satisfies these conditions given the θ and B. That value is A = 1.048

other purchases. In the data, k measures will rise with plot size but that occurs because higher income people on average live on larger plots and thus invest more.

Price per unit land by plot size

In this equilibrium the price per square meter on a large plot of 5.4 (= 8635/1600) is smaller than on a small plot at 6.3 (= 5039/800). This suggestions a relative over-supply of large plots. While avoiding specifying the exact income distribution relative to the supply of big versus small plots (N_2/N_1), by raising $w(N_2)$ we can mimic a reduction in the relative supply of large plots. At $w(N_2)$ = 20000, in this example the price per square meter on small versus large plot is approximately the same. Of course for the same number of consumers in $[w_m, \bar{w}]$, we have implicitly reduced the land area of the 20k settlement since plot size for most plots is cut in half. If we held the land supply fixed, we would lower the income of the marginal entrant at w_m in order to sell the increase in plot supply and that would lower the $w(N_2)$ where prices are equalized by the marginal person indifferent between a large and small plot.

Many plot sizes

The generalization to many plot sizes is direct, given the monotonic relationship between income and plot size. Suppose there 25 plot sizes, we then order plot sizes and incomes, allocating the largest plots to the highest income, and thus defining 24 margins like $w(N_2)$, now $w(N_2), w(N_3), w(N_4)....w(N_{24})$. We start with the first margin as in equation (8) in Appendix B and solve for that price for the second smallest size plot and then proceed sequentially to solve for successive prices as we move up the plot size and income cut-off scale. Prices and income intervals will rise with plot sizes and a figure corresponding to Figure 5 there would be 24 crossing points.

4.3 Amenity differentials

This section is critical to interpreting amenity effects on time of development and level of investment. For any plot size, amenities such as being near a mosque or having higher elevation will vary across plots even in the same insula and more widely across insula. How do we model the impact of amenity differentials? The key technical aspect is that, holding plot size fixed, higher income people will outbid lower income for plots with higher B's. So again there will be a marginal consumer. For example if we are looking at large plots, the base amenity, B_1 allocated to the lowest income interval of those on large plots. Suppose that the supply of larger B_1 plots is such that we cross into B_2 plots at $w(B_2)$. $w(B_2)$ is

then the marginal person on a big plot between a smaller versus bigger B. As before the price of a B_1 plot is set by the person marginal between small and large plots. Using the methodology in equation 8, we equate the utility from being on a large plot with B_1 and $w(B_2)$ with known price $R(l_2, B_1)$ to the utility from being on a plot with a higher B to solve out $R(l_2, B_2)$, as discussed in the Appendix.

Armed with this simple principle we show an equilibrium where that are four values of B (0.691, 0.705, 0.719, and 0.733) on each of big and small plots. To be able to better graph the equilibrium we increase the relative supply of small plots so that the cross-over in income to bigger plots is at 12000. An equilibrium exists as illustrated in Figure 6, where we make the cross over points on small plots into successively higher B's be 6000, 8000, and 10000 relative to the base value consumed by people with income 4000 to 6000. At 12000, there is cross over to bigger plots starting with the low B_1 value. After that B rises at 14000, 16000, and 18000. In part (a) of the figure, different color lines refer to how τ varies by income for different plot size-B combinations. The solid parts of the curves show the equilibrium τ's. At each successive point of increase in B for the same plot size, people develop earlier (or τ drops), as illustrated by the horizontal dashed straight lines at the switch point from B_1 to $B_@$ at w = 6000. At the cross-over into bigger plots as above, τ again drops, as it does at each successive B switch point. However between switch points as income rises, so does τ. Overall there is an ambiguity if one can't control for income. As B rises in the left part of the graph on average it looks like τ rises, while on the right part with the bigger plots it looks like τ declines on average as B rises. The conjecture is that if B varies almost continuously with many more than 4 values, as it does in reality, then the jump down in τ when B increases will dominate.

In Part B, we show that the equilibrium exists. The solid parts of the colored curves plot the outer envelope of realized utilities, where no person wants to switch from their plot size-B combination at equilibrium prices to a different combination. Note in particular, that higher income people on a large plot with low amenities, B_1, do not want to bid away small plots with high amenities, B_4.[13]

Finally we note we do not show the optimal k graph; we know that is dominated by income. k does drop by a tiny amount as B or plot size increases (holding w fixed) in our examples, but increases strongly with income.

---

[13]We explored this aspect with a number of examples asking whether higher income people with low amenities on a big plot would be willing to pay more for a high amenity small plot, than the lower income occupants. The answers in all examples was no.

4.4 Non-monotonicity in the income plot size relationship

Single crossing property models like this had consumers only differing by incomes. With the advent of empirical implementation of structural models, having a stochastic component is essential. Here we introduce differences in "tastes" as differences in θ, or differences in the rate at which conditions deteriorate in the center city for individuals. We know from the comparative statics that a higher θ results in a lower τ for any plot size, price and income. But higher θ people will be willing to pay more for plots compared to living in the city center. As in Epple and Sieg (1999), at the margin of entry to the 20k plot area there is a locus of combinations of w and θ, where theta rises (more eager to leave) as income declines, for people willing to pay the same price for a small plot. That is there is a heterogeneous set people who are at the margin between being in a 20k area and staying in the center city. [14]

4.5 Summary

Below we will run a number of regressions. Prices will rise with plot size and amenities as illustrated in the model outcomes. On the quantity side we will be looking at whether plots are developed by 2021 or not and at investment levels if developed, as key outcomes. It should be clear the model has something to say about what we should expect and that depends critically on what effects we are looking at: those of plot size versus and those from amenity variation.

What to expect in either case seems more clear for results on the time of development. Holding plot size fixed, if income is constant, then as amenities increase plots are developed sooner. However incomes do vary across the same plot size, where higher income means later development, all else equal. But if amenities vary at high frequency then we think the amenity and τ association will be negative. The effects of plot size on τ, holding amenities fixed, again are ambiguous as we can see in Figures 5 and 6. For the same amenities some small plots are developed before big ones and some later.

In terms of investment, given the strong association between investment and income, even if an increase in plot size for w fixed can in principle reduce investment by a tiny amount, given a lack of ability to control for income and the strong plot size- income association, we expect investment will rise with plot size. The effect of amenities is less clear. Holding plot size and income fixed, an increase in amenities reduces investment by a tiny amount. Again we have this positive association, holding plot size fixed, between income and amenities. However if amenities vary pretty continuously with R rising at each change, it could be that

---

[14]So above the margin is w = 4000 and θ = 0.01, with a price of 5039. At that price, for example, people with income's and θ's of about 5075 and 0.0095 respectively or 2925 and 0.011 respectively would be indifferent between staying in the city and moving to the 20k area.

the the investment amenity association is, if not negative, at least weak.

5 Research design and empirical findings

5.1 Methodology

Our empirical analysis aims to uncover the consequences of the planners' decisions to treat gridcells differently, for example by designating them as parts of residential plots of different sizes or exposing them to different planned amenities. Our analysis is aided by the fact that the project areas were largely agricultural areas (greenfields) circa 2000, which limits pre-existing difference in their uses. We typically begin with OLS regressions, which control for area fixed effects and observable physical controls, to mitigate the potential for confounding factors within project areas. We complement this approach using a spatial regression discontinuity (RD) design, which compares gridcells that are in very close proximity, and differ only in their treatment (e.g, whether they are part of a small or a large plot).

In our OLS analysis, we use the gridcell dataset to estimate regressions of the type:

$$y_i = \beta_1 \text{Plot\_size}_i + \textbf{Program\_area}'_i \gamma_1 + \textbf{Controls}'_i \delta_1 + \epsilon_{1i}, \qquad (2)$$

where y_i is the outcome (e.g., logarithm of plot price or plot price per square meter or some measure of housing) in the gridcell of its plot and Plot_size_i is the logarithm of size in sq m of the plot in which gridcell i's centroid falls. $\textbf{Program\_area}'_i$ is a vector of program area fixed effects, which focuses the analysis within relatively small areas, within which the initial government-set price per sq m was identical. $\textbf{Controls}_i$ include, in some specifications, a full set of interactions of program area fixed effects with 2012 enumeration areas and indicators for proximity to natural and planned amenities.[15] In price regressions, time period interactions by source of data (real estate agents or residents) are included. ϵ_{1i} is an error term.

We cluster the standard errors by insulae - the main units of plot size assignment (Abadie et al., 2023) - of which there are 3,231 in our full sample. To justify this approach, we note that insulae fixed effects have high R-squared - typically around 0.8 - in explaining variation in plot size assignment within project areas. Our estimates are, however, broadly similar when we cluster on smaller plot identifiers (of which there are roughly 36,000) or larger units, such as 158 interactions of program areas with enumeration areas in the 2012 census, 34 mitaa (local administrative units), or even the 12 project areas.

---

[15]As discussed above, it is possible that some "natural" amenities, such as ruggedness, may have been altered in the process of preparing the areas for construction, but typically the inclusion of these controls does not change the main estimates much.

In addition to the OLS regressions, we estimate spatial regression discontinuity (RD) models, which facilitate tight identification of own plot size effects. To implement spatial RD we define, as illustrated in Appendix Figure A.2, the boundary between insulae pairs, as a one-meter-wide line (typically on a road). Each gridcell's adjacent residential insula is defined by nearest boundary segment to the gridcell's centroid. The adjacent gridcell is the nearest gridcell (as measured by the distance between centroids) within the gridcell's adjacent insula.

We then estimate spatial regression discontinuity (RD) models of the type:

$$y_i = \beta_2 \text{Own\_larger}_i + \textbf{Program\_area}'_i \gamma_2 + \textbf{Dist}'_i \delta_2 + \textbf{Boundary}'_i \rho_2 + \textbf{Controls}'_i \kappa_2 + \epsilon_{2i}, \quad (3)$$

where Own_larger_i is indicator for gridcell i's plot being larger than the plot of the adjacent gridcell, as defined above; $\textbf{Dist}_i$ is vector of distances to insula's boundary, where we allow for separate distance measures for "large" and "small" side of each boundary; $\textbf{Boundary}_i$ is vector of fixed effects for boundary segments corresponding to each gridcell i; and ϵ_{1i} is an error term. In most specifications we estimate semi-parametric RD, restricting the analysis to gridcells within 100m or even 50m from their insulae boundary. In some cases, we focus on RDs with large (or small) plot size differences across adjacent gridcells' plots, or alternatively add an interaction of $Own\_l \arg er_i$ with the log plot size difference between the gridcell's plot and that of its adjacent gridcell's plot.

Our strategy is related to earlier research using spatial regression discontinuity (e.g., Dell, 2010; Turner et al., 2014; Michaels et al., 2021). We differ in our use of very small and uniform spatial units, in analyzing a greenfields setting, and in focusing on spatial discontinuities within small administrative units, where we can separately identify the effect of plot size from other urban planning instruments.

5.2 Empirical findings

Our discussion of the empirical findings begins with an examination of the aggregate land value gains from the 20k project, where we explore the appreciation of bare land prices over time and then compare the price of 20k plots to those of nearby non-20k plots. We then examine the effect of each plot's size on its value and ask whether the size of neighboring plots also matters. We also look at the effects of the sizes of pots and their neighbors on quantity outcomes, including the share of a gridcell that is built upon and whether a grid square's plot is built upon or not. This leads to a general discussion of the valuation of natural and planned amenities. Finally, we conclude with a discussion of the sorting of owners and other residents between and within the 20k project areas.

5.2.1 Aggregate land value gains from the 20,000 Plot Project

We begin with evidence about the appreciation of land values (prices) in the 20k project. Panel A of Table 1 shows significant real price appreciation across project areas - the mean increase in real log prices was almost 2, corresponding to a mean increase of over 600%. The panel also shows that there was considerable dispersion in price increases across project areas, although areas that were initially expensive did not seem to experience systematically different trends from those that were initially cheaper.[16]

While the land value gains discussed above are impressive, it is possible that some of them would have taken place even without the 20k project. To shed more light on the impact of the project on land values, Panel B of Table 1 compares bare land values in 20k project areas to those in nearby non-20k areas, which were sold by the same set of real estate agents. Column 1 shows that project area prices were about 125-130% higher than in nearby non-project areas, for both informal (unsurveyed and not titled) and "formal" (surveyed and titled) areas. Column 2, which adds fixed effects for the nearest 20k project area to control for neighborhood, shows the same gains relative to informal outside areas but smaller gains of about 70% relative to surveyed plots. While the number of plots sizes outside 20k project areas for which we have bare land transaction prices is relatively small (65 surveyed, 41 unsurveyed), the regression findings are broadly consistent with separate interviews of 34 mitaa leaders, who also estimated that the price of bare land in 20k plots within their mtaa was approximately 100% higher than in non-20 plots within their mtaa. These gains are very large, especially if we consider that proximity to the project areas may have increased the value of non-project areas.

To understand why land in 20k areas is more valuable within their mtaa jurisdictions, we asked the mtaa leaders "What factors or characteristics do you think determine the difference in the price of land in 20k versus non-20k areas? What are the main drivers?". We received answers from 31 out of 34 leaders, and these tended to emphasize the importance of two factors: property rights and access. First, among the leaders who replied, 24 mentioned property rights (of which 21 mentioned land titles explicitly). The main explanations that those leaders mentioned were reductions in boundary conflicts, increased tenure security, and increased access to financial credit (since formal titles can serve as collateral). Second, 23 leaders mentioned better access in 20k areas (20 mentioned roads specifically and 3 others mentioned access). The leaders who mentioned this suggested that non-20k areas tend to clog up over time, and some mentioned that this made it difficult to improve local

---

[16]We found also that inflation-adjusted prices were even higher (by about 20-30%) during the first decade of the project (2000-2010), before declining a bit and stabilizing at the high levels reported in Panel A of Table 1. We inflate historical prices up to the year 2021 using annual inflation rates all in Tanzanian Shillings as detailed in Appendix A.3.

service provision. Other explanations, apart from property rights and access, were much less common in the mtaa leaders' interviews.

5.2.2 The implications of own plot size

To explore the implications of a plot's own size, we begin with regressions of specification (2), focusing again on plots that were sold as bare land. While we are interested in the elasticity of log price per square meter with respect to plot size, which we explore below, we begin with regressions where the outcome is the logarithm of plot price. We do this to avoid potential concerns about division bias, which may affect the estimates if we normalize price by area, and classical measurement error in log plot size on enters with opposite signs on either side of the estimating equation. Panel A of Table 2 shows that when we control for a broad set of fixed effects, the elasticity of plot price with respect to plot size is around 0.5, suggesting that the elasticity of plot price per square meter with respect to plot size is around -0.5. We find very similar results in columns (1) and (2) of Panel B of the same table, where we use as the outcome the logarithm of plot price per square meter. Columns (3) and (4) use the official size cutoffs for plot sizes, and show that compared to small plots, the price per square meter is lower on middle-sized plots and lower still on larger plots. Figure 7 shows the relationship between the logarithm of price per square meter and plot size non-parametrically, using 100-meter bins. The downward-sloping relationship is evident throughout the plot size distribution (at least where there is a sufficient number of plots).

To shed more light on the causal effect of log plot size on log plot price per square meter, we now turn to RD specifications, to control for possible unobserved consistent differences in neighborhood features among plots of different sizes. In Table 3 we show results using a 100-meter bandwidth and estimating the regression separately for discontinuities with different gaps in the plot size between adjacent small and large plots. Where the gap in sizes across the discontinuity is large (over 400 sq m) as in panel A, column 3 there is a large discount (0.37 log points) per sq meter for the larger lots. Where the gap in neighboring insulae plot sizes is small (panel B), the estimates are small and imprecise. In Appendix Table A.1, we report estimates of a specification where own larger is interacted with the log size of the gap between neighboring insulae plots. With the full set of controls, the price difference increases strongly as the gap size increases. In summary, it appears grid squares are strongly misallocated between larger and smaller plots and that there would be considerable gains from increasing the supply of small plots.[17]

---

[17]To make this concrete, consider a hypothetical decision by the initial planners to split a marginal 1600 sqm plot into four plots of 400 sqm. The mean value of a 1600 sqm plot in 2021 was about 39 million TZS, compared to about 14.8 million TZS for each 400 sqm plot. A conservative estimate of the cost of splitting one plot into four at planning stage is about 0.87 million TZS. This leaves a gain of about 19.3 million TZS, or almost 50% of the mean

Plot size may affect not only land values, but also land use. From the data and model we are interested primarily in three outcomes: share of the grid square that is built upon, which factors in two considerations: whether the grid square is in a plot that is built upon (has a building larger than 30 sq m) or not by 2021, or timing (τ); and if built upon, what is the size of footprint (where we add up the sizes of the three largest building footprints) in the plot the grid square is in, which captures the level of investment (k). We also consider two other outcomes, the footprint size of the largest building of the plot of the grid square and whether there are multiple buildings (over 30 sq m). In Panel B of Table 4 where we control for amenities, there is no significant effect of plot size on whether a plot is yet built upon, a result that the theory suggested. Some larger plots are built early and some later, compared to small ones. However, the share built declines significantly with plot size reflecting the increased ratio of green space to footprint as plot size rises, a feature of other work, perhaps reflecting preferences for greater relative green space by the higher income consumers of larger plots. The footprint size if built upon increases significantly with plot size, as do size of the largest building, and whether there are multiple buildings or not all. While the theory suggested that holding income constant, investment would decline as plot size increases, we cannot control for income. This table suggests that the unobserved increased in income that goes on average with increased plot size dominates investment decisions in the model and in the data.

Figure 8 shows non-parametrically that the share of gridcell built declines in plot size (so larger plots have more outdoor space). The footprint size, if built upon, rises with plot size (or investment goes up with income); and the size of the largest building and whether the plot has multiple buildings also rises with plot size. When we get to very large plots estimates can be imprecise given the limited sample size for some bins. In terms of multiple buildings the positive association with plots size could be evidence of backyarding in bigger plots (Brueckner et al., 2019), or just that bigger plots have more outbuildings. Whether a plot is built upon or not, which has a zero coefficient for the log of plot size in OLS, in the non-parametrics is a little more nuanced. Middle size plots compared to small have some bins with lower probabilities of being built upon.

To further examine this relationship, we estimate the RD specification (3) using the same five outcomes in Table 5, to control for unobserved neighborhood features. Panel A restricts the analysis to discontinuities where the gap in plot sizes exceeds 400 sqm. While it shows that the share of a gridcell that is built declines in plot size and the share of plots with multiple buildings rises with plot size, the footprint of development (k) and the size

value of a 1,600 sqm plot. Even if splitting entails allocating more land to roads, the hypothetical gain in land value is still substantial. We note, however, based on our conversation with a former director of the 20,000-plot project, that nowadays splitting plots is difficult, due to legal and procedural barriers.

of the largest building now no longer increase with plot sizes. In fact the coefficient on the footprint variable is negative, which theory suggests is the case if we hold income constant while changing plot size. This suggests that the RD helps control for income: even though plot sizes vary across neighboring insula, adjacent insula are likely to have similar income people. In Panel B, with the small gap in plot sizes, as expected, all the effects of plot size diminish. For a final look, Appendix Table A.2 shows interacted regressions, where we allow the RD coefficient to vary with the logarithm of the size gap. TO FOLLOW

Finally for plots size results, we consider the implications of plot size for population density. Appendix Table A.3 shows, using our questionnaire data, that mean population per built residential plot barely increases (from 5.3 to 5.6) as we move from small plots to large ones. To assess the implications for overall population density, we consider the share of plots of each size that are built and assume that a roughly equal share (one half) of the total area is taken up by residential plots, as we observe in our data. The resulting projects suggest that the overall population density per sq km varies from around 2,100 for small plots to around 700 for large plots, compared to a mean of around 4,000 per square km for Dar es Salaam as a whole (MLHHSD, 2018). The 20k project areas are thus relatively sparsely populated compared to the city as a whole, owing in part to their peripheral locations and the fact that only half the plots are built. But the small plots are roughly three times more densely populated. This could have implications for neighborhood agglomeration effects.

5.2.3 Spillovers: Does the size of neighboring plots matter?

We now examine whether the (non)homogeneity of sizes of neighboring plots affects an owner's plot. For this we define "super-insula" areas, which we divide into 3 types of neighborhoods: those that have fairly uniformly small, medium and large plot sizes. Recall that within insulae plots are of similar size. Super-insulae are sets of contiguous insulae of the same size category (small, medium, and large), with the derivation discussed in Appendix section C.2.[18]. This yields a very nice RD formulation, looking across borders of super-insulae that differ by definition in composition. At a border between super-insulae of different types, people experience a mixed neighborhood such as 50%-50% small and large plots. As they move interior from the border they experience increasing homogeneity of their own type.

For price comparisons we had too few observations where plots were in super-insulae border areas. For quantities, controlling for own plot size, we want to know whether neighbor plot sizes are an amenity, based on theory that says better amenity plots are developed sooner and, if income is the same, may have lower investment (due to higher cost of the land). Here unlike moving across many insula borders, we might expect incomes to vary more across

---

[18] "Super insulae" is a concept that we – rather than the projects' planners – define.

super-insula borders, although the covariate, own larger may capture part of that effect. The key is to look at how outcomes vary as owners move away from mixed neighbors at the border into more homogeneous smaller or larger plots neighborhoods. We start with a simple formulation in Table 6, where we interact own smaller with distance to the border and own larger with distance to the border. We restrict the sample to borders that are no more than 30 meters apart, losing about 2% of the sample. As owners move into neighborhoods with smaller houses and homogeneity increases, we see a greater likelihood of a plot being built upon, as well as the share built rising, suggesting that for smaller plot owners, having more small plot neighbors is valued. Footprint size declines with the improved amenity (the case in the theory holding income constant) but the effect is insignificant; and the effect on the size of the largest building is positive. However as we move away from the border on the larger plot side, effects are all insignificant. So perhaps surprisingly, it is small plot owners who value uniformity of the neighborhood, not large plot owners.

We explored this result in two ways. First we tried OLS, using more general within-area variation. We calculate the share of each gridcell's neighboring gridcells within 100 meters, which are on residential plots (excluding the own plot's gridcells). This allows us to calculate the share of each gridcell's neighbors that are small, meaning neighboring plots with plots less than 800 sq meters. We then estimate specification (2), adding as covariates the share small and the share small interacted with own plot size. Given the formulation we can also have price regressions as well as quantity ones. The results are reported in Appendix Table A.4. Results are generally insignificant, but there is a pattern where the coefficient of share neighbors small is positive and the coefficient of share small interacted with plot size is negative. For the share built outcome where coefficients are significant, at a plot size of 400 sq m, the marginal effect of a 0.1 point increase in share small on share built is (0.1)*(0.28 - .037 * 6) = 0.006. This would potentially suggest, small plots value being surrounded by small neighbor plots. This positive effect declines to 0 at plot size of about 1900 sq m; and after that the suggestion is that share small is bad for large plots. A similar pattern exists for plot built or not but results are only significant at the 10% level. We think the results are somewhat weak because the super-insulae RD suggests large plot owners do not care about there being a higher share of small plots nearby.

We then detailed the effects using the super-insula RD by distinguishing between the three types at the border: small-medium, small-large and medium-large, not just smaller versus larger. TO FOLLOW

5.2.4 Valuation of amenities

We next turn to the valuation of amenities more generally, examining the value of both planned amenities, arising from road and non-residential plots, and natural ("first nature") amenities. Starting with natural amenities, Table 7 shows that elevation is highly prized. Conditional on the full set of covariates, a one-meter increase in elevation is associated with an increase of almost one percent in the value of land, as well as increases in the share of the gridcell that is built, the likelihood a plot is built upon, and other quantity outcomes. At the same time, ruggedness and proximity to rivers or streams - and water or wetlands - appear to be disamenities, as reflected in the various measures of housing construction. This evidence is consistent with residents seeking to mitigate the significant risk of flooding in Dar es Salaam (Jaupart et al., 2017) by preferring higher ground that is less likely to flood, while avoiding rugged terrain that is costly to build on. Note that in the model, we argued that in equilibrium it seemed likely that disamenities would delay construction time, but here that is not the case. However greater ruggedness, river/stream, and wetlands may raise the price of construction reducing investment , where there is a cost to leveling land, building on slopes, draining wet areas, or building to avoid indoor flooding (see, Henderson, Regan and Venables (2021)).

Next are two consumption amenities: distance to a pre-existing major paved road and a z-index of three characteristics. A shorter distance to pre-existing roads leads to a a greater share built, earlier development, and greater investment in all dimensions, with some notion that holding plot size fixed higher income people will bid for access. A kilometer closer to a main thoroughfare increases the likelihood of being developed by 0.04 (mean 0.49). The underlying characteristics of the z-index tell us the degree to which insula are uniform in plot size, rectangular and laid out on a grid fashion. Uniformity of plot size within the insula is measured as (1- coefficient of variation of insula plot sizes) with a maximum value of 1. Rectangularity is the area of the insula divided by the area of the minimum bounding rectangle of the insula, with a maximum value of 1 when an insula is perfectly rectangular. Finally, the degree to which the road structure might be more grid-like is measured by the degree of alignment of the insula vis-a-vis the nearest neighboring insula. For the last, we define the minimum bounding rectangle of each insula, then for each define the closest in terms of edges, and then look at the angle of the edges to each other. If the edges are parallel, the difference in angle is 0 and $\tan(0)=0$ and the index is given the value 1- $\tan(0)$, the maximum then of 1. This implies the insula are parallel to each other as in a gird-road structure. Then to measure deviations from this, think of holding one insula fixed, and then start rotating the other clockwise. As the angle between them increases, misalignment increases up to 45 degrees, where $\tan(45) = 1$ and the index is then 0 at the maximum

misalignment. Further rotation brings another side of the rotating insula nearer alignment with the fixed insula. We add these three measures together to form a z-index. An increase in the z-index leads to a greater share built and an earlier time of development, but as an amenity an insignificant effect on overall investment and likelihood of multiple buildings. When we break the z-index out, the critical components as seen in Appendix Table A.5 are rectangularity and grid-like alignment, rather than uniformity of plot size.

Turning to planned amenities in terms of non-residential land use, proximity to planned open spaces, educational facilities, or other listed planned amenities in Table 7, have only two significant coefficients out of 45 suggesting that the presence of an amenity planned on paper was not valued. However, this does not mean that owners do not value these amenities; but, rather, it seems that the planned amenities were usually not delivered.

This is shown in Figure 9, which reports two things. TO BE FINALIZED One is the implementation rate for nine different planned amenities into their intended use (requiring at least 100 planned sites in the price data for any type to appear in the figure). Only cemetery use approaches 50%, religious sites about 40%, and educational uses about 30%. Other uses are well under 10%. However for any planned use, apart from its intended use, it could be put into a different use, with in many cases high rates of other use (like planned recreation sites). And in all cases there is a significant and often large fraction of sites left unused. The other part of the figure shows for the 8 defined categories the share planned for each use and the realized share of that type over all plots. The figure reports that 40% of planned non-residential plots lie fallow, and that is pretty evenly split between being kept-up by the community (but not used for recreation) versus being left unkept. Finally, the figure reports the share of plots that went into farming and residential use. Farming is about 15% of all non-residential plots and residential 10%. Thus rather than lying fallow like 40% of the non-residential planned plots about 25% went into some non-public use. While maybe that is good to have the plots in a productive use, it is not a triumph for planning. All this suggests that planners were overly optimistic when proscribing non-residential uses, which have yet to materialize about two decades after the project began and more than one decade after the government finished selling the land to private owners.

Finally we examine correlations associated with implemented uses in Table 8. There we do not report price data regressions due to lack of actual implemented uses near many of the transacted sites. We report on 14 types of actual: now in 18 of the 70 cases there is a significant correlation with quantities. What is most apparent is that unkept has clear negative associations. That could be because either unkept is a disamenity; less developed areas are less able to raise volunteer labor to do the upkeep; and/or the local leader focuses organizational efforts on more densely developed areas. In Appendix Table A.6 we show that

being near a site that is implemented as planned is significantly associated with a greater likelihood of being developed by 2021, again with the chicken and egg problem.

5.2.5 Sorting of owners and other residents

We conclude our discussion of the results with evidence on the sorting of residents by education, which we calculate using the residents' questionnaire. We focus on years of schooling, since this is widely used as a proxy for lifetime earnings (OECD 2022), and one that most residents are happy to share - more so than current earnings. Panel A of Figure 10 shows two aspects of the sorting of plot owners, which reside in around half of the plots we surveyed. First, the mean education (around 13-14 years) of these owner-occupiers is 4 to 5 years higher than that of the mean heads of household in Dar es Salaam as a whole.[19] Second, owners of plots in 20,000 plot project areas that were initially more expensive (based on prices set by the government) are more educated. Panel B shows evidence on non-owners, around one-third of which are renters and the other two-thirds usufructuaries (typically relatives or friends of owners who do not pay any rent). Non-owners are quite similar in their schooling to the mean in Dar es Salaam, and those living in initially more expensive 20k project areas are not differentially more educated.

The strong sorting pattern on owners by education is also reflected within 20k plot areas. As shown in Appendix Table A.7, owners of larger plots are more educated, as are owners of plots whose land is more expensive and owners of plots which are currently more valuable (including housing), as reported by the owners. The sorting pattern for non-owners within project areas is again more mixed.

6 Concluding Remarks

For millennia, urban planners have shaped cities, seeking solutions for practical problems through a balance of market-based approaches, which emphasize individuals' preferences and knowledge, with centralized ones, which seek to define the rules of the game (e.g., property rights), provide public goods, and mitigate externalities and inequalities. Finding appropriate balance is particularly challenging in developing country cities, which are growing very rapidly, underscoring tensions between rigid formal planning standards and a reality dominated by unplanned informal neighborhoods with low-quality housing. In such settings, a key policy option is 'de novo' urban planning, whereby greenfield agricultural land on urban fringes is purchased and partitioned into formal surveyed and titled plots with roads, which

---

[19]In 2014, the average years of schooling for heads of households in for Dar es Salaam was around 8.7 years (World Bank, 2019).

people can buy and build their homes on. In this paper we study the consequences of concrete decisions made by de novo planners – what sizes of residential plots to select, how to lay them out, and how to allocate land to roads and other public spaces, as well as commercial ones. We study the implications of these choices in the context of the '20,000 Plot' project, implemented by the Tanzanian government in the first decade of this millennium.

The 20k project delivered around 36,000 residential de novo plots in 12 project areas on or near the fringes of Dar es Salaam, taking up a total of 75 square km. We find that the project cost of just under $1 (USD 2021) per square meter of residential plot were rapidly recouped by the purchasers' payments. Bare land values then increased sharply in real terms and are now double those in nearby informal areas.

Despite the overall successes of the project, however, gains in land values have been uneven both across and within project areas, and currently only about half of the plots are built on, reflecting potentially inefficient use of a scare resource - planned land.

To shed light on the consequences of the project's design, we develop new data from circa 2020, combining questionnaires and information from satellite images. Using variation in plot sizes within project areas and spatial regression discontinuity design, we find that the elasticity of price per sqm with respect to plot size is approximately -0.5, so land in larger plots is much less valuable. We also find that larger plots have much more open space and only slightly more built area, attract better educated owners, and their overall population densities are much lower. These results are consistent with an over-provision of large plots, probably due to implementation of colonial-era planning rules, which have largely persisted until the present day.

We also find evidence that plots are more likely to be built on - and are built on more intensively - when they are located next to plots of similar size, suggesting a role for the planned layout of the neighborhoods and the potential segmentation of areas into blocks of homogeneous plot sizes.

We find that land values and construction patterns reflect valuation of natural amenities (elevation) and disamenities (ruggedness as well as proximity to rivers, streams, and water and wetlands, in a city prone to flooding). And while proximity to large roads is valued, most planned amenities are not. The likely explanation for this is that most planned amenities, other than roads, have not yet been provided, more than a decade after the project was completed and its plots were sold.

Taken together, our findings suggest that while the 20k project has been a success, better planning could result in larger and more equitably distributed gains from urban planning.

References

Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. M. (2023), 'When should you adjust standard errors for clustering?', **138**(1), 1–35.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001), 'The colonial origins of comparative development: An empirical investigation', **91**(5), 1369–1401.

Ahmad, A. M., Sultan, A. E.-I. M., Abdalla, M. M., Malik, A. O., Said, M. A., Mukhtar, E.-S. M. and El-Mufti, N. M. (2002), 'Low-cost housing projects in khartoum with special focus on housing patterns', **26**(2), 139–157.

American Planning Association (1950), Urban Land Use, Information Report No. 14, American Society of Planning Officials, Chicago, Illinois.

Angel, S. (2012), *Planet of Cities*, Lincoln Institute of Land Policy.

Asabere, P. K. (1990), 'The value of a neighborhood street with reference to the cul-de-sac', **3**(2), 185–193.

Asabere, P. K. and Colwell, P. F. (1984), 'Zoning and the Value of Urban Land', *Real Estate Issues* **9(1)**, 22–27.

Baruah, N., Henderson, V. and Peng, C. (2021), 'Colonial legacies: Shaping african cities', *Journal of Economic Geography* **21**(1), 29–65.

Bayer, P., Ferreira, F. and McMillan, R. (2007), 'A unified framework for measuring preferences for schools and neighborhoods', **115**(4), 588–638.

Bertaud, A. (2018), *Order without Design: How Markets Shape Cities*, MIT Press.

Besley, T. and Persson, T. (2014), 'Why do developing countries tax so little?', **28**(4), 99–120.

Bolton, L. (2020), Sites and Services, and In-Situ Slum Upgrading, Technical report, Institute of Development Studies, Brighton, UK.

Brownstone, D. and De Vany, A. (1991), 'Zoning, returns to scale, and the value of undeveloped land', **73**(4), 699–704.

Brueckner, J. K., Rabe, C. and Selod, H. (2019), 'Backyarding: Theory and evidence for South Africa', *Regional Science and Urban Economics* **79**, 103486.

Buckley, R. M. and Kalarickal, J. (2006), *Thirty Years of World Bank Shelter Lending : What Have We Learned?*, Washington, DC: World Bank.

Castells-Quintana, D. (2017), 'Malthus living in a slum: Urban concentration, infrastructure and economic growth', **98**, 158–173.

Chiumenti, N., Kulka, A. and Sood, A. (2022), 'How to Increase Housing Affordability: Understanding Local Deterrents to Building Multifamily Housing'.

Choi, N., Berrisford, S. and Alexander, C. (2020), Housing solutions for low-income urban communities in rwanda : Main report, Technical report, World Bank Group.

Collier, P. and Venables, A. J. (2014), Housing and urbanization in africa : Unleashing a formal market process, Policy Research Working Paper WPS6871, World Bank Group.

Colwell, P. F. and Sirmans, C. F. (1993), 'A comment on zoning, returns to scale, and the value of undeveloped land', **75**(4), 783–786.

Combes, P.-P., Duranton, G. and Gobillon, L. (2021), 'The production function for housing: Evidence from france', **129**(10), 2766–2816.

Davis, O. A. and Whinston, A. (1962), 'Externalities, welfare, and the theory of games', **70**(3), 241–262.

Davis, O. A. and Whinston, A. (1964), 'The economics of complex systems: The case of municipal zoning', **17**(3), 419–446.

De Soto, H. (1989), *The Other Path: The Invisible Revolution in the Third World*, Harper & Row.

Dell, M. (2010), 'The persistent effects of peru's mining "Mita"', **78**(6), 1863–1903.

Diamond, R. (2016), 'The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000', **106**(3), 479–524.

Dixit, A. (1973), 'The optimum factory town', **4**(2), 637–651.

Duranton, G. (2017), 'The death and life of great american cities / the economy of cities', **51**(12), 1871–1873.

Epple, D. and Sieg, H. (1999), 'Estimating equilibrium models of local jurisdictions', **107**(4), 645–681.

Fujita, M., Krugman, P. and Venables, A. J. (1999), *The Spatial Economy: Cities, Regions, and International Trade*.

Fuller, B. and Romer, P. (2014), 'Urbanization as Opportunity', (2439696).

Gibbons, S., Machin, S. and Silva, O. (2013), 'Valuing school quality using boundary discontinuities', **75**(C), 15–28.

Glaeser, E. (2012), *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier and Happier*, Penguin.

Glaeser, E. L., Gyourko, J. and Saks, R. E. (2005), 'Why Have Housing Prices Gone Up?', *American Economic Review* **95**(2), 329–333.

Gottlieb, J. D. (2018), How minimum zoning mandates can improve housing markets and expand opportunity, Technical report, The Aspen Institute Economic Strategy Group.

Grimes, O. F. (1976), Housing for Low-Income Urban Families, World Bank Country Economic Reports, The Johns Hopkins University Press, Washington, D.C.

Gulyani, S. and Connors, G. (2002), Urban upgrading in africa: A summary of rapid assessments in ten countries.

Hall, P. and Tewdwr-Jones, M. (2019), *Urban and Regional Planning*, 6th edition edn, Routledge.

Harari, M. and Wong, M. (2017), Long-term impacts of slum upgrading: Evidence from the kampung improvement program in indonesia.

Henderson, J. V., Nigmatulina, D. and Kriticos, S. (2021), 'Measuring urban economic density', **125**, 103188.

Henderson, J. V., Regan, T. and Venables, A. J. (2021), 'Building the city: From slums to a modern metropolis', **88**(3), 1157–1192.

Howard, E. (1902), *Garden Cities of To-Morrow*, S. Sonnenschein & Co., Ltd., London.

Jacobs, J. (1962), *The Death and Life of Great American Cities*, Random House.

Jamal, K. (2018), 'Minister disbands dar city lands unit'.

Jaupart, P., Chen, Y. and Picarelli, N. (2017), Cholera in times of floods weather shocks and health in dar es salaam.

Kironde, J. M. L. (1994), 'The evolution of the land use structure of dar es salaam 1890-1990: A study in the effects of land policy'.

Kironde, J. M. L. (2006), 'The regulatory framework, unplanned development and urban poverty: Findings from dar es salaam, tanzania', **23**(4), 460–472.

Kironde, J. M. L. (2015), 'Good governance, efficiency and the provision of planned land for orderly development in african cities: The case of the 20,000 planned land plots project in dar es salaam, tanzania', **3**(4), 348–367.

Kolbe, J., Schulz, R., Wersing, M. and Werwatz, A. (2021), 'Real estate listings and their usefulness for hedonic regressions', *Empirical Economics* **61**(6), 3239–3269.

Konadu-Agyemang, K. (2001), *The Political Economy of Housing and Urban Development in Africa: Ghana's Experience from Colonial Times to 1998*, Bloomsbury Academic.

Kulka, A. (2019), 'Sorting into neighborhoods: The role of minimum lot sizes'.

Lamson-Hall, P., Angel, S., DeGroot, D., Martin, R. and Tafesse, T. (2019), 'A new plan for african cities: The ethiopia urban expansion initiative', **56**(6), 1234–1249.

Lawler, A. (2008), 'Indus collapse: The end or the beginning of an asian culture?', **320**(5881), 1281–1283.

Libecap, G. D. and Lueck, D. (2011), 'The demarcation of land and the role of coordinating property institutions', **119**(3), 426–467.

Mayo, S. K. and Gross, D. J. (1987), 'Sites and services—and subsidies: The economics of low-cost housing in developing countries', **1**(2), 301–335.

Michaels, G., Nigmatulina, D., Rauch, F., Regan, T., Baruah, N. and Dahlstrand, A. (2021), 'Planning ahead for better neighborhoods: Long-run evidence from tanzania', **129**(7), 2112–2156.

MLHHSD (2018), Dar es Salaam City Master Plan 2016-2036, Technical report, Tanzania Ministry of Lands, Housing and Human Settlements Development, Tanzania.

Moretti, E. (2012), *The New Geography of Jobs*, Houghton Mifflin Harcourt.

Msangi, D. E. (2011), 'Land acquisition for urban expansion'.

Mwiga, B. G. (2011), Evaluating the Effectiveness of the Regulatory Framework in Providing Planned Land in Urban Areas: the case of Dar es Salaam city 20,000 plots project, Tanzania, PhD thesis, University of Twente, Enschede, Netherlands.

OpenStreetMap contributors (2017), 'Planet dump retrieved from https://planet.osm.org on March 30, 2022', https://www.openstreetmap.org.

Paden, R. (2001), 'The two professions of hippodamus of miletus', **4**(1), 25–48.

Scruggs, G. (2015), 'How Much Public Space Does a City Need?', *Next City* .

Shertzer, A., Twinam, T. and Walsh, R. P. (2018), 'Zoning and the economic geography of cities', **105**, 20–39.

Smith, A. (1759), *The Theory of Moral Sentiments*, A. Millar, A. Kincaid and J. Bell in Edinburgh.

Smith, M. E. (2007), 'Form and meaning in the earliest cities: A new approach to ancient urban planning', **6**(1), 3–47.

Solow, R. M. and Vickrey, W. S. (1971), 'Land use in a long narrow city', **3**(4), 430–447.

Song, J. (2021), 'The effects of residential zoning in U.S. housing markets'.

Storey, G. R. (2004), 'The meaning of "Insula" in roman residential terminology', **49**, 47–84.

Symonds, Q. (2023), 'QS Ranking of Urban Planning Programs'.

Tiba, A., Mwarabu, G., Sikamkono, W., Kenekeza, H., Sarehe, S., Mwakalinga, V. and Fyito, V. (2005), 'The implication of 20,000 plots project on the emerging form of dar es salaam city'.

Tipple, G. (2015), 'Housing policy-making in africa: Ten common assumptions', **49**, 413–418.

Turner, M. A., Haughwout, A. and van der Klaauw, W. (2014), 'Land use regulation and welfare', **82**(4), 1341–1403.

UN-Habitat (2013), Streets as Public Spaces and Drivers of Urban Prosperity, Technical report, UN-Habitat, Nairobi.

United Republic of Tanzania (2021), The first disbursement of fund for implementation of land use planning, surveying and titling program in the country. kilosadc.go.tz.

United States Geological Survey (2000), 'SRTM1 Arc-Second Global'. DOI: /10.5066/F7PR7TFT.

World Bank (1974), Appraisal of national sites and services project, Technical Report 337a-TA.

World Bank (1977), Tanzania: The second national sites and services project, Technical Report 1518a-TA.

World Bank (1984), Completion report: Tanzania - first national sites and services project, Technical Report 4941.

World Bank (1987), Tanzania: Second national sites and services project, Technical Report 6828.

World Bank (2019), Measuring living standards within cities, dar es salaam 2014-2015, Technical report, World Bank, Development Data Group.

Yamasaki, J., Nakajima, K. and Teshima, K. (2023), 'From samurai to skyscrapers : How transactions costs shape tokyo'.

Figures

Figure 1: Map of 20k project areas in Dar es Salaam

(a) Initial prices per square meter

(b) Current prices per square meter
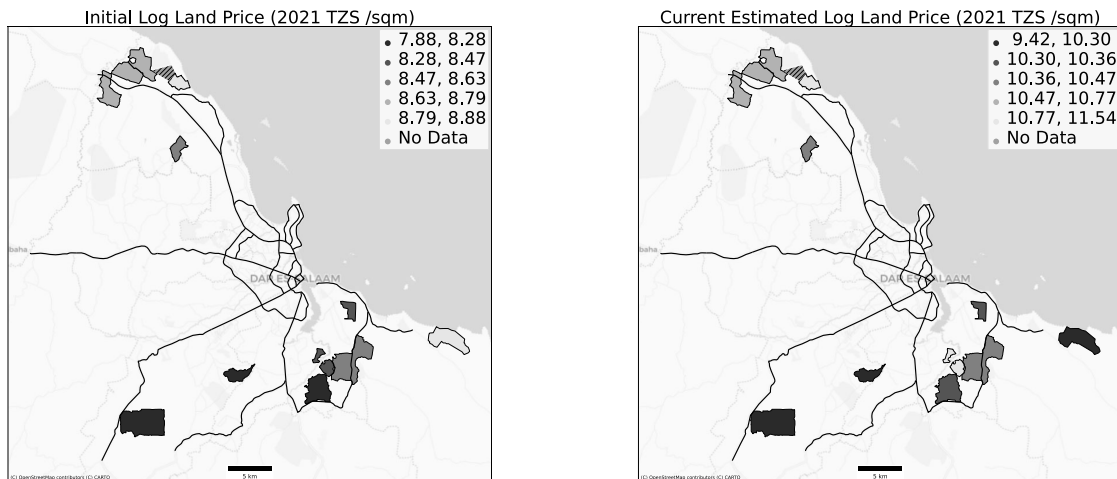
Notes: This figure maps locations of 20K areas in Dar es Salaam along with the Central Business District (CBD) with (OpenStreetMap contributors, 2017) in the background. In Panel A, each area is colored by its initial government charged price per sqm. In Panel B, each area is colored by its predicted current transaction price per sqm.

Figure 2: Example of landuses in Mbweni Mpiji

(a) Planned plot boundaries and non-residential uses

(b) Planned plot boundaries and satellite imagery
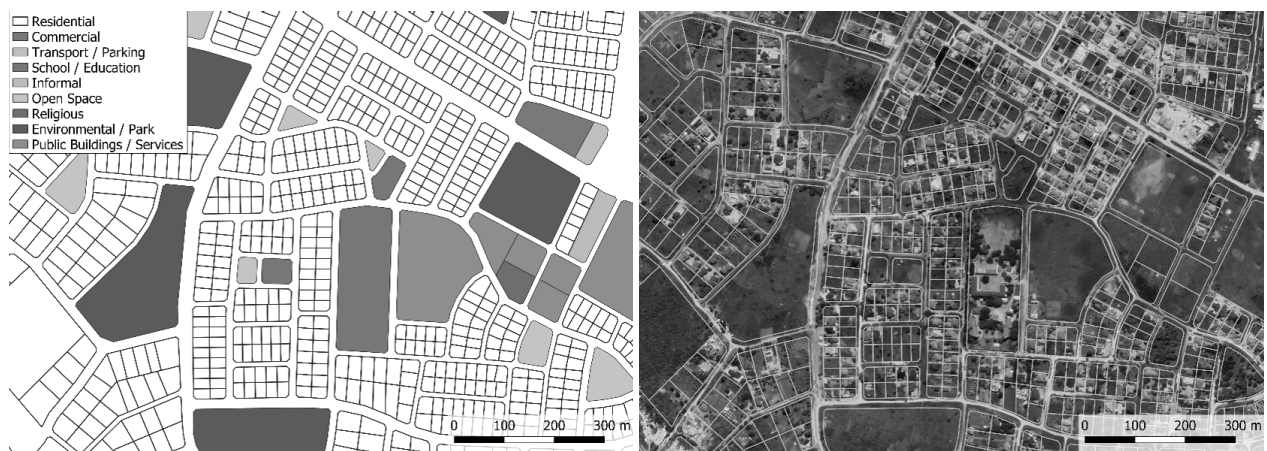
Notes: This figure plots an example of planned plot boundaries in Mbweni Mpiji. In Panel A, each plot is colored by it's planned use. In Panel B, satellite imagery is displayed in the background.

Figure 3: Example of 20k boundary in Tuangoma

(a) Satellite imagery in 2001 (pre-implementation)   (b) Satellite imagery in 2021 (post-implementation)

Notes: This figure plots an example of a 20k project boundary in Tuangoma. In Panel A, background satellite imagery is from 2001 (pre-implementation). In Panel B, background satellite imagery is from 2021 (post-implementation).

Figure 4: Histogram of residential plots sizes
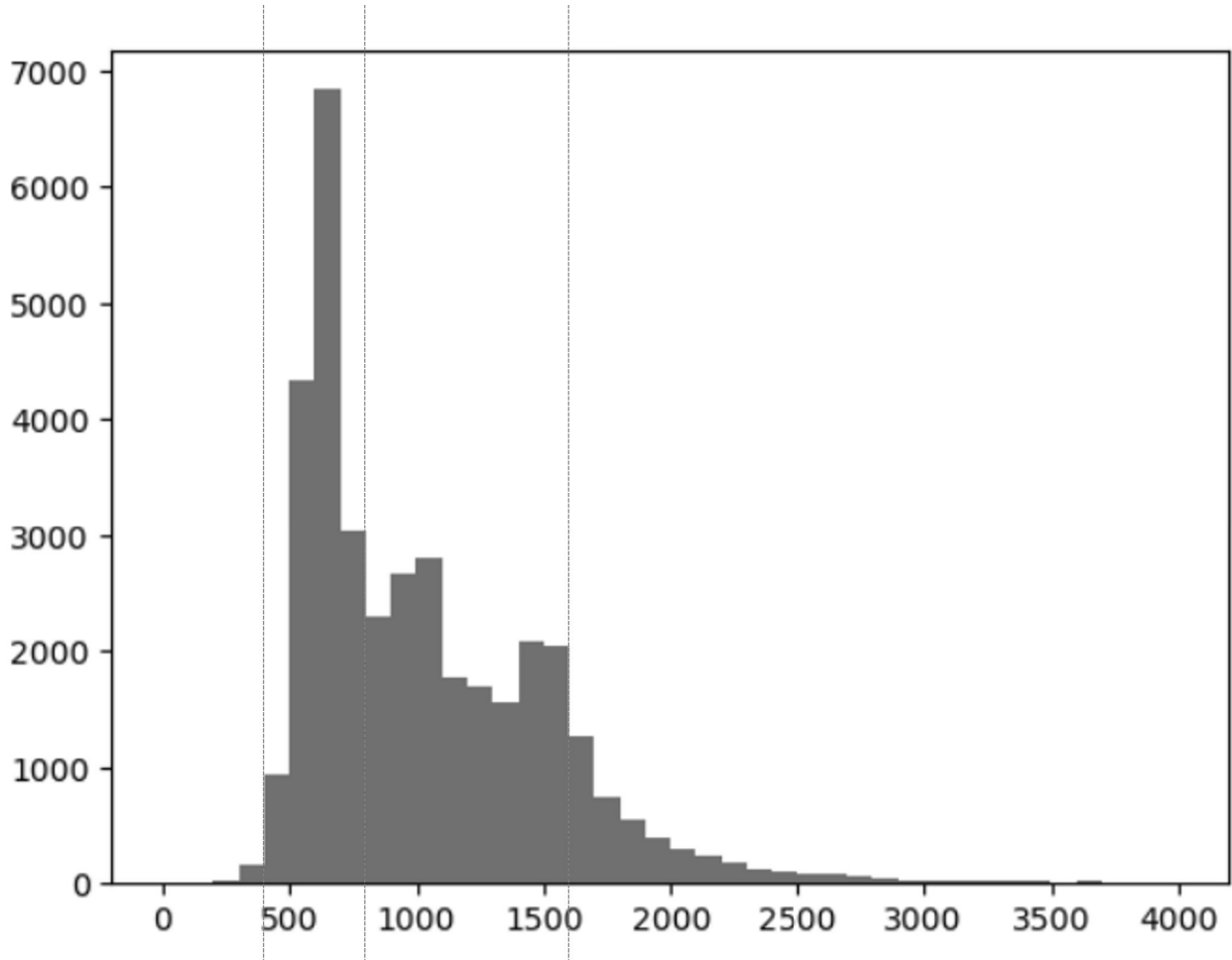
Notes: This figure plots the histogram of residential plot sizes in our sample. Vertical red lines denote cutoffs between formal size categories: Small (400-800 sqm), Medium (800-1600 sqm), and Large (1600-4000) sqm.

Figure 5: Equilibrium

(a) Plot of lifetime utilities
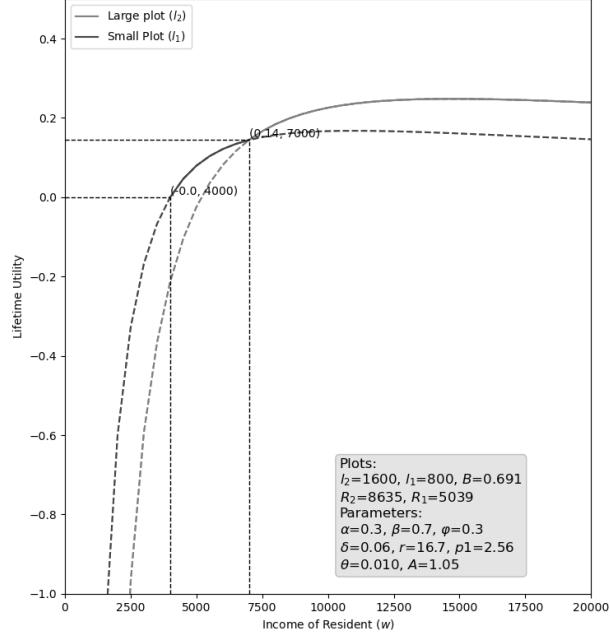
(b) Optimal k in equilibrium
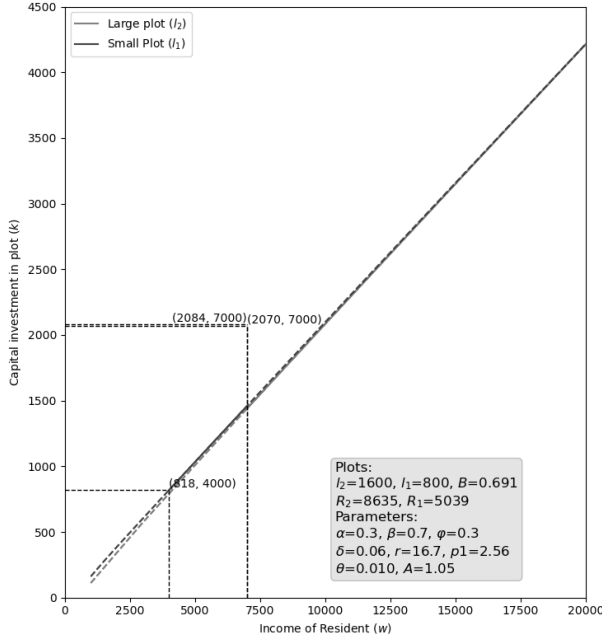
(c) Optimal τ in equilibrium
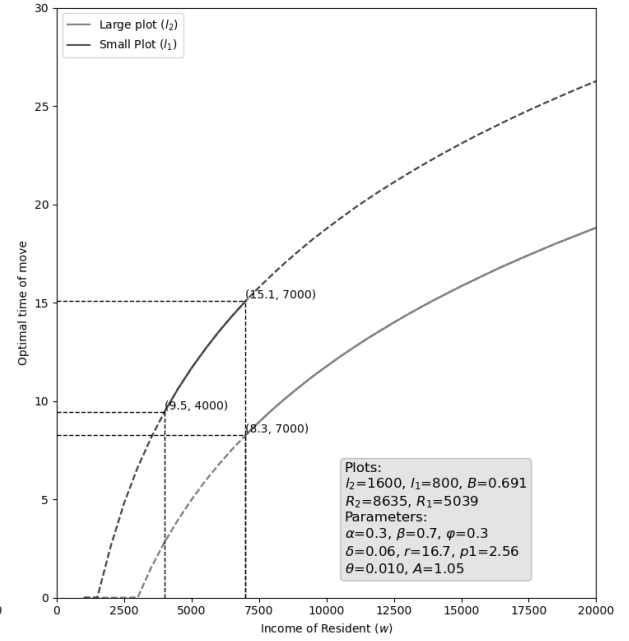
Figure 6: Equilibrium with varying amenities

(a) Optimal τ in equilibrium
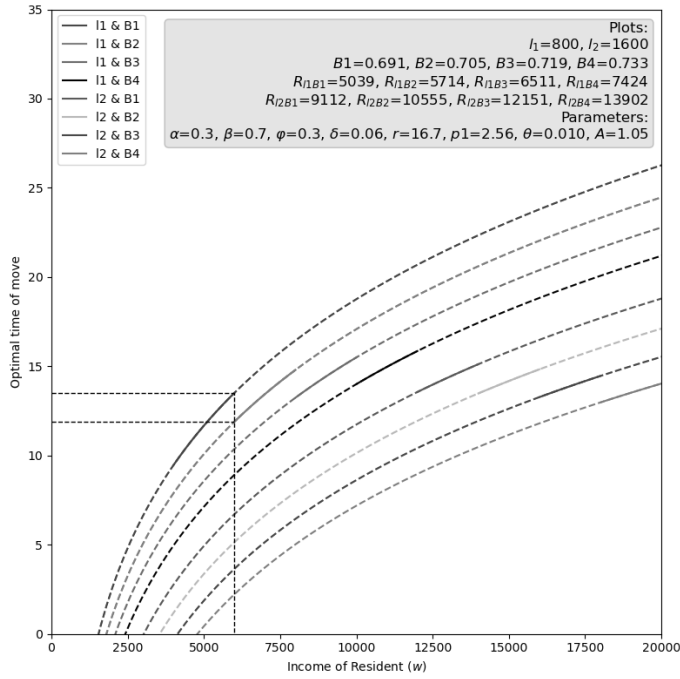
(b) Plot of lifetime utilities
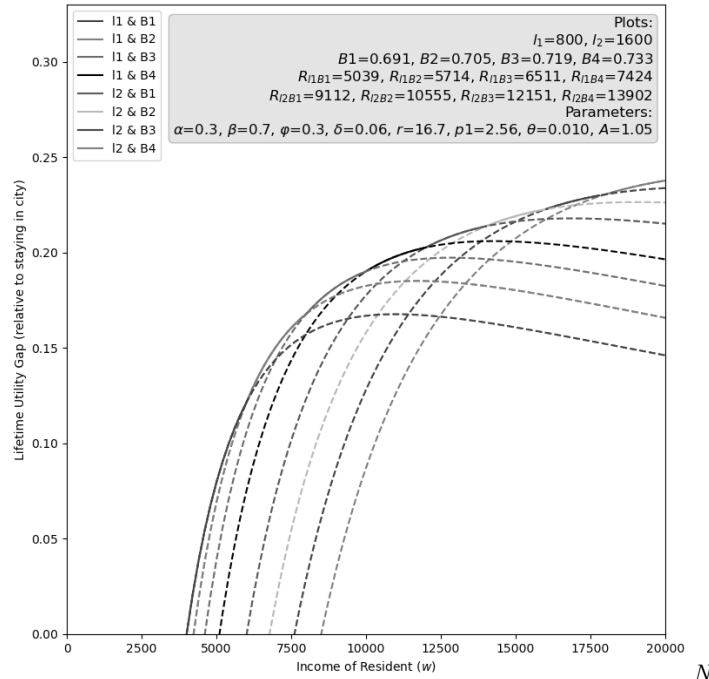
Notes:

In part (a), the different color lines show how the optimal τ varies by income for each plot size-B combination. The solid parts of the lines show the realized τ's, as they vary in equilibrium with income and changes in B's and plot sizes. Part (b) uses the same color scheme and use of solid versus dashed parts to lines. Note the solid parts form an outer-envelope of realized net utilities, which satisfies the equilibrium property that no income person could be better off choosing a different plot size B combination.

Figure 7: OLS regressions of ln plot price per square meter on ln plot size

Plot size in 100sqm bins

Mean outcome for omitted category (500-600sqm)=10.2
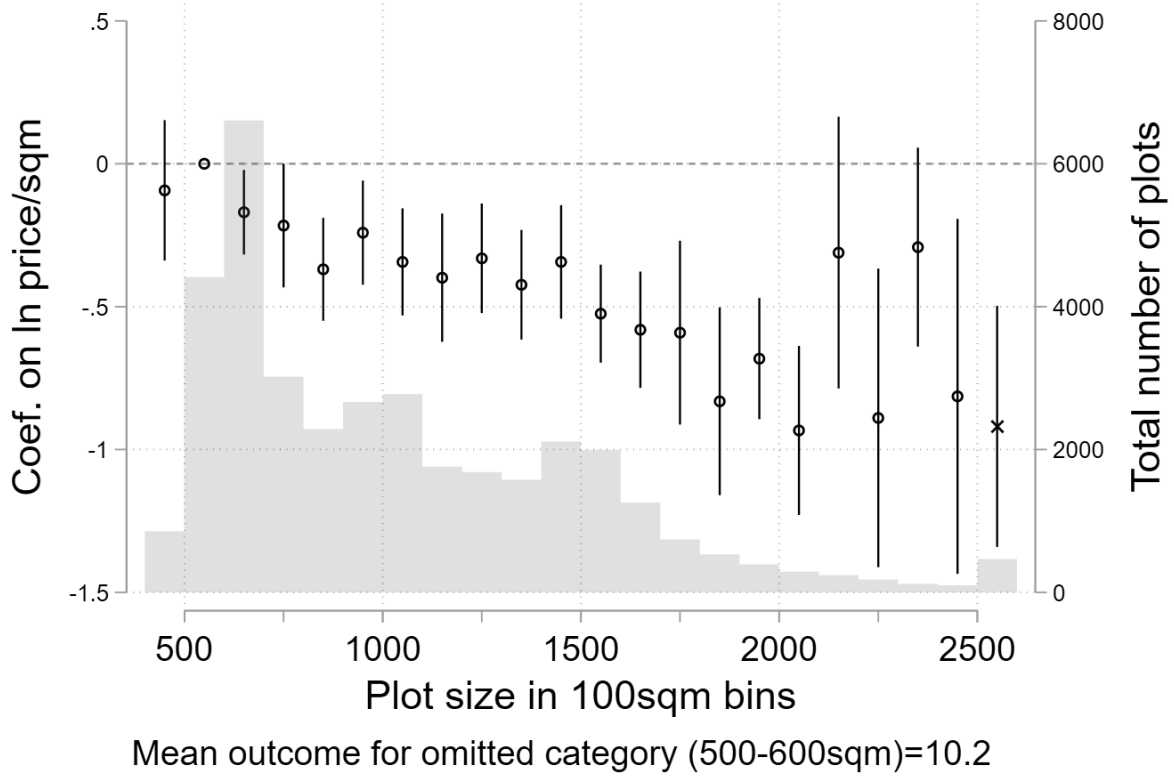
Notes: This figure plots coefficients and their 95% confidence intervals for the regression of log price per square meter on 100sqm plot size bins (plots with size above 2500sqm are pooled into one bin, marked by an 'x'). The omitted bin is 500-600sqm. Controls include transaction period by source (dalali and household questionnaires) FEs and MTAA by 20K area FEs and amenities. Observations are gridcells and standard errors are clustered by insula. Coefficients below 400sqm are not displayed, but are included in the regression.

Figure 8: OLS regressions of building quantity outcomes on ln plot size (gridcell-level)

(a) Share of gridcell built

(b) Plot has a building

(c) Size of largest building

(d) Log area of 3 largest buildings

(e) Plot has multiple buildings
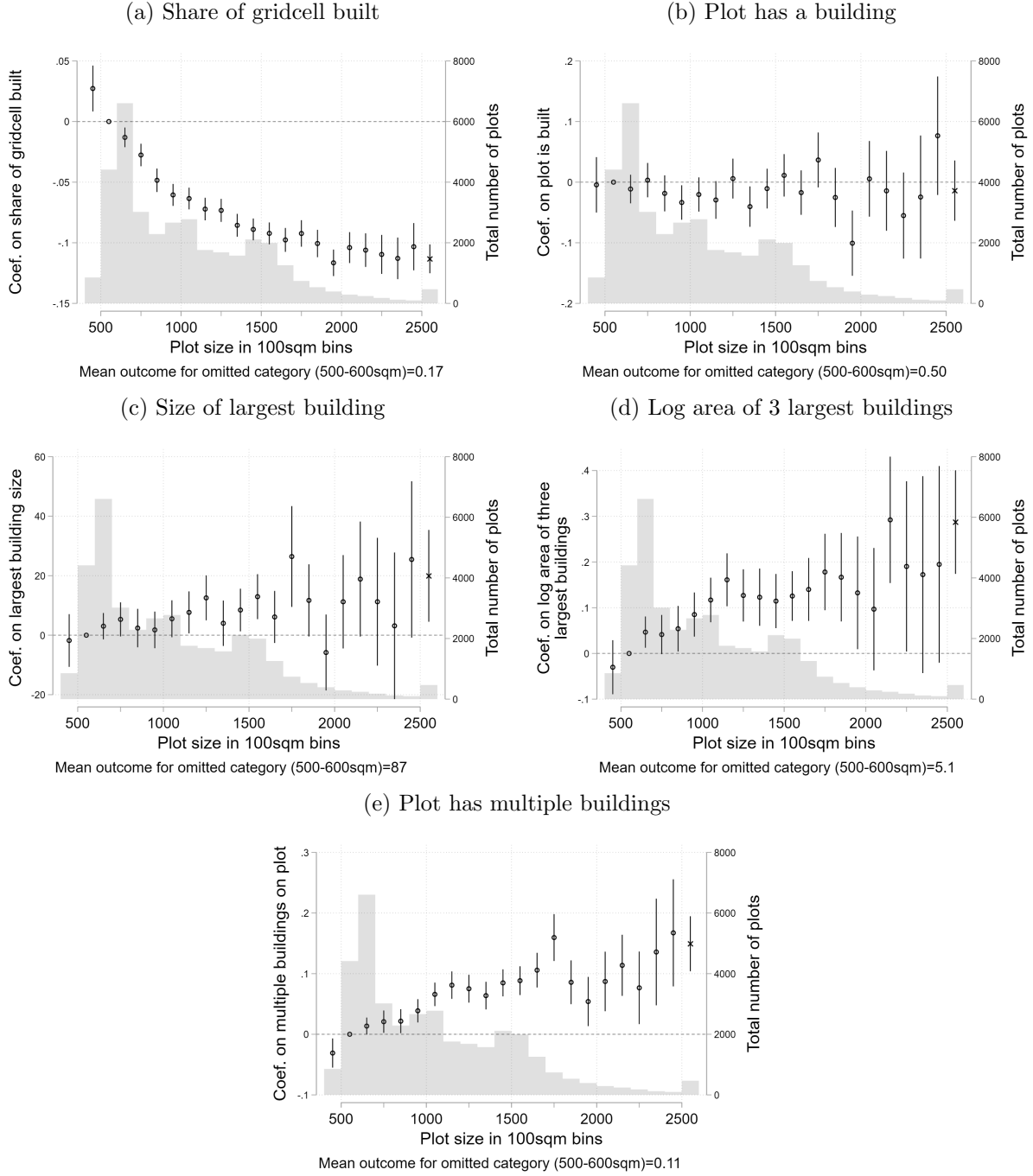
Notes: This figure plots coefficients and their 95% confidence intervals for regressions of building quantity measures on 100sqm plot size bins (plots with size above 2500sqm are pooled into one bin, marked by an 'x'). The omitted bin is 500-600sqm. Outcomes vary by panel: the share of the gridcell's area, restricted to own insula and residential plots which is built upon (a), an indicator if the gridcell's plot has a centroid of a building whose footprint area is at least 30sqm (b), the log of the sum of the areas of the three largest buildings whose centroid falls into the gridcell's own plot, condition on the largest building's footprint area being at least 30sqm (c), the size (sqm) of the largest building on the gridcell's plot where missing values are included as zeros (d), and an indicator if there are multiple buildings in the gridcell's plot (e). Controls include the standard list of amenities and MTAA by 20K area FEs. Observations are gridcells and standard errors are clustered by insula. Coefficients below 400sqm are not displayed, but are included in the regression.

41

Figure 9: Non-residential plots by planned and implemented uses

(a) Non-residential land uses

(b) Non-residential land uses
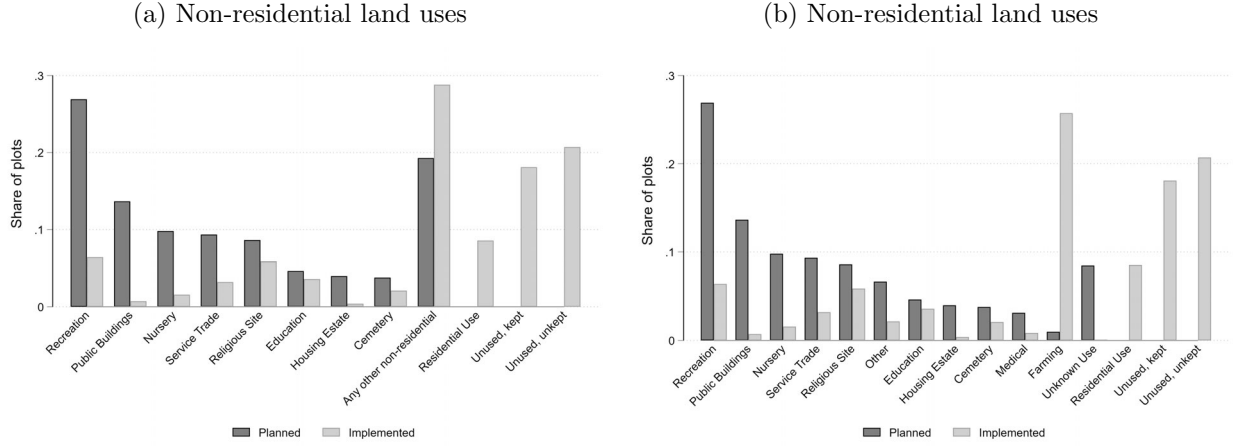
Notes: this figure plots the share of plots by planned use (dark grey) and current/implemented use (light grey).

Figure 10: Sorting by education across 20k areas

(a) Sorting of owners

(b) Sorting of non-owners
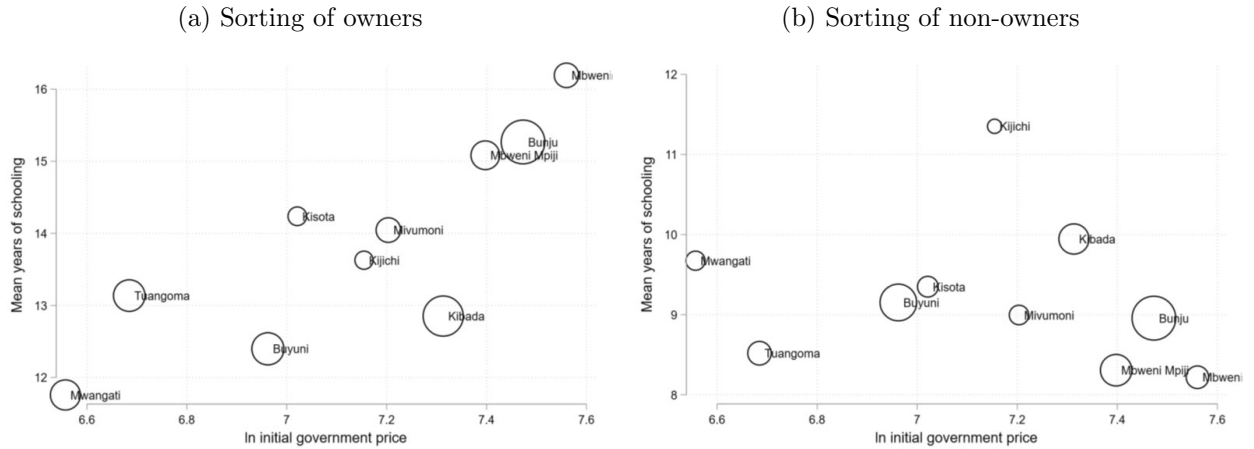
Notes: this figure plots mean years of schooling by initial price per sqm across 20k areas. In panel a, the sample is restricted to owner residents. In panel b, the sample is restricted to non-owner residents. Not that Mwongozo is missing because it was not included in the resident questionnaire, and Mbweni Malindi is missing since we have no data on initial prices there.

Tables

Table 1: Price appreciation and 20k land prices

(a) Price appreciation in 20k areas

Area	Ln Initial Price (Nominal)	Δ Ln Price (2021 TZS)
Bunju	7.47	1.98
Buyuni	6.96	1.14
Kibada	7.31	1.84
Kisota	7.02	2.02
Mbweni JKT	7.56	2.66
Mbweni Mpiji	7.40	2.05
Mivumoni	7.20	1.93
Mwangati	6.56	2.42
Mwongozo	7.56	0.99
Tuangoma	6.68	2.31
Kijichi	7.15	2.42
.	.	.
Average	7.17	1.98

Note: initial prices are based on Mwiga (2011) Table 6.4 which was sourced from the Tanzanian Ministry of Lands in 2010. In the second column, we inflate the initial prices starting from 2000. Current period (2021) prices are 20K area FE estimates + intercept from a regression of log price per square meter on period dummies (with 2021 as base), a dummy for dalali vs. resident questionnaire (with dalali questionnaire as base), and 20K area fixed effects. The price data are from bareland transaction prices from both the dalali and resident questionnaires. This table reports estimates using both to provide lower and upper bounds on appreciation by area.

(b) Ln Price inside and nearby 20k areas

	(1)	(2)
Ln plot size	0.71	0.73
	(0.04)	(0.03)
Non-20K Surveyed	-0.83	-0.53
	(0.07)	(0.07)
Non-20K Unsurveyed	-0.82	-0.85
	(0.21)	(0.19)
Mean Outcome	17	17
20K FE		✓
Municipality FE	✓	✓
Time Period FE	✓	✓
N (plots)	812	812

Note: This table presents regressions of log price on log plot size (dalali estimates) and planning/survey status. The outcome is always the log price of a bareland transaction from the Dalali survey. Each observation is a transaction. Plots outside of 20k areas are included. The sample is made of 706 plots inside 20K areas, 65 plots outside 20K areas and surveyed, and 41 outside 20K areas and unsurveyed. Controls include Municipality (Ilala, Temeke, Kigamboni, Kinondoni) FEs and time period of transaction FEs. Column 2 additionally controls for 20K area FEs, taken as the nearest for transactions outside of 20K boundaries. Standard errors in parentheses are clustered by 20K area.

Table 2: Prices and Plot Sizes (OLS)

(a) Ln plot price on ln plot size (gridcell level)

	(1)	(2)	(3)
Ln plot size	0.66	0.47	0.51
	(0.083)	(0.068)	(0.078)
Mean Outcome	17	17	17
Period*Source FE	✓	✓	✓
20k*MTAA FE		✓	✓
Amenities			✓
N (gridcells)	2885	2885	2885
N (plots)	998	998	998

Note: This table presents regressions of log price on log plot size. The outcome is always the log price of a bareland transaction from the Dalali or Occupier survey. Controls vary across columns: transaction period by source (Dalali or Occupier survey) FEs (cols 1-3), 20K*MTAA Area FEs (cols 2-3), and amenities (col 3). Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of: river, wetland, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Note that in this specification, the dummy for wetland within 100m is co-linear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

(b) Ln plot price per square meter on plot size (gridcell level)

	(1)	(2)	(3)	(4)
Log plot size	-0.53	-0.49		
	(0.068)	(0.078)		
Medium			-0.27	-0.23
			(0.049)	(0.052)
Large			-0.60	-0.55
			(0.078)	(0.081)
Mean Outcome	10	10	10	10
Period*Source FE	✓	✓	✓	✓
20k*MTAA FE	✓	✓	✓	✓
Amenities		✓		✓
N (gridcells)	2885	2885	2885	2885
N (plots)	998	998	998	998

Note: This table presents regressions of log price per square metre on log plot size or binned size (small, medium, and large). The outcome is always the log price per square metre of a bareland transaction from the Dalali or Occupier survey. Controls vary across columns: transaction period by source (Dalali or Occupier survey), and 20K*MTAA Area FEs (cols 1-4); and amenities (cols 2 and 4). Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of: river, wetland, planned open space, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Note that in this specification, the dummy for wetland within 100m is co-linear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

45

Table 3: Ln Price per square meter (RD by gap size, gridcell level)

	(1)	(2)	(3)	(4)
Panel A: gap>400sqm				
Own Larger	-0.34	-0.32	-0.37	-0.31
	(0.15)	(0.14)	(0.14)	(0.15)
Mean Outcome	9.9	9.9	9.9	9.9
Period*Source FE	✓	✓	✓	✓
20k*MTAA FE	✓	✓	✓	✓
Amenities		✓	✓	✓
Segment FEs			✓	
Segment Sample			✓	✓
N (gridcells)	805	805	744	744
N (plots)	283	283	240	240
Panel B: gap<100sqm				
Own Larger	-0.052	-0.024	-0.16	-0.057
	(0.10)	(0.089)	(0.087)	(0.10)
Mean Outcome	10	10	10	10
Period*Source FE	✓	✓	✓	✓
20k*MTAA FE	✓	✓	✓	✓
Amenities		✓	✓	✓
Segment FEs			✓	
Segment Sample			✓	✓
N (gridcells)	824	824	681	681
N (plots)	415	415	304	304

Note: This table presents RD regressions across neighbouring insula boundaries. All panels restrict the sample to insula pairs with at least one observation on each side of the insula-pair boundary. Panel A and B restrict the sample to within 100m of the insula-pair boundary for insula pairs with at least a 400sqm gap (pnl A) or 100sqm gap (pnl B) in mean size. The RD specification takes an indicator for whether a gridcell is in an insula with mean plot size larger than the nearest neighbouring insula, and always controls for linear distance to the boundary between insula pairs on each side of the boundary. For Panel A, mean distance to the boundary is 36m, median 32m, 75th percentile 48m, and 95th percentile 80m. For Panel B, the mean distance to the boundary is 29m, median 26m, 75th percentile 39m, and 95th percentile 60m. The outcome is always the log price per square metre of a bareland transaction from the Dalali or Occupier survey. Controls vary across columns: transaction period by source (Dalali or Occupier survey) FEs (cols 1-4), 20K*MTAA Area FEs (cols 1-4), and amenities (cols 2-4), and insula-segment FEs (col 3). Column 4 uses the same sample as column 4, but does not include segment FEs. Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of: river, wetland, planned recreation, planned education, planned nursery school, planned religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Note that in this specification, the dummy for wetland within 100m is co-linear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table 4: Built outcomes on ln plot size (gridcell level)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Size of building on plot	(5) Multiple buildings on plot
Panel A: 20k*MTAA FE					
Ln plot size	-0.087	-0.031	0.11	2.71	0.079
	(0.0025)	(0.0091)	(0.017)	(2.42)	(0.0070)
Mean Outcome	0.11	0.49	5.3	96	0.19
N (gridcells)	94789	94789	46465	94789	94789
N (plots)	36215	36215	17822	36215	36215
Panel B: 20k*MTAA FE +Amenities					
Ln plot size	-0.078	-0.0014	0.14	10.3	0.091
	(0.0026)	(0.0093)	(0.018)	(2.60)	(0.0072)
Mean Outcome	0.11	0.49	5.3	96	0.19
N (gridcells)	94789	94789	46465	94789	94789
N (plots)	36215	36215	17822	36215	36215

Note: This table presents regressions of five quantity outcomes on log plot size. The outcomes are log price of a bareland transaction from the Dalali or Occupier survey (col 1), share of the plot area that is built (col 2), an indicator for whether the plot is built (col 3), log of own plot area that is intersected with buildings that are larger than 30 square meters (col 4), size of the largest building where missing values are treated as zeros (col 5), and an indicator for multiple buildings on the plot (col6). Controls vary across columns and panels: transaction period by source (Dalali or Occupier survey) FEs (col 1), 20K*MTAA Area FEs (panel A), and amenities (panel B). Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of: river, wetland, planned open space, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Standard errors in parentheses are clustered by insula.

Table 5: Built outcomes per square meter (RD by gap size, gridcell level)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Size of building on plot	(5) Multiple buildings on plot
Panel A: gap>400sqm					
Own Larger	-0.037	0.0064	-0.011	2.71	0.043
	(0.0051)	(0.016)	(0.042)	(4.43)	(0.016)
Mean Outcome	0.093	0.47	5.3	92	0.20
20k*MTAA FE	✓	✓	✓	✓	✓
Amenities	✓	✓	✓	✓	✓
Segment FEs	✓	✓	✓	✓	✓
N (gridcells)	22483	22483	10526	22483	22483
N (plots)	9066	9066	4219	9066	9066
Panel B: gap<100sqm					
Own Larger	-0.0094	-0.0030	-0.0096	-0.69	-0.0080
	(0.0042)	(0.013)	(0.027)	(3.46)	(0.011)
Mean Outcome	0.12	0.50	5.2	97	0.18
20k*MTAA FE	✓	✓	✓	✓	✓
Amenities	✓	✓	✓	✓	✓
Segment FEs	✓	✓	✓	✓	✓
N (gridcells)	30460	30460	15079	30460	30460
N (plots)	15157	15157	7469	15157	15157

Note: This table presents RD regressions across neighbouring insula boundaries. Panel A restricts the sample to within 100m of the insula-pair boundary, and for insula pairs with at least a 400sqm gap in mean size. Panel B restricts the sample to within 100m of the insula-pair boundary, and for insula pairs with no more than a 100sqm gap in mean size. The RD specification takes an indicator for whether a gridcell is in an insula with mean plot size larger than the nearest neighbouring insula, and always controls for linear distance to the boundary between insula pairs on each side of the boundary. For Panel A, the mean distance to the boundary is 33m, median 30m, 75th percentile 44m, and 95th percentile 69m. For Panel B, the mean distance to the boundary is 29m, median 26m, 75th percentile 39m, and 95th percentile 63m. The outcomes vary across columns. Controls always include 20K*MTAA Area FEs, amenities, and insula-segment FEs. Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of: river, wetland, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Standard errors in parentheses are clustered by insula.

Table 6: Built outcomes: super insula RD (gridcell level)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Size of building on plot	(5) Multiple buildings on plot
Own Larger	-0.0013	-0.00057	-0.0036	1.52	0.0080
	(0.0026)	(0.011)	(0.021)	(2.81)	(0.0091)
Own Smaller × Dist. (km)	0.050	0.20	-0.062	31.4	0.058
	(0.017)	(0.066)	(0.12)	(15.7)	(0.051)
Own Larger × Dist. (km)	-0.033	0.014	0.081	-3.92	-0.0046
	(0.018)	(0.070)	(0.13)	(17.0)	(0.057)
Ln plot size	-0.066	0.025	0.18	15.1	0.11
	(0.0032)	(0.013)	(0.027)	(3.37)	(0.011)
Mean Outcome	0.11	0.49	5.3	96	0.19
20k*MTAA FE	✓	✓	✓	✓	✓
Amenities	✓	✓	✓	✓	✓
Segment FEs	✓	✓	✓	✓	✓
N (gridcells)	92753	92753	45559	92753	92753
N (plots)	35525	35525	17474	35525	35525

Note: This table presents RD regressions across neighbouring Super-Insula boundaries. We discard super-insula pairs where the minimum distance between the two is more than 30m (allowing for no more than a large road to pass between the two). The RD specification takes an indicator for whether a gridcell is in a Super-Insula with mean plot size larger than the nearest neighbouring Super-Insula, and always controls for linear distance to the boundary between Super-Insula pairs on each side of the boundary. The mean distance to the boundary is 51m, median 41m, 75th percentile 76m, and 95th percentile 130m. share of the plot area that is built (col 1), an indicator for whether the plot is built restricted to the largest building size being at least 30sqm (col 2), log of the total area of the three largest buildings on the plot if the largest building is at least 30sqm (col 3), size of the largest building where missing values are treated as zeros (col 4), and an indicator for multiple buildings on the plot (col 5). Controls always include 20K*MTAA Area FEs and super-insula pair FE ("segment FEs"). Amenities include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies indicating within 100m of the following: river, wetland, planned open space, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Standard errors in parentheses are clustered by insula.

Table 7: Land values & housing development by planned amenities (gridcell level)

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Size of building on plot	(6) Multiple buildings on plot
Ln plot size	0.51	-0.078	-0.0014	0.14	10.3	0.091
	(0.078)	(0.0026)	(0.0093)	(0.018)	(2.60)	(0.0072)
Elevation (m)	0.0027	0.00090	0.0028	0.0032	0.70	0.0010
	(0.0028)	(0.000097)	(0.00043)	(0.00067)	(0.098)	(0.00032)
Ruggedness	-0.016	-0.0059	-0.017	-0.012	-3.95	-0.0083
	(0.032)	(0.00097)	(0.0038)	(0.0089)	(0.89)	(0.0031)
River/stream 100m	-0.0092	-0.028	-0.12	-0.072	-27.0	-0.057
	(0.18)	(0.0052)	(0.022)	(0.058)	(4.62)	(0.019)
Water/wetland 100m	0.0	0.0	-0.1	-0.1	-14.5	-0.0
	(.)	(0.0)	(0.0)	(0.2)	(6.3)	(0.0)
Dist (km) paved major road	-0.13	-0.015	-0.041	-0.062	-12.5	-0.031
	(0.038)	(0.0016)	(0.0070)	(0.012)	(1.73)	(0.0053)
Z-index: 3 Ins. Characteristics	0.031	0.0037	0.016	0.0096	4.43	0.0053
	(0.035)	(0.0013)	(0.0058)	(0.010)	(1.43)	(0.0041)
Pln. recreation in 100m	-0.0013	-0.00059	-0.0077	-0.0081	-2.03	-0.0063
	(0.054)	(0.0019)	(0.0071)	(0.012)	(1.78)	(0.0053)
Pln. nursery school in 100m	0.060	0.0063	0.018	0.030	5.13	0.010
	(0.055)	(0.0026)	(0.0097)	(0.017)	(2.33)	(0.0071)
Pln. religious site in 100m	0.052	0.0022	0.016	-0.0058	3.49	-0.0012
	(0.069)	(0.0030)	(0.012)	(0.020)	(3.41)	(0.0086)
Pln. education in 100m	0.15	-0.0044	-0.0074	-0.022	-4.27	-0.0075
	(0.085)	(0.0030)	(0.011)	(0.021)	(2.67)	(0.0082)
Pln. service trade in 100m	-0.071	-0.0012	-0.0025	-0.0046	0.57	-0.0038
	(0.12)	(0.0043)	(0.016)	(0.030)	(3.65)	(0.011)
Pln. housing estate in 100m	-0.14	0.00071	0.0069	0.0037	2.03	-0.019
	(0.12)	(0.0075)	(0.031)	(0.048)	(7.42)	(0.023)
Pln. public building in 100m	0.0066	-0.0052	-0.0094	-0.043	-4.67	-0.017
	(0.11)	(0.0044)	(0.016)	(0.029)	(3.64)	(0.012)
Pln. cemetery in 100m	0.080	0.0041	0.038	-0.046	4.44	0.015
	(0.15)	(0.0051)	(0.019)	(0.034)	(4.49)	(0.015)
Pln. any other non-res in 100m	0.13	-0.0022	-0.021	0.0071	-4.01	0.0028
	(0.098)	(0.0031)	(0.012)	(0.023)	(2.65)	(0.0083)
Mean Outcome	17	0.11	0.49	5.3	96	0.19
Period*Source FEs	✓					
20k*MTAA FEs	✓	✓	✓	✓	✓	✓
N (gridcells)	2885	94789	94789	46465	94789	94789
N (plots)	998	36215	36215	17822	36215	36215

Note: This table presents regressions of both price and built outcomes on log plot size. The outcomes vary across columns: This table presents regressions of the following outcomes: log price (col 1), share of gridcell built (col 2), an indicator for plot built (col 3), log of own plot area intersected with buildings greater than 30 sqm (col 4), size of the largest building on the plot where missing values are treated as zeros (col 5), and an indicator for multiple buildings on the plot (col 6). Controls vary across columns: transaction period by source (dalali or household questionnaires) FEs (col 1), and 20K*MTAA Area FEs (cols 1-6). All specifications show coefficients for amenities, which include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies indicating within 100m of the following: river, wetland, planned open space, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Each observation is a gridcell. Standard errors in parentheses are clustered by insula.

Table 8: Housing development by implemented amenities (gridcell level)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Size of building on plot	(5) Multiple buildings on plot
Ln plot size	-0.078 (0.0025)	0.00023 (0.0093)	0.14 (0.018)	10.6 (2.58)	0.091 (0.0072)
Elevation (m)	0.00089 (0.000098)	0.0028 (0.00043)	0.0032 (0.00067)	0.70 (0.098)	0.00099 (0.00032)
Ruggedness	-0.0055 (0.00097)	-0.016 (0.0038)	-0.011 (0.0090)	-3.68 (0.89)	-0.0077 (0.0031)
River/stream 100m	-0.028 (0.0053)	-0.12 (0.023)	-0.078 (0.058)	-27.3 (4.61)	-0.061 (0.019)
Water/wetland 100m	0.0058 (0.0088)	-0.074 (0.032)	-0.088 (0.16)	-16.4 (6.03)	-0.039 (0.038)
Dist (km) paved major road	-0.015 (0.0016)	-0.041 (0.0071)	-0.061 (0.012)	-12.1 (1.65)	-0.030 (0.0053)
Z-index: 3 Ins. Characteristics	0.0036 (0.0013)	0.016 (0.0057)	0.0098 (0.010)	4.33 (1.41)	0.0052 (0.0041)
Impl. farming in 100m	0.00015 (0.0023)	0.0079 (0.0086)	-0.0084 (0.015)	0.22 (1.97)	0.00049 (0.0063)
Impl. recreation in 100m	0.0068 (0.0031)	0.0035 (0.011)	0.037 (0.019)	7.04 (3.48)	0.0032 (0.0091)
Impl. religious site in 100m	0.011 (0.0043)	0.029 (0.016)	0.015 (0.027)	9.53 (5.88)	0.016 (0.012)
Impl. education in 100m	0.00037 (0.0044)	0.0025 (0.015)	-0.024 (0.027)	-2.02 (4.10)	-0.0024 (0.011)
Impl. cemetery in 100m	0.0031 (0.0054)	0.021 (0.018)	-0.018 (0.034)	4.77 (4.58)	0.029 (0.016)
Impl. service trade in 100m	0.016 (0.0052)	0.049 (0.019)	0.050 (0.029)	8.20 (4.18)	0.044 (0.018)
Impl. nursery school in 100m	0.0071 (0.0072)	0.036 (0.027)	0.0013 (0.039)	5.79 (6.52)	0.016 (0.020)
Impl. other non-res in 100m	-0.0050 (0.0076)	-0.032 (0.025)	0.0066 (0.041)	1.20 (7.01)	-0.00038 (0.018)
Impl. public building in 100m	-0.0064 (0.010)	-0.035 (0.040)	-0.015 (0.060)	-9.04 (9.17)	0.031 (0.033)
Impl. housing estate in 100m	0.050 (0.011)	0.20 (0.063)	0.17 (0.095)	42.5 (11.3)	0.17 (0.062)
Impl. unknown non-res in 100m	-0.0023 (0.061)	0.13 (0.23)	-0.15 (0.044)	9.99 (35.4)	-0.14 (0.042)
Unused, kept in 100m	0.0033 (0.0026)	0.00055 (0.010)	-0.020 (0.016)	-2.21 (2.45)	-0.0032 (0.0077)
Unused, unkept in 100m	-0.0082 (0.0027)	-0.034 (0.010)	-0.041 (0.017)	-9.97 (2.38)	-0.024 (0.0070)
Impl. as residential in 100m	-0.0033 (0.0033)	-0.0017 (0.014)	-0.012 (0.023)	-1.69 (3.41)	-0.0068 (0.011)
Mean Outcome	0.11	0.49	5.3	96	0.19
20k*MTAA FEs	✓	✓	✓	✓	✓
N (gridcells)	94789	94789	46465	94789	94789
N (plots)	36215	36215	17822	36215	36215

Note: This table presents regressions of both price and built outcomes on log plot size. The outcomes vary across columns: This table presents regressions of the following outcomes: share of gridcell built (col 1), an indicator for plot built (col 2), log of own plot area intersected with buildings greater than 30 sqm (col 3), size of the largest building on the plot where missing values are treated as zeros (col 4), and an indicator for multiple buildings on the plot (col 5). Controls include 20K*MTAA Area FEs and amenities including average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies indicating within 100m of a river, wetland. For the within 100m of an implemented land use, we pick those whose corresponding implemented use has at least 100 gridcells in the largest sample (all gridcells) plus implemented as residential, unused kept, and unused unkept.

Supplementary Data and Appendix Evaluating Urban Planning: Evidence from Dar es Salaam

Vernon Henderson (*LSE*)   Francisco Libano-Monteiro (*LSE*)   Martina Manara (*Sheffield*)

Guy Michaels (*LSE*)   Tanner Regan (*GWU*)

Table of Contents

A Data source descriptions

A.1 Project maps and planning treatments

We collected three types of project maps. First are town planning drawings (TPDs) made by planners, and which we have for all the project areas, except Mwongozo. The TPDs are also called "neighborhood layouts", since they depict residential plots, non-residential plots with their planned use, and roads with markers of road reserve width. These drawings were created in hardcopy format and approved by the Town Planning Department of the Ministry of Lands, Housing and Human Settlements Development (MLHHSD) between 1997 and 2009. The hardcopies were scanned and shared with us, and we georeferenced them.

Second are survey maps (SMs), which were prepared by the MLHHSD after approval of TPDs, and which we again have for all project areas except Mwongozo. SMs show how surveyors physically demarcate land into plots, based on TPDs layouts. In practice, this entails putting down beacons in the ground, typically at the block corners, to determine exact coordinates (latitude and longitude) using theodolites. Thereafter, more beacons are placed in alignment with each plot's corners. Each beacon number is then associated with its coordinates, which enables the plot boundaries to be precisely recorded in software. The SMs were also given to us as digital copies. We transformed them from vectoral drawings (.dwg) into polygon shapefiles (.shp) and georeferenced them.

Finally, there are cadastral data, which we obtained from the MLHHSD for the municipalities of Kigamboni, Kinondoni, and Ubungo. These data cover all our project areas except for Buyuni, Mwanagati, Tuoangoma, and Kijichi. The cadastral database contains registered SMs, which are approved and recorded in GIS software by the MLHHSD, ready for issuance of title deeds and land rent bills. Therefore, while SMs are implemented town planning drawings, cadastral drawings constitute the legally registered version of SMs.[20].

Given the incompleteness of our three sources, we carefully designed a procedure to assemble a dataset as complete and accurate as possible. Our procedure involved discussions with the '20,000 plots' project secretary and town planners, who approved of our methodology. Our procedure can be summarized as follows. We use the SMs as the basis for our dataset of plot boundaries (polygons), since they are more up to date than the TPDs and more complete than cadastral data. Where SMs are not available (i.e., Mwongozo) we instead use cadastral data. To make sure that the cadastral data is restricted to plots implemented as part of the 20K project, we restrict them to plots, which were registered between 2000 and 2010, and which fall within the known boundaries of the Mwongozo project area. [21]

---

[20]The cadastral data also contain plots that are not easily distinguishable from those implemented as part of the project

[21]Table A.8 below summarizes the main data source used for each project area to conduct analysis on residential

We draw on the TPDs for two purposes. First, we use them to update the planned use of non-residential plots in our plot boundary data where the SMs are missing this information. [22] Second, we digitize planned road reserves and their widths by manually tracing the georeferenced TPDs. [23]

A.2 Data derived from satellite imagery

To study the quantity and quality of housing we use Worldview satellite images purchased from Airbus Defense and Space Limited. These data provide pansharpened color images at approximately 0.5-meter resolution. [24] The images cover all of the project areas with a 500m buffer outside them. While we aimed to obtain the most recent image of each area, the precise dates vary: Kibada, Kijichi, Kisota, Mwongozo, and Tuangoma (July 2019); Buyuni (July 2020); Mivumoni (Sept 2020); Bunju, Mbweni Mpiji, Mwanagati, Mbweni JKT, and Mbweni Malindi (March 2021).

A.2.1 Buildings and fences

We employed a company, Ramani Geosystems, which specializes in geospatial digitization, to trace buildings from the imagery. The building data include: (i) building footprints, (ii) roof quality (painted metal or tiled, unpainted metal, and rusted metal), and (iii) whether the building is under construction or not. The company also traced out fences and hedges.

A.2.2 Roads

We also used the satellite imagery to trace and classify modern roads. We did this using trained research assistants from our field staff team in Dar es Salaam. The modern roads were constructed as follows. The digitized planned roads were taken as a starting point. First, we add road extensions (polylines) wherever a road appeared in the modern image, but not in the plan. Second, we segment roads wherever they intersect. Third, we classify each segment's road type (none, footpath, dirt road, or paved) and road width (in meters). Road segments that were planned but do not appear in a modern image were assigned type 'none' and width of zero.

---

plots.

[22]In Mwongozo, where we lack SMs, we use the cadastral data definition of a plot's planned use.

[23]To compare planned and implemented non-residential uses, we combine these planning data with data from (i) satellite images capturing road implementation (see Section A.2) and (ii) fieldwork enumeration of implemented non-residential uses and current maintenance (see Section D.5).

[24]These images combine panchromatic images at a resolution of 0.5 meters with multispectral images at a resolution of 2.5 meters.

A.3 Additional data sources

A.3.1 Digital elevation and ruggedness

We measure elevation and ruggedness using SRTM elevation at a horizontal resolution of 1 arc-second, or approximately 30 meters (United States Geological Survey, 2000). This directly gives us a measure of elevation above sea level. Following Nunn and Puga (2012), we use the data to compute the local ruggedness as the standard deviation of elevation across the eight neighbors of each approximately 30m by 30m cell (in the SRTM data).

A.3.2 Openstreetmap (local geographic features)

We measure proximity to natural features using data from OpenStreetMap contributors (2017). We select data from Openstreetmap in Dar es Salaam to create two collections of geospatial data: (i) rivers and streams, and (ii) water bodies and wetlands. We then measure proximity for our units of analysis to the nearest feature in each of these two collections.

A.3.3 Inflation Rates

Throughout the paper we report prices in 2021 Tanzanian Shillings (TZS). Unless explicitly noted, we inflate prices to the year 2021 for any source data reported in Tanzanian Shillings for earlier years. To do so we use annual inflation rates from Statista (2022) which compiles data published by the IMF. For price data reported in Tanzanian shillings for the year $y_0 < 2021$ we inflate by using the product $\prod_{y=y_0+1}^{2021}(1 + i_y)$, where i_y is the inflation rate for year y.[25]

A.3.4 Initial price data of government-sold plots

To measure the initial price that the Tanzanian government charged when it sold the plots, we obtained data from the project secretary, which is also partially reported in Mero (2008, 2009). These are also used in Mwiga (2011) and Kironde (2015). The initial prices for each 20K area are reported in A.9.

A.3.5 Price of plots sold in market transactions

We collected data on the prices of plots sold in market transactions from questionnaires we gave to (i) real estate agents and (ii) current residents. We also obtained estimates of sale prices for plots of various sizes from (i) interviews with local leaders and (ii) real estate agent questionnaires.

---

[25]According to https://www.exchangerates.org.uk/USD-TZS-spot-exchange-rates-history-2021.html, accessed on 21 June 2023, the mean exchange rate in 2021 was about 2314.5 TZS per US Dollar (USD).

A.4 Questionnaire, interview, and enumeration data

With the aid of our research assistants, based in Dar es Salaam, we administered questionnaires, interviews, and enumerations, which we describe below. Precautions were taken to ensure the safety of the enumerators (research assistants), for instance by having them work in pairs and report to the local mtaa office daily.

A.4.1 Preliminary interviews with experts

From July 2021 - October 2022 we held ten interviews with eight experts involved in key aspects of the project, including government officials and academics. These interviews focused mostly on obtaining institutional details about the planning and execution of the project.

A.4.2 Local leader interviews

Sampling frame

The mtaa (plural, mitaa) is the smallest administrative unit in urban Tanzania, equivalent to a sub-ward, and their boundaries do not coincide with the boundaries of 20k areas.

Each mtaa has a local government office composed of one elected mtaa chairperson (mwenyekiti), one government appointed executive officer (mtendaji), and five members of the mtaa committee. Collaborating with branch leaders (wajumbe) – elected political figures who are not formally integrated into the local government structure – the mtaa office performs several governance functions, including supervision of land transactions, land disputes, and community life. We liaised with all the mtaa offices in the areas covering '20,000 plots' project, for the purposes of collecting relevant research permits, ensuring research stakeholder buy-in, and collecting preliminary information, through a questionnaire to local leaders.

To identify the relevant mitaa, we overlapped the map of project areas boundaries (Section A) with a government map of Dar es Salaam's mtaa boundaries. We found a total of 38 mitaa containing all the planned project areas. Two research assistants visited these mitaa to verify that the '20,000 plots' project has been implemented and conduct interviews with mtaa leaders, and we found that the program provided private residential plots in (parts of) 34 mitaa, whose leaders we interviewed.[26]

---

[26]We found that the project was not implemented in three mitaa (Kibaga, Kinyerezi and Kifuru) of Ilala municipality (corresponding to Kinyerezi project area), which we confirmed with past leaders of those mitaa and one land officer of Ilala municipality. Furthermore, we found that one mtaa in Kigamboni municipality had only 37 plots, of which 25 are owned by a public agency (National Social Security Fund, NSSF), and the remaining 12 were designated for public uses.

Interview details and protocol

Interviews took place from September 2021 - October 2021 and the data were recorded in double copy, through a paper questionnaire and an ODK app.[27] Two research assistants conducted interviews in the local language (Swahili), with one of the authors participating remotely. Each interview lasted between one and a half hours to three hours. The target respondent was the mtaa chairperson, whose responses we recorded, but executive officers and wajumbe were occasionally present. Given the objective and 'non-political' nature of the questionnaire, the presence of multiple respondents was deemed useful to triangulate and complement information. Overall, our interviewees included the 34 mtaa chairpersons, 22 mtaa executive officers, and 18 wajumbe.

The interviews that we conducted with the mtaa leaders provided information that was directly useful, and which we also used to design the subsequent questionnaires with real estate agents and residents, as described in next sections. Furthermore, we asked the mtaa leaders to provide lists and contacts of real estate agents operating in their mitaa, which was essential for sampling and recruitment of these respondents (see section 2.1). Finally, the mtaa leaders confirmed the mtaa boundary and '20,000 plots' project area boundary within the respective mtaa starting from our printed A1 maps. Any amendments were then digitized in updated boundary layers.

Interview questions

The interview questions were structured in 11 sections. Section 1 gathered information on respondents, including demographics. Section 2 asked information about residential plots in the mtaa, for example, statistics on land use, built construction, and processes and opinions on opportunities and constraints to land development. Section 3 enquired about other formal plots with each mtaa, outside of the 20k areas. Section 4 focused on land markets in 20K and non-20K informal plots, asking about volumes of land sales and predictions of bare land current prices for different plots sizes. This is one of the sources used for current price data (see Section C). We also asked questions on local leaders' involvement in land sales and collected contacts for our real estate agent questionnaire (see Section D.3). Section 5 focused on residents' profiles, for instance, asking questions on household income in 20K and non-20K areas within the mtaa. Section 6 asked about land titles and other documentation held by landowners. Further sections asked about infrastructure provision in 20K and non-20K areas including roads and open space (Section 7), electricity, water, and sanitation (Section 8). Section 9 asked about housing units provided by real estate firms and obtained those

---

[27]ODK is an open-source mobile data collection platform.

firms' contact information. Section 10 asked about other services, such as public safety, transport and schools, and Section 11 concluded by asking local leaders to confirm the mtaa boundaries on our map.

A.4.3   Real estate agent questionnaire

Sampling frame

The real estate agent questionnaire was carried out in two phases: a phone questionnaire (phase one) and a field questionnaire (phase two). In early November 2021, we contacted 48 real estate agents whose contacts we had obtained from the mtaa leaders (section 2). We obtained from the real estate agents preliminary information including the mitaa in which they operate; whether they operate in 20,000 plots areas, non-20,000 plots, or both; and whether their work covered rentals, sales, or both. For the field questionnaire itself we targeted real estate agents who (i) had some experience of sales of 20,000 plots (at least 20 transactions); (ii) had experience with of sales of non-20,000 plots in the same mitaa, when these exist (at least 5 transactions). [28]. In addition, real estate agents who achieved the highest Likert score (based on the enumerators' assessment of the real estate agents' knowledge and reliability) were targeted regardless of the number of transactions they reported. Through this process, 20 real estate agents were targeted for the field questionnaire. However, only 12 of these real estate agents participated in the study. In fact, some real estate agents have other primary occupations, and so many could not take entire days off work to assist us. However, through a process of snowballing we recruited 6 additional real estate agents meeting our criteria. This gave us a final pool of 18 real estate agents respondents.

Questionnaire details and protocol

From November 2021 to December 2021, our research assistants (RAs) enumerated all land transactions facilitated by the real estate agents they interviewed. The research assistants were supplied with A1 printed maps displaying mtaa boundary and '20,000 plots' project area boundary overlaid with satellite imagery. Using these maps, the real estate agents were asked to identify the plots whose sale they facilitated and take the RAs to the actual plots. The RAs recorded the sales using a paper questionnaire and an ODK App. In some cases, the RAs also recorded the plot boundaries manually on our A1 map. For example, if the transaction involved subdivision, the RAs demarcated original and subdivided plot

---

[28]The thresholds of 20 and 5 sales were selected since we anticipated that one day of fieldwork would enable us to visit at most 25 plots

boundaries. Furthermore, the RAs demarcated boundaries of informal transacted plots. Finally, the data on the sold plots were digitized and added to our digital project map.

In total, we collected information on 1326 transactions, including: 1126 on 20K plots, 110 on non-20K formal plots, and 90 on non-20K informal plots. For clarity, non-20K formal plots are surveyed, included in a town planning drawing, and eligible for land titles (as 20K plots); however, they were not supplied as part of the 20K project. Typically, they result from ex-post regularization of informal plots. Thus, they may be formal plots in predominantly informal neighborhoods. We note that most real estate agents were able to read maps and were familiar with the mitaa in which they operate, which made the process of data collection relatively smooth.

Next, we create a set of bare land transactions including inside and outside of 20k areas. We keep only the transactions for plots that were empty at the transaction date, or in the case of listing prices, to those that were empty at the time of the questionnaire. This leaves us with 1142 transactions, including: 967 on 20K plots, 96 on non-20K formal plots, and 79 on non-20K informal plots.

To match the plot-level price data uniquely to the 20k plots, we impose further restrictions on the data. For plots with multiple bare land records, we keep only the latest transaction, and if there are no transactions, we keep the listing price. We further verify that land transactions inside the 20k project areas (i.e., excluding non-20k formal and informal plots) correspond to residential plots within 20k areas. At the end of this process, we match 830 20K plots to recent bare land transactions, and 689 of these contain data on prices.

A.4.4 Questionnaire content

Questionnaire 1 – phone questionnaire

The phone questionnaire asked questions about real estate agents' demographics and experience of supervising sales in the mtaa. For example, we asked whether the real estate agents worked in 20K or non-20K areas, or both, and in sale or rental markets or both. We asked about volumes of sales and current prices of bare land for plots of different sizes in 20K versus non-20K areas, which is also used as a source of current prices (see Section C). We also asked questions about rental prices for unfurnished properties of different sizes in 20K versus non-20K areas.

Questionnaire 2 – field questionnaire

The field questionnaire recorded transaction id and area (e.g., 20K versus non-20K), estimated plot size, period and year of transaction, presence of written record of time, price in

million TZS and written record of price, development status at the time of sale, real estate agents' assessment of information reliability (e.g., quality of recollection), and enumerator assessment. Further open-ended questions asked real estate agents to tell us about processes and stages of land transactions in the mtaa, including in 20K and non-20K areas. We also asked questions on the involvement of the mtaa office or formal lawyers in the ratification of bills of sale in 20K and non-20K areas.

A.4.5 Resident questionnaire

Sampling frame

For the resident questionnaire, we started with the universe of 17,333 residential plots where the processed satellite imagery showed at least one building. One of the '20,000 plots' areas (Mwongozo) was excluded from the resident questionnaire, due to cost-effectiveness considerations: the area has low development rate and high transport costs. Similarly, we excluded a small exclave of Kijichi, which has only about 30 plots, most of which are undeveloped.

Our assessed questionnaire capacity was about 3,300 interviews (19% of the population), requiring each of our seven enumerator teams to complete 15 interviews per week. To ensure that this target of 15 interviews per week was feasible, we assigned each team a weekly questionnaire cluster of randomly selected plots, through a process that we hereby describe. Of the above-mentioned 17,333 plots we randomly selected 5,900, and grouped them in questionnaire clusters of approximately 35 plots each.[29]

Of the 5,900 randomly selected plots, we ended up dropping two clusters, with a total of 70 plots, which we used for a pilot. Of the remaining 5,830 plots, 4,613 plots were eligible for interview (for reasons explained below).[30] Our enumerators completed 3,231 questionnaires, reaching 98% of the maximum achievable sample we had aimed at (3,300), and covering 18.64% of the initial universe of 17,333 plots.[31]

Interview details and protocol

In June 2022, a team of fourteen local town planning graduates working as our enumerators was given four weeks of training on the questionnaire, including two weeks under the

---

[29]Each questionnaire cluster was designed to contain plots that were in spatial proximity and fully contained within one program area. Consequently, some clusters contained fewer than 35 plots.

[30]Two additional questionnaire clusters with 35 plots each were dropped during the questionnaire's implementation - one due to a local land conflict and another due to personal circumstances of enumerators.

[31]Given the complexity of the questionnaire protocol and questionnaires (see next sections), we did not collect statistics on the reasons why interviews could not be administered in given plots. We had no respondent dropping halfway through an interview.

supervision of one of the co-authors. They conducted the questionnaire from July 2022 - February 2023, working in pairs and residing in their respective project areas for the duration of data collection. This allowed enumerators to avoid long commutes and to embed in local areas, and secure support from local leaders when necessary. A fieldwork supervisor periodically visited each team, ensuring adherence to protocols and accuracy in the delivery of questionnaires. Each team also reported daily to one of the co-authors.

Each interview team worked from Wednesday to Sunday each week, to maximize the likelihood of finding the landholders at home. At the start of each workweek (typically on Wednesday), each team visited its designated plots accompanied by a local leader (mjumbe), and completed an ODK report confirming that they did so. These visits also enabled enumerators to identify plots that were ineligible for data collection or those whose eligibility was undetermined.[32]

To all the plots that were eligible for interview (4,683) and those deemed undetermined (23), enumerators delivered a leaflet written in the local language (Swahili) and signed by the mtaa chairperson. This provided introductory information on the research project and the interview planned for the weekend. When possible, enumerators spoke to people living on the plots, and otherwise left the leaflet attached to the gate or under the door. [33] Prospective respondents could contact the enumerators using details provided on the leaflet to ask the enumerators for clarifications and book interview times. These weekly preparatory activities took place in parallel with the enumeration of non-residential plots (see next section 4).

Within each interview plot, the target respondent was designated as one of the following: (i) the landowner (named on any property document); or (ii) if no landowner lived in the plot, the head of a resident usufructuary household (i.e., a person who is not part of the landowner household, but allowed to live there for free), or (iii) if none of the above lived in the plot, the head of a resident tenant household (i.e., not part of the landowner household, but allowed to live there in exchange for rent). In cases where there were multiple people in each category (e.g., joint landowners, multiple usufructuary households, or multiple renting households), we interviewed only one. Guardians and servants (those not part of the landholder household but paid to live and/or work on the plot) were not interviewed. Wherever possible, the enumerators tried to interview their target rather than another respondent.

Every four weeks, a catch-up week was organized, enabling enumerators to revisit plots

---

[32] plots were ineligible because: (i) they were undeveloped – possibly due to changes in land use since the imagery was taken or measurement error in the imagery processing or the project maps (129 plots); (ii) under construction (398 plots); (iii) built but uninhabited (280 plots); (iv) built but inhabited only by guardians, staff, or housekeepers (149 plots); (v) other reasons (238 cases). Therefore, we had 1,194 ineligible plots in total. In addition, 23 plots were deemed undetermined (23 plots), as enumerators were unable to establish if the building was inhabited or if residents were eligible for interview.

[33] We decided to not leave leaflets with neighbors to avoid undue concerns or interference.

assigned in prior weeks, where they did not find the target respondent at home. If the target respondent was still unavailable, enumerators were allowed to interview a proxy - an adult member of the target respondent's household, ideally the spouse or partner. In total, we interviewed 215 proxies, including current or former spouses and partners (117), children (54), child-in-law (1), grandchildren (2), siblings (33) or other household members (8). Therefore, proxies constitute 6.7% of the plots where we interviewed respondents.

The next section explains the questionnaire administered to target respondents and their proxies.

Questionnaire content

The questionnaire was structured in 13 sections. Section 1 asked questions about the residents, and identified respondents, including the target and (where needed) the proxy'. Section 2 collected information on current land uses, while section 3 focused on road access and plot characteristics (e.g., counts of buildings with residential and non-residential use). Section 4 asked about infrastructure, including sanitation, sources of water and energy, and garbage disposal. Section 5 asked about the main (largest footprint) residential building - its construction and finishing materials for the walls and roof, and the presence of indoor toilet and kitchen facilities. Section 6 enquired about rental income, while Section 7 asked questions about the history of plot acquisition and development, such as the year and mode of acquisition, and the timing of construction. Sections 8, 9 and 10 asked about the respondent's education and employment, including current work or last work before retirement, while section 11 asked about household wealth and how it is held. Section 12 contained questions about the neighborhoods' amenities and disamenities, the residents' contributions to public goods, and perceptions of the local mtaa office. Finally, section 13 recorded respondents' own assessment of the current property value, and enumerators' assessment of building materials and maintenance condition.

A.4.6 Non-residential plots enumeration

Sampling frame

To enumerate the non-residential plots in the 20k project areas, we first referred to Town Planning Drawings collected by the Ministry of Lands, Housing and Human Settlement Development. Two research assistants transferred information on non-residential planned land uses from these georeferenced drawings to our shapefile of 20k plots. In total there were 1,562 plots with non-residential planned land uses, of which we enumerated 1,530 (98%).

Enumeration details and protocol

The data on non-residential plots were collected from June 2022 - February 2023, in parallel with the resident questionnaire described above. Each enumerator team received a map of non-residential plots in their respective areas. Accompanied by a local leader (mjumbe), they visited these plots and collected information on their actual use, state of maintenance, and ownership status. The information the enumerators collected drew on their own observations and information they gathered from others -primarily (but not exclusively) local leaders.

Enumeration questions

The enumerators determined whether each non-residential plot was fenced, currently used for residential activities, or currently used for non-residential activities either in its entirety or in part. For plots with non-residential activity, enumerators then sought to identify the specific use from a list of 16 pre-coded activities (outlined in table A.10). Finally, enumerators noted the maintenance condition of the plot (very well kept, reasonably well kept, abandoned, encroached by squatters), its ownership (e.g., government or public institutions, private individuals or firms), and the source(s) of information they (the enumerators) had used in (e.g., own observation, people who live or work in this plot, local leaders, or neighbors).

A.4.7 Enumeration of public transport nodes

Sampling frame

We conducted an enumeration of all the public transport access points (e.g., bus stops and others described below) available to the residents of the 20k project areas.

Enumeration details and protocol

This enumeration took place from December 2022 - February 2023. We started by asking a representative of the local government of each mtaa (typically the chairperson, who resides locally) to list all the public transport access points used by residents in their mtaa, including bus and minibus (daladala), auto rickshaw (bajaj), and motorcycle (bodaboda). If any of these three access modes was missing in a given mtaa, we asked about the nearest relevant point outside of the mtaa (i.e., the closest minibus collection point). Enumerators visited each access point and asked questions to drivers and passengers and recorded their findings using ODK.

Enumeration questions

Our objective was to enumerate the locations of all public transport access points (motorcycle, auto-rickshaw 'bajaj', bus, and minibus). For each access point with a bus or a minibus, we asked whether there is a direct route to Kariakoo (the most central location accessible by informal transport, beyond which only formal transport can enter the city center). Alternatively, we asked how many different buses (transfers) were required to reach Kariakoo. Furthermore, for any transport mode we asked: 'If a resident wanted to reach Kariakoo (the closest station) on a typical working day, how many [of given transport mode] would depart from here from 6am to 8am?", "If a resident managed to leave by 7am, how long would it take overall, from this station to the closest one in Kariakoo, taking the fastest route by [given transport mode]?", and "If this resident managed to leave by 7am, how much would he/she pay overall, from this station to the closest one in Kariakoo, taking the fastest route by [given transport mode]?".

B   Model details

B.1   The optimization problem

The optimization problem as described in the text is

$$
\max_{h_1, z_1, k, z_2, \tau} \int_0^\tau [\varphi ln h_1 + \beta ln z_1 + Ae^{-\theta t}]e^{-\rho t}dt + \int_\tau^\infty [\varphi ln(l^\alpha k^{1-\alpha}) + \beta ln z_2 + B]e^{-\rho t}dt
$$
$$
+ \omega \left(\int_0^\infty we^{-\delta t}dt - \int_0^\tau (ph_1 + z_1)e^{-\delta t}dt - \int_\tau^\infty z_2 e^{-\delta t}dt - rke^{-\delta \tau} - R(0) \right), \quad (4)
$$

where h_1 and z_1 are housing and all other goods consumed while in the city, z_2 is all other goods consumed upon moving to the 20k area, k is the amount of housing capital invested at the time of move and τ is the date of move. l is the given (for the moment) plot for sale at price $R(0)$ in time 0. r is the purchase cost of capital; z is the numeraire, and p is the rental price of housing in the city. We specify a constant wage, w. ρ is the personal discount rate and δ is the interest rate. We equate ρ and δ, so that z consumption is constant over the lifetime. A is the initial amenity level in the city at time 0 which declines at the rate θ and B is the amenity level in the city.

The first order conditions are

1. $\beta = \omega z_1 = \omega z_2, \ \rightarrow z_1 = z_2 \equiv z$

2. $h_1 = \frac{\varphi}{\omega p}, \ \rightarrow h_1 = \frac{\varphi z}{\beta p}$

3. $k = \frac{\varphi(1-\alpha)}{\omega r \delta}$

4. $\frac{w-z}{\delta} - \frac{ph_1}{\delta}(1 - e^{-\delta\tau}) = R(B, l_2) + rke^{-\delta\tau}$

5. $\varphi ln h_1 + Ae^{-\theta\tau} - \varphi ln(l_2^\alpha k^{1-\alpha}) - B + \omega[\delta rk - ph_1] = 0$

Substituting in the budget constraint (item 4) gives an expression for the multiplier: $\omega(\tau, R) = \frac{\beta+\varphi(1-\alpha e^{-\delta\tau})}{w-\delta R})$.

Through substitution using FOC and ω, we can derive useful expressions for τ and k:

$$ln(w - \delta R) = -\frac{1}{\alpha\varphi}[Ae^{-\theta\tau} - B - \varphi\alpha ln(\beta + \varphi(1 - \alpha e^{-\delta\tau}))$$
$$+ \varphi ln(\frac{\varphi}{p}) - \varphi\alpha lnl - \varphi(1 - \alpha)ln(\frac{\varphi(1 - \alpha)}{r\delta}) - \alpha\varphi] \tag{5}$$

$$k = \frac{\varphi(1 - \alpha)(w - \delta R)}{r\delta(\beta + \varphi(1 - \alpha e^{-\delta\tau}))} \tag{6}$$

B.2 An equilibrium with two plots sizes

If B and θ are the same throughout, there are two margins defining prices. One margin is for the lowest income person in 20k areas at $w = 4000$ who drives the price of small plots and the other is the lowest income person on a large plot set at $w = 70000$ who defines the price of a large plot. We need to solve for each of these prices.

The margin of entry to the 20k area

For the last entrant to the 20k area, we must equate the utility from staying in the city forever and in our problem always consuming the same h_1 and z_1 forever to the utility from moving. The utility from staying forever given $h_1 = \frac{\varphi w}{p(\beta+\varphi)}$ and $z_1 = \frac{\beta w}{\beta+\varphi}$ is defined as $U_1 = (\varphi ln\frac{\varphi w}{p(\beta+\varphi)} + \beta ln\frac{\beta w}{\beta+\varphi})/\rho + \frac{A}{\rho+\theta}$. The utility from staying comes from plugging back in the FOC's into the expression for the present value of utility in (4) including conditions for τ and the expression for ω. This yields $U_2 = Z/\rho + (1 - e^{-(\theta+\rho)\tau})\frac{A}{\theta+\rho} - ln(\beta + \varphi(1 - \alpha e^{-\delta\tau}))\frac{\beta+\varphi(1-\alpha e^{-\delta\tau})}{\rho} + ln(w - \delta R_1)\frac{\beta+\varphi(1-\alpha e^{-\delta\tau})}{\rho}$. Equating the utility from staying to that for leaving gives

$$ln(w - \delta R_1) = \frac{1}{\beta + \varphi(1 - \alpha e^{-\delta\tau})}\left[\varphi ln\frac{\varphi w}{p(\beta + \varphi)} + \beta ln\frac{\beta w}{\beta + \varphi} + \frac{A\rho}{\rho + \theta}e^{-(\theta+\rho)\tau} - Z\right]$$
$$+ ln(\beta + \varphi(1 - \alpha e^{-\delta\tau})) \tag{7}$$

where for convenience $Z \equiv \beta ln(\beta) + \varphi(1 - e^{-\delta\tau})ln(\frac{\varphi}{p}) + \varphi e^{-\delta\tau}(\alpha ln(l_2) + (1 - \alpha)ln(\frac{\varphi(1-\alpha)}{r\delta})) + e^{-\delta\tau}B$.

For this marginal lowest income person moving to the 20k area, we then have two equations (5 and 7) in 2 unknowns, R_1 and τ. A solution is depicted in Figure A.3 for our parameter values. The blue line plots (7) and the red (5). Note the they intersect at the maximal R this consumer is willing to pay to live in a 20k area. While (5) depicts the FOC for τ, (7) also reflects the condition for the τ that allows the maximal R based on the gap between U_2 and U_1, given U_1 is a fixed amount, exogenous to our problem.

The margin between a big and small plot

If $w(N_2)$ defines the income of the lowest income person on a large plot, to solve for what they are willing to pay for a large plot, we need to equate $U(l_1, R_1, w(N_2)) = U(l_2, R_2, w(N_2))$, where R_1 is defined above by equations 5 and 7. The utility from a small plot $U(l_1, R_1, w(N_2)) =$
$[\beta ln(\beta) + \varphi(1 - e^{-\delta\tau_1})ln(\varphi/p) + \varphi e^{-\delta\tau_1}(\alpha ln(l_1) + (1-\alpha)ln(\varphi(1-\alpha)/r\delta)) + e^{-\delta\tau}B]/\rho$
$+ ln\left(\frac{w(N_2) - \delta R_1}{\beta + \varphi(1 - \alpha e^{-\delta\tau_1})}\right)[\beta + \varphi(1 - \alpha e^{-\delta\tau_1})]/\rho + \frac{A}{\rho+\theta}(1 - e^{-(\theta+\rho)\tau_1})$. The utility from a large plot
$U(l_2, R_2, w(N_2)) = [\beta ln(\beta) + \varphi(1 - e^{-\delta\tau_2})ln(\varphi/p) + \varphi e^{-\delta\tau_2}(\alpha ln(l_2) + (1-\alpha)ln(\varphi(1-\alpha)/r\delta)) + e^{-\delta\tau}B]/\rho$
$+ ln\left(\frac{w(N_2) - \delta R_2}{\beta + \varphi(1 - \alpha e^{-\delta\tau_2})}\right)[\beta + \varphi(1 - \alpha e^{-\delta\tau_2})]/\rho + \frac{A}{\rho+\theta}(1 - e^{-(\theta+\rho)\tau_2})$, where τ_1 and τ_2 come from the application of eq (5) to the relevant plot size optimization problem. Equating and simplifying we get

$$
\begin{aligned}
0 =& [\alpha[e^{-\delta\tau_2}ln(l_2) - e^{-\delta\tau_1}ln(l_1)] + [e^{-\delta\tau_2} - e^{-\delta\tau_1}][(1-\alpha)ln(\varphi(1-\alpha)/r\delta) - ln(\varphi/p)]]\varphi/\rho \\
&+ [e^{-\delta\tau_2} - e^{-\delta\tau_1}]B/\rho + [ln\left(\frac{w(N_2) - \delta R_2}{\beta + \varphi(1 - \alpha e^{-\delta\tau_2})}\right)(\beta + \varphi(1 - \alpha e^{-\delta\tau_2})) \\
&- ln\left(\frac{w(N_2) - \delta R_1}{\beta + \varphi(1 - \alpha e^{-\delta\tau_1})}\right)(\beta + \varphi(1 - \alpha e^{-\delta\tau_1}))]/\rho \\
&+ \frac{A}{\rho + \theta}(e^{-(\theta+\rho)\tau_1} - e^{-(\theta+\rho)\tau_2}) \qquad\qquad\qquad\qquad\qquad (8)
\end{aligned}
$$

A solution is depicted in Figure A.4.

B.3   Solving equilibria when B's differ

As an example suppose some large plots have the base B_1 of 0.69 but some have a higher B_2, say, 0.76. Suppose the supply of B_2 large plots is such that the income cut-off is $w(B_2) = 12000$, assuming as in equilibrium that all better B plots are allocated to the highest income people. As before the price of a B_1 plot is set by the person marginal between small and large plots, where all small plots have B_1. Using the methodology in equation 8, we equate the utility from being on a large plot with B_1 and $w = 12000$ with known

66

price $R(l_2, B_1)$, $U(w(B_2), R(l_2, B_1), l_2)$, to the utility from being on a plot with a higher B, $U(w(B_2), R(l_2, B_2), l_2)$ to solve out $R(l_2, B_1)$, noting as always there are endogenous τ's involved which are given by the relevant applications of equation 5.

C   Data set construction

C.1   Insula construction

Insulae (singular insula) are contiguous groups of plots, which can be thought of as city blocks, defined by the planners of the 20k project. [34] Insulae are typically separated by roads, or in some cases by natural spaces that cannot be built on (e.g., streams). Insulae typically contain either residential plots or non-residential ones, but a few contain a mix of residential and non-residential plots.

We often characterize residential insula by their plot size, which is measured as the median size of residential plots within the insula. Residential insula can thus be classified by into three size groups where 'small' insula have a median plot size of less than 800sqm; 'medium' insula have a median plot size between 800sqm and 1600sqm; and 'large' insulae have a median plot size above 1600sqm. These classifications follow the official planning definitions of high, medium, and low density, where higher density corresponds to smaller plot sizes.

C.2   Super-insula construction

"Super-insulae" is a term that we (not the planners) define to group together insulae that are similar and close to each other. We create super-insula by aggregating spatially "proximate" residential insula of the same type of plot size (small, medium, and large). We treat any two insula as spatially "proximate" if a straight line can connect them across open space without intersecting any other insula.

Programmatically, we create such super-insula as follows. First, we define a grid (raster) of 1m x 1m cells, each classified as small, medium, large, non-residential, or open space. Second, we expand our set of small, medium, large, and non-residential cells by iteratively replacing any open-space cell by the class of its adjacent cell. This is a morphological operation called dilation, common in image processing. We continue this process until no open space remains. In the end, each set of contiguous cells becomes a super-insula with a unique classification (small, medium, large, or non-residential).

---

[34]We use the term insulae, since in Tanzania "blocks" refer to groups of nearby insulae. Insulae typically contain multiple plots, but some insulae may contain only a single plot.

Appendix References

Kironde, J. M. L. (2015), 'Good governance, efficiency and the provision of planned land for orderly development in african cities: The case of the 20,000 planned land plots project in dar es salaam, tanzania', **3**(4), 348–367.

Mero, C. S. (2008), Planning Influence on Urban Land Delivery Systems and Urban Land Markets: Challenges and Opportunities for Provision of Serviced Land in Urban Tanzania. The Case of "Dar es Salaam 20,000 Plots Project".

Mero, C. S. (2009), 'Towards Eradication of Unplanned Areas in Dar es Salaam city: Innovative Dar es Salaam 20,000 Plots Project- A success story', *Town Planners Newsletter, TAPT* (6).

Mwiga, B. G. (2011), Evaluating the Effectiveness of the Regulatory Framework in Providing Planned Land in Urban Areas: the case of Dar es Salaam city 20,000 plots project, Tanzania, PhD thesis, University of Twente, Enschede, Netherlands.

Nunn, N. and Puga, D. (2012), 'Ruggedness: The Blessing of Bad Geography in Africa', *Review of Economics and Statistics* **94**(1), 20–36.

OpenStreetMap contributors (2017), 'Planet dump retrieved from https://planet.osm.org on March 30, 2022', `https://www.openstreetmap.org`.

Statista (2022), 'Tanzania: Inflation rate from 1988 to 2027 retrieved on Oct. 9, 2022', `https://www.statista.com`.

United States Geological Survey (2000), 'SRTM1 Arc-Second Global'. DOI: /10.5066/F7PR7TFT.

D   Appendix Figures

Figure A.1: Historical examples of urban planning

(a) The urban plan of Miletus, Ancient Greece
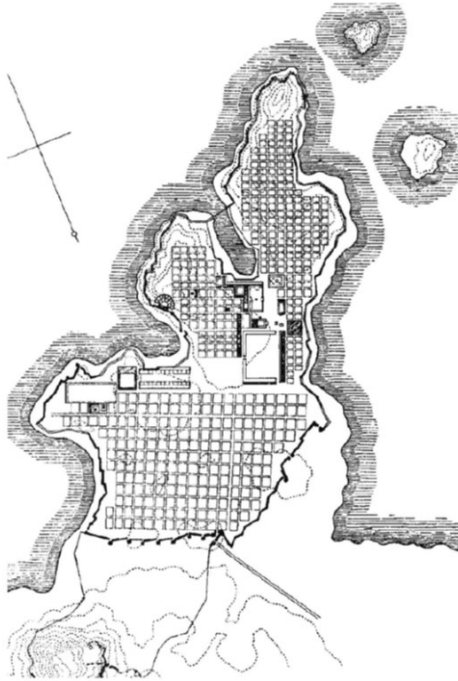
Figure 1. *Miletus.*[21]

(b) Diagram from Howard (1902) "Garden Cities of To-Morrow"
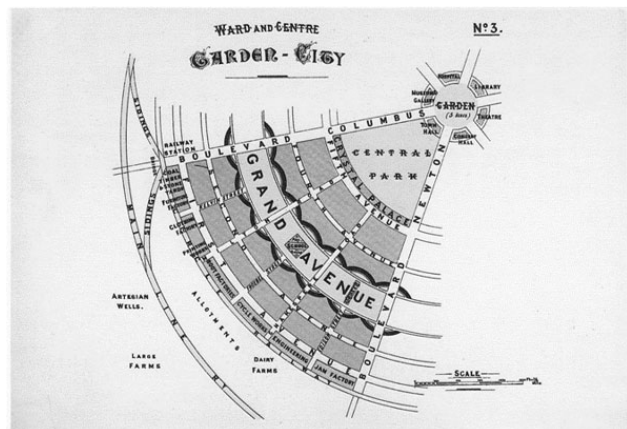
Notes: This figure plots historical examples of urban planning.

Figure A.2: Diagram of insula, plots, gridcells, and boundaries
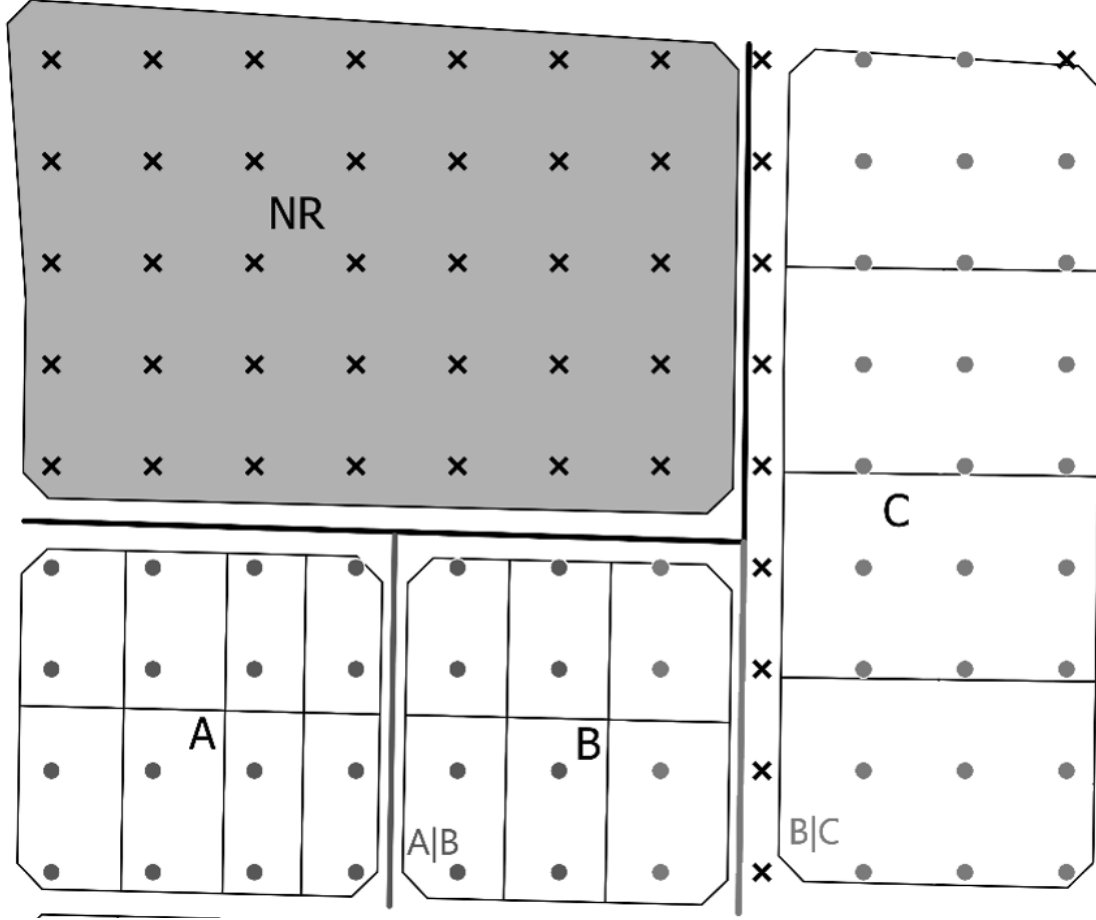
Notes: This figure provides a diagram demonstration of our data construction of insula, plots, gridcells, and boundaries. Plots are denoted by black outlines, and are colored white (residential), and grey (non-residential "NR"). Insulae are made up of the contiguous plots, each with a unique ID (A, B, C). Gridcell centroids spaced 20m apart and we take only cells with centroids that fall in plots (dots), ignoring cells that fall between ('x's). Boundaries fall equidistant between insulae, and we only use residential-residential boundaries (blue and red), ignoring non-residential boundaries (black). Gridcells are assigned based on the boundary that they are nearest to (blue to blue and red to red).

Figure A.3: R and τ for marginal entrant to 20K

Legend:
- $R(\tau)$ from EQ1-6
- $R(\tau)$ from EQ1-5&7

(6038, 15.0)

Land Price (R_0)

Optimal Time to Move (τ)

EQ: k^*=1447
Plot:
l_2=800, B=0.691
Residents: w=7000
Parameters:
α=0.3, β=0.7, φ=0.3
δ=0.06, r=16.7, $p1$=2.56
θ=0.010, A=1.05
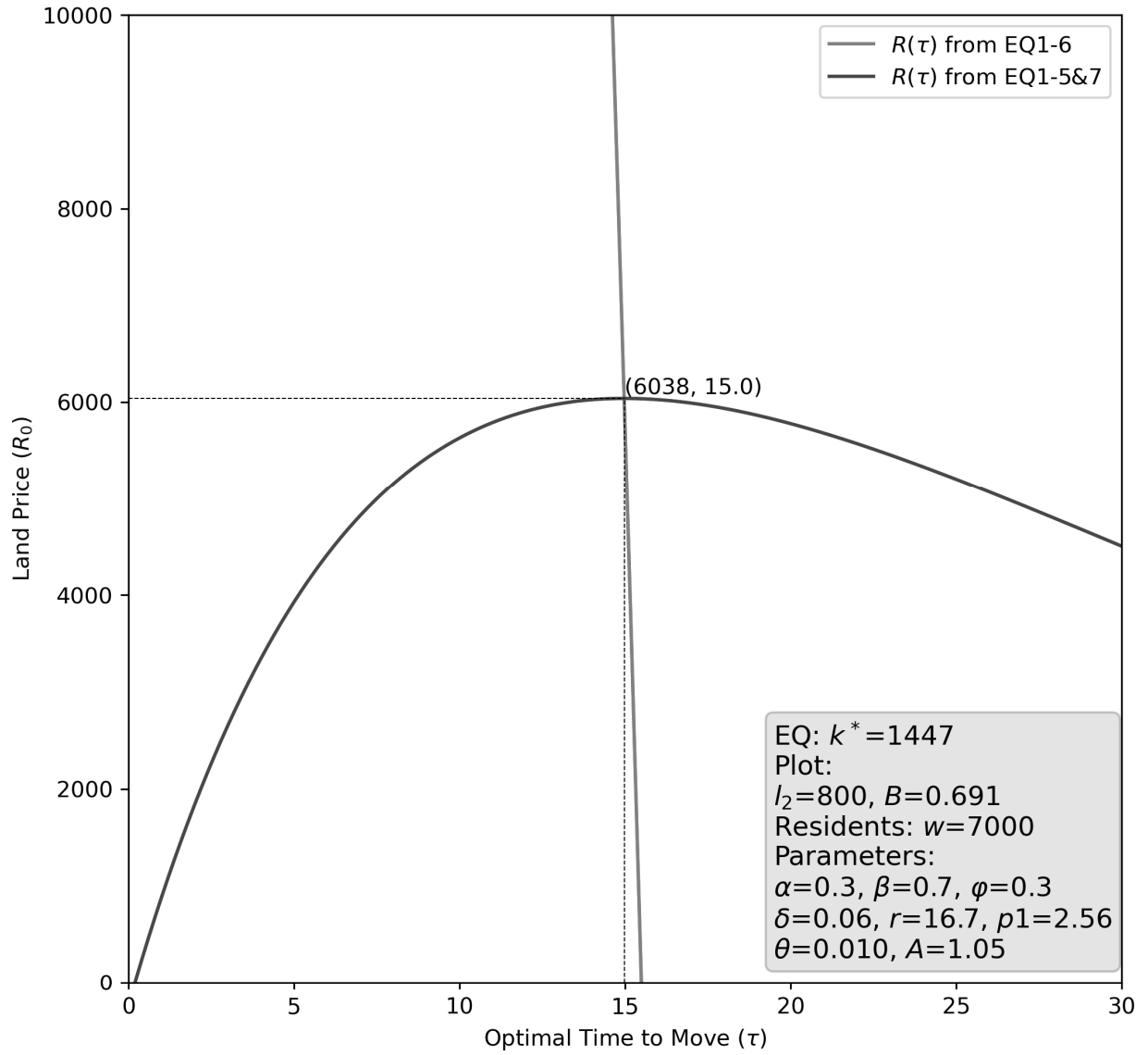
Notes: add figure notes.

Figure A.4: R and τ for margin between large and small plots

Legend:
- $R(\tau)$ from EQ1-6
- $R(\tau)$ from EQ1-5&8 with $R_1 = 5039$ and $\tau_1 = 15.08$

(8635, 8.26)

Plot:
$l_2=1600$, $l_1=800$, $B=0.691$
Residents:
$w_2=7000$, $w_1=4000$
Parameters:
$\alpha=0.3$, $\beta=0.7$, $\varphi=0.3$
$\delta=0.06$, $r=16.7$, $p1=2.56$
$\theta=0.010$, $A=1.05$

Land Price (R_0)

Time of Move (τ)
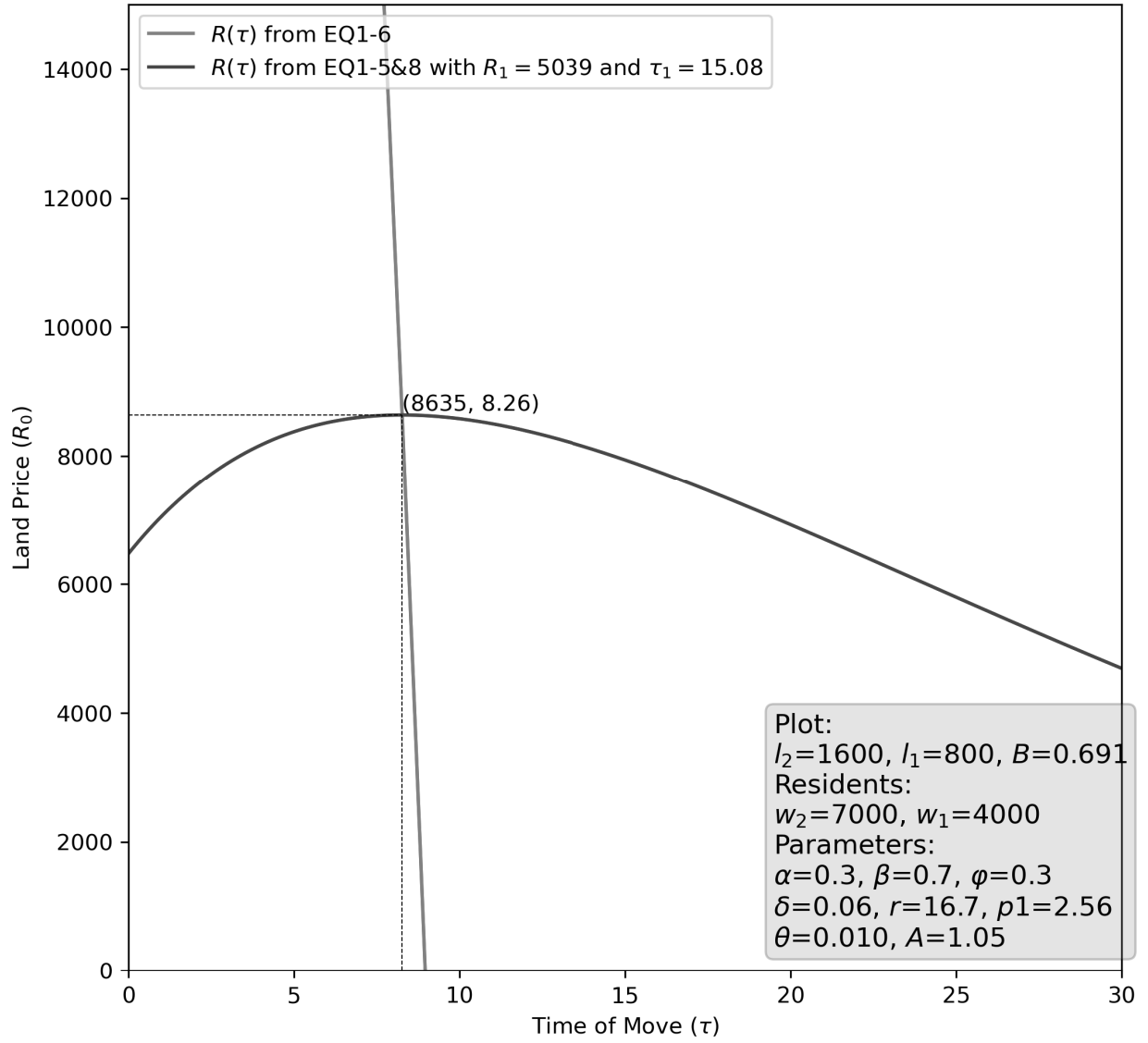
Notes: add figure notes.

E   Appendix Tables

Table A.1: OLS ln price per sqm on ln size & equivalent RD with ln size interaction (gridcell)

	(1) RD	(2) OLS	(3) RD	(4) OLS
Own Larger=1	-0.14		-0.08	
	(0.10)		(0.10)	
Own Larger=1 × ln GAP	-0.58		-1.08	
	(0.51)		(0.52)	
Ln plot size		-0.45		-0.47
		(0.07)		(0.08)
Mean Outcome	10	10	10	10
Period*Source FE	✓	✓	✓	✓
20k*EA FE	✓	✓	✓	✓
Amenities			✓	✓
Segment FEs	✓		✓	
Segment Sample	✓	✓	✓	✓
N (gridcells)	2449	2449	2243	2243
N (plots)	834	834	754	754

Note: This table presents OLS along with RD regressions across neighbouring insula boundaries restricting the sample to within 100m of the insula-pair boundary. The RD specification takes an indicator for whether a gridcell is in an insula with mean plot size larger than the nearest neighbouring insula, and always controls for linear distance to the boundary between insula pairs on each side of the boundary. The outcome is always the log price per square metre of a bareland transaction from thedalali or household questionnaires. Some controls are the same across columns: transaction period by source (dalali or household questionnaires) FEs and 20K*Enumeration Area FEs. Other controls vary across columns: amenities (cols 3-4), and insula-segment FEs (cols 1&3). Columns (cols 2&4) use the same samples as columns (cols 1&3). Amenities include average elevation, average ruggedness, and dummies for within 100m of: river, wetland, planned open space, planned school, planned non-resdential use, and large (20m+) planned road within 10m. Note that in this specification, the dummy for wetland within 100m is co-linear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table A.2: OLS built outcomes on ln size & equivalent RD with ln size interaction (gridcell)

	(1) share gridcell built (RD)	(2) share gridcell built (OLS)	(3) share gridcell built (RD)	(4) share gridcell built (OLS)	(5) plot is built (RD)	(6) plot is built (OLS)	(7) plot is built (RD)	(8) plot is built (OLS)
Own Larger=1	0.00 (0.00)		0.00 (0.00)		0.02 (0.01)		0.02 (0.01)	
Own Larger=1 × ln GAP	-0.08 (0.01)		-0.08 (0.01)		-0.03 (0.02)		-0.03 (0.02)	
Ln plot size		-0.09 (0.00)		-0.08 (0.00)		-0.04 (0.01)		-0.01 (0.01)
Mean Outcome	0.11	0.11	0.11	0.11	0.49	0.49	0.49	0.49
20k*MTAA FE	✓	✓	✓	✓	✓	✓	✓	✓
Amenities				✓				✓
Segment FEs	✓		✓		✓		✓	
N (gridcells)	87569	87789	87569	87789	87569	87789	87569	87789
N (plots)	33613	33648	33613	33648	33613	33648	33613	33648

Note: This table presents OLS and RD regressions across neighbouring insula boundaries restricting the sample to within 100m of the insula-pair boundary. The RD specification takes an indicator for whether a gridcell is in an insula with mean plot size larger than the nearest neighbouring insula, and always controls for linear distance to the boundary between insula pairs on each side of the boundary. The mean distance to the boundary is 31m, median 28m, 75th percentile 41m, and 95th percentile 66m. The outcomes vary across columns. is always the log price per square metre of a bareland transaction from the dalali or household survey. Controls always include 20K*MTAA Area FEs, amenities, and insula-segment FEs. Amenities include average elevation, average ruggedness, a three-way Z-Index of insula characteristics (rectangularity, alignment, regularity), and dummies for within 100m of: river, wetland, planned recreation, planned nursery school, planned religious site, planned education, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential use. Standard errors in parentheses are clustered by insula.

Table A.3: Population density by plot size

	Mean Pop. per built residential plot	Share of Plots Built	Mean Pop. per residential plot	Mean Plot Size (sqm)	Pop. dens. residential plots (sqkm)	Pop. dens. (sqkm)
Small (<800sqm)	5.3	0.50	2.6	629	4167	2083
Medium (800-1600sqm)	5.4	0.49	2.6	1179	2233	1116
Large (>1600sqm)	5.6	0.49	2.7	1965	1391	696
All Plots	5.4	0.49	2.7	1041	2551	1276

Note: This table presents population statistics by plot size. All statistics are calculated over residential plots in 20K areas. The first column is the average number of residents on built plots from the household questionnaire. The second column is the share of plots built, and the third column is the average number of residents per residential plot including unbuilt plots. We assume that unsurveyed, but built, plots have the same average number of residents as surveyed built plots, and further, than unbuilt plots have zero population. The fourth column is the average size of residential plots. The fifth column is population density on residential plots. The sixth column is population density rescaling for non-residential land (50% of all land).

Table A.4: Effects of ln plot size and share neighbors small (gridcell level)

	(1)	(2)	(3)	(4)	(5)	(6)
		Share	Plot	Log	Size of	Multiple
	Ln Price	gridcell	is	area of	building	buildings
		built	built	buildings	on plot	on plot
Ln plot size	0.41	-0.053	0.026	0.16	15.2	0.10
	(0.13)	(0.0037)	(0.016)	(0.032)	(4.47)	(0.014)
Shr. Neighbs Small	0.16	0.28	0.36	0.30	59.8	0.14
	(1.51)	(0.052)	(0.21)	(0.48)	(56.4)	(0.16)
Ln plot size × Shr. Neighbs Small	-0.059	-0.037	-0.050	-0.042	-8.13	-0.020
	(0.22)	(0.0077)	(0.030)	(0.071)	(8.42)	(0.024)
Mean Outcome	17	0.11	0.49	5.3	96	0.19
Period*Source FE	✓					
20k*MTAA FE	✓	✓	✓	✓	✓	✓
Amenities	✓	✓	✓	✓	✓	✓
N (gridcells)	2885	94785	94785	46464	94785	94785
N (plots)	998	36215	36215	17822	36215	36215

Note: This table presents regressions of log price on log plot size and the fraction of small neighbouring plots (within 100m). The outcomes include: log price (col 1), share of gridcell built (col 2), an indicator for plot built (col 3), log of own plot area intersected with buildings greater than 30 sqm (col 4), size of the largest building on the plot where missing values are treated as zeros (col 5), and an indicator for multiple buildings on the plot (col 6). Controls vary across columns: transaction period by source (Dalali or Occupier survey) FEs (col 1), 20K*MTAA Area FEs (cols 1-6), and amenities (col 1-6). Amenities include average elevation, average ruggedness, and dummies for within 100m of: river, wetland, planned recreation, planned nursery school, planned road religious site, planned service trade, planned housing estate, planned public building, planned cemetery, and planned any other non-residential land use. Note that in the col 1 specification, the dummy for wetland within 100m is co-linear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table A.5: Land values & housing development by planned amenities (gridcell level)

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Size of building on plot	(6) Multiple buildings on plot
Ln plot size	0.51	-0.078	-0.000055	0.14	10.5	0.092
	(0.076)	(0.0026)	(0.0094)	(0.018)	(2.63)	(0.0072)
Elevation (m)	0.0029	0.00088	0.0027	0.0031	0.68	0.00095
	(0.0027)	(0.000098)	(0.00043)	(0.00068)	(0.098)	(0.00033)
Ruggedness	-0.021	-0.0058	-0.016	-0.011	-3.82	-0.0079
	(0.031)	(0.00098)	(0.0039)	(0.0089)	(0.89)	(0.0031)
River/stream 100m	-0.0079	-0.027	-0.11	-0.073	-26.6	-0.056
	(0.17)	(0.0051)	(0.022)	(0.057)	(4.61)	(0.019)
Water/wetland 100m	0	0.0096	-0.067	-0.070	-14.3	-0.036
	(.)	(0.0092)	(0.031)	(0.16)	(6.11)	(0.037)
Dist (km) paved major road	-0.13	-0.015	-0.041	-0.062	-12.5	-0.031
	(0.038)	(0.0016)	(0.0070)	(0.012)	(1.73)	(0.0053)
Insula rectangularity (standardized)	-0.037	0.0030	0.017	0.0057	3.82	0.0083
	(0.035)	(0.0013)	(0.0054)	(0.010)	(1.38)	(0.0042)
Insula alignment (standardized)	0.015	0.0025	0.0035	0.012	1.42	0.0013
	(0.029)	(0.00095)	(0.0041)	(0.0080)	(0.98)	(0.0032)
Insula regularity (standardized)	0.049	-0.0017	-0.0046	-0.0075	-0.82	-0.0042
	(0.035)	(0.0012)	(0.0048)	(0.0094)	(1.35)	(0.0039)
Pln. recreation in 100m	0.0019	-0.00059	-0.0084	-0.0076	-2.14	-0.0066
	(0.053)	(0.0019)	(0.0071)	(0.012)	(1.78)	(0.0053)
Pln. nursery school in 100m	0.060	0.0064	0.018	0.031	5.21	0.010
	(0.055)	(0.0026)	(0.0097)	(0.017)	(2.33)	(0.0071)
Pln. religious site in 100m	0.051	0.0024	0.016	-0.0051	3.60	-0.00088
	(0.068)	(0.0030)	(0.012)	(0.020)	(3.41)	(0.0086)
Pln. education in 100m	0.15	-0.0044	-0.0080	-0.022	-4.37	-0.0078
	(0.085)	(0.0030)	(0.011)	(0.021)	(2.68)	(0.0083)
Pln. service trade in 100m	-0.052	-0.0011	-0.0027	-0.0041	0.57	-0.0039
	(0.12)	(0.0043)	(0.016)	(0.030)	(3.63)	(0.011)
Pln. housing estate in 100m	-0.14	0.00091	0.0098	0.0045	2.53	-0.017
	(0.12)	(0.0075)	(0.030)	(0.048)	(7.36)	(0.023)
Pln. public building in 100m	-0.00078	-0.0052	-0.0090	-0.043	-4.62	-0.017
	(0.11)	(0.0044)	(0.016)	(0.029)	(3.62)	(0.012)
Pln. cemetery in 100m	0.066	0.0045	0.040	-0.044	4.88	0.017
	(0.15)	(0.0052)	(0.019)	(0.034)	(4.48)	(0.015)
Pln. any other non-res in 100m	0.13	-0.0022	-0.021	0.0067	-4.11	0.0025
	(0.099)	(0.0031)	(0.012)	(0.023)	(2.65)	(0.0083)
Mean Outcome	17	0.11	0.49	5.3	96	0.19
Period*Source FEs	✓					
20k*MTAA FEs	✓	✓	✓	✓	✓	✓
N (gridcells)	2885	94789	94789	46465	94789	94789
N (plots)	998	36215	36215	17822	36215	36215

Note: This table presents regressions of both price and built outcomes on log plot size. The outcomes vary across columns: This table presents regressions of the following outcomes: log price (col 1), share of gridcell built (col 2), an indicator for plot built (col 3), log of own plot area intersected with buildings greater than 30 sqm (col 4), size of the largest building on the plot where missing values are treated as zeros (col 5), and an indicator for multiple buildings on the plot (col 6). Controls vary across columns: transaction period by source (dalali or household questionnaires) FEs (col 1), and 20K*MTAA Area FEs (cols 1-6). All specifications show coefficients for amenities, which include average elevation, average ruggedness, distance to paved major road, and standardized measures of insula characteristics: rectangularity, regularity, and alignment. It includes dummies indicating within 100m of a river and wetland. Each observation is a gridcell. Standard errors in parentheses are clustered by insula.

Table A.6: Housing development by implemented amenities (gridcell level)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Size of building on plot	(5) Multiple buildings on plot
Ln plot size	-0.077	0.0015	0.14	11.2	0.092
	(0.0025)	(0.0093)	(0.018)	(2.55)	(0.0072)
Elevation (m)	0.00088	0.0027	0.0031	0.67	0.00095
	(0.000097)	(0.00043)	(0.00067)	(0.098)	(0.00032)
Ruggedness	-0.0057	-0.016	-0.011	-3.73	-0.0080
	(0.00097)	(0.0038)	(0.0089)	(0.89)	(0.0031)
River/stream 100m	-0.028	-0.12	-0.068	-26.7	-0.057
	(0.0052)	(0.022)	(0.058)	(4.61)	(0.019)
Water/wetland 100m	0.0058	-0.074	-0.090	-16.7	-0.040
	(0.0088)	(0.031)	(0.16)	(5.97)	(0.037)
Dist (km) paved major road	-0.015	-0.041	-0.064	-12.6	-0.032
	(0.0016)	(0.0071)	(0.011)	(1.68)	(0.0053)
Z-index: 3 Ins. Characteristics	0.0038	0.016	0.011	4.42	0.0051
	(0.0013)	(0.0058)	(0.0100)	(1.42)	(0.0041)
Impl. as planned in 100m	0.0085	0.029	0.014	8.66	0.012
	(0.0028)	(0.010)	(0.016)	(2.83)	(0.0083)
Mean Outcome	0.11	0.49	5.3	96	0.19
20k*MTAA FEs	✓	✓	✓	✓	✓
N (gridcells)	94789	94789	46465	94789	94789
N (plots)	36215	36215	17822	36215	36215

Note: This table presents regressions of both price and built outcomes on log plot size. The outcomes vary across columns: This table presents regressions of the following outcomes: log price (col 1), share of gridcell built (col 2), an indicator for plot built (col 3), log of own plot area intersected with buildings greater than 30 sqm (col 4), size of the largest building on the plot where missing values are treated as zeros (col 5), and an indicator for multiple buildings on the plot (col 6). Controls vary across columns: transaction period by source (dalali or household questionnaires) FEs (col 1), and 20K*MTAA Area FEs (cols 1-6). All specifications show coefficients for amenities, which include average elevation, average ruggedness, distance to paved major road, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies indicating within 100m of the following: river, wetland, and implemented into planned non-residential land use. Each observation is a gridcell. Standard errors in parentheses are clustered by insula.

Table A.7: sorting by years of schooling within 20k areas

	(1)	(2)	(3)
Panel A: Landowners only			
Ln plot size	0.73		
	(0.36)		
Ln property value		1.19	
		(0.18)	
Ln Price			0.69
			(0.26)
Mean Outcome	14	14	14
Period*Source FE			✓
20k*MTAA FE	✓	✓	✓
Amenities	✓	✓	✓
N (gridcells)	5018	4085	1026
N (plots)	1648	1351	338
Panel B: Usufructuary and Tenants			
Ln plot size	-0.57		
	(0.30)		
Ln property value		0.83	
		(0.16)	
Mean Outcome	9.1	9.4	
Period*Source FE			
20k*MTAA FE	✓	✓	
Amenities	✓	✓	
N (gridcells)	5134	3594	
N (plots)	1562	1117	

Note: This table presents regressions of education of household heads on log plot size (col 1), estimated property value (col 2), and log plot price (col 3). The outcome is always the number of schooling years completed by the head of household from the occupier survey. Controls vary across columns: 20K*MTAA Area FEs and amenities (cols 1-3), transaction period by source (Dalali or Occupier survey) (col 2). Amenities include average elevation, average ruggedness, and dummies for within 100m of: river, wetland, and planned non-residential land uses. Panel A is uses only household heads who own the plot, and panel B uses household heads that are either usufructuary or tenants on the plot. Standard errors in parentheses are clustered by insula.

Table A.8: Project data source by 20K area

Project Area	Data source
Bunju	Survey map
Buyuni	Survey map
Kibada	Survey map
Kijichi	Survey map
Kisota	Survey map
Mbweni JKT	Survey map
Mbweni Malindi	Survey map
Mbweni Mpiji	Survey map
Mivumoni	Survey map
Mwanagati	Survey map
Mwongozo	Cadastre (survey map registered from gov)
Tuangoma	Survey map

Table A.9: Initial government prices by 20K area

Project Area	Initial price (TSh/sqm)
Bunju	1,760
Buyuni	1,056
Kisota	1,120
Mbweni JKT	1,920
Mbweni Mpiji	1,632
Mivumoni	1,344
Mtoni Kijichi	1,280
Mwanagati	704
Mwongozo-Gezaulole (Dungu Farm)	1,920
Tuangoma	800
Kibada	1,500
Mbweni Malindi	.

Table A.10: Categories of landuse from non-residential plot enumeration

Code	Landuse
1	OPEN SPACE
2	FARMING
3	LIVESTOCK
4	PUBLIC BUILDING
5	NURSERY SCHOOL
6	PRIMARY SCHOOL
7	PLAY GROUND
8	CEMETERY
9	RELIGIOUS SITE
10	SERVICE TRADE (including any business)
11	MARKET
12	DISPENSARY
13	HOTEL SITE
14	PARKING AREA
15	HOUSING ESTATE
16	INFORMAL SECTOR
99	OTHER (specifiy)
999	DON'T KNOW