Evaluating Urban Planning: Evidence from Dar es Salaam*

Vernon Henderson (*LSE*) Francisco Libano-Monteiro (*LSE*)
Martina Manara (*Sheffield*) Guy Michaels (*LSE*) Tanner Regan (*GWU*)

April 17, 2024

PRELIMINARY AND INCOMPLETE

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. We study the consequences of such planning decisions in 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. We study this project using detailed maps, questionnaires, and satellite imagery, and we combine within-neighborhood analysis and spatial regression discontinuity designs. We find that overall, the project secured property rights and access, thus boosting land values, and attracting highly educated owners; small plots, which command higher land values and are built more intensively, are under-provided; access to main paved roads is prized; and development rates are higher where plot layout is more gridded and small plots are bunched together. But planned non-residential amenities are ignored due to low implementation rates and about half the plots are still unbuilt, suggesting that despite the project's success, significant improvements are possible.

KEYWORDS: Urban Planning, Economic Development, Africa.

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

---

*Corresponding author: Michaels, g.michaels@lse.ac.uk. We thank our field managers, Erick Makori and Prosper Kaigarula and our team of enumerators for their excellent fieldwork and Claire Coker for excellent research assistance. For helpful comments we thank our discussants, Simon Franklin and Maisy Wong, and many colleagues, especially Shlomo Angel, Joseph Kironde, Sarah Kyessi, and Clemence Mero, and seminar and conference participants at the University of Chicago, Penn-Wharton, Galway, Bank of Italy, USC, UQAM, the Online Spatial and Urban Seminar, and the 8th Urbanization and Poverty Reduction Research Conference. We gratefully acknowledge financial support from the International Growth Centre grant TZA-23006, the Wheeler Institute for Business, and the ESRC's Centre for Economic Performance. Research approvals: LSE Research Ethics Committee and the Tanzanian Commission for Science and Technology (COSTECH).

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 is regulated by zoning.[1] But in developing countries, and especially in Africa's large and growing cities, much of the land is unplanned or ineffectively regulated.(e.g., Castells-Quintana, 2017; Henderson, Regan and Venables, 2021). The resulting informal settlements may reduce private investments, lower tax bases, and exacerbate urban disamenities (Scruggs, 2015; UN-Habitat, 2013). Therefore, projects that offer effective planning for portions of developing country cities 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 from Tanzania indicates that the de novo approach was cost effective and promoted higher quality housing an dland values in the long run than in neighboring laissez-faire informal developments or slums that were upgraded ex-post (Michaels et al., 2021). However, we know little about the *economic* consequences of the choices made by the de novo planners, including the allocation of land for private uses (e.g., the sizes and layout of residential and commercial plots) and 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). This paper starts to fill the gap by asking: how does de novo layout affects housing outcomes?

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 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

---

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

issues. While urban planners and economists could learn from each other about how to improve city design, such mutual learning is as yet limited (e.g., Bertaud (2018)). 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).

To shed light on the consequences of de novo urban planning, we develop a model of planning choices, outcomes, and residential sorting. We complement this model with an empirical investigation of the effects of planners' choices on residential plots' size and their layout relative to each other and on non-residential land uses, including roads and public and commercial uses. The outcomes we study include bare land values measured using novel data on 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 utilize detailed project maps, which allow us to run OLS regressions using within-neighborhood variation and spatial regression discontinuity design.

The context of our investigation is the "20,000 Plots" project, which the Tanzanian government implemented on the fringes of Dar es Salaam from around 2000-2010.[3] This project delivered around 36,000 residential de novo plots (almost twice the number initially intended) in 12 project areas, 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, as evidenced both by our regression analysis and by estimates of local leaders.

The local leaders indicated that the 20k plots' price premium reflects more secure individual property rights and better access, and our empirical evidence supports their claims. First, within non-20k areas near 20k areas, the minority of plots that were formalized (and are thus more secure) are costlier than the informal plots, though still cheaper than in 20k areas. Second, while local unpaved road access is ubiquitous in 20k areas (but not outside them), proximity to preexisting main paved roads within 20k boosts land values.

Preserving private property rights and access, while valuable, seems to entail a rigidity, where plot size and layout cannot be readily reconfigured to accommodate market needs. This rigidity may reflect limited administrative and technological capacity to implement changes without undermining property rights or access.[4] It is therefore important to investigate whether the planned layout is efficient or can be improved - and if so how.

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

---

[3] For brevity, we often refer to the "20,000 Plot Project" as the "20k Project" or simply "20k".

[4] In contrast, local leaders allow changes in layout and size of informal plots.

open space, attract better educated owners, and have lower population densities. These results are consistent with an over-provision of large plots, probably due to colonial-era planning rules, which have largely persisted until the present day.

We also find evidence that when small plots (though not larger ones) are bunched together they are more likely to be built on and are built on more intensively, suggesting a potential role for segmentation of areas into blocks of homogeneous plot sizes. Other planning decisions also matter, as plots are built on more and more intensively when residential "blocks" (which we refer to as "insulae") are more rectangular and follow a more gridded alignment.

Finally, we find that construction rates and intensity are consistent with positive valuation of a natural amenity (elevation) and negative valuation of natural disamenities (ruggedness as well as proximity to rivers, streams, water and wetlands) in a city prone to flooding. In contrast to these (and to the proximity to main roads), most planned amenities are not valued, most lively because (other than roads), they have not yet been provided, more than a decade after the project was completed and its plots were sold.

Besides the research noted above, our paper is related to 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 also recent work on land-use regulation differences at boundaries (Turner et al., 2014; Kulka, 2019; Chiumenti et al., 2022; Shertzer et al., 2018), to which we contribute by studying specific variation in plot sizes and configurations (rather than variation in bundled regulations), and by studying de novo projects. 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 strand of 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 a suburban settings, where large tracts of land are abundant. 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 on 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 de novo planning. Finally, 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 features (Lawler, 2008; Smith, 2007), as did some ancient cities in Mesopotamia, Assyria, and Egypt (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 and public spaces with multiple public buildings. Gridded cities spread across 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 urban planning is studied in hundreds of graduate programs worldwide (Symonds, 2023). Despite the rich history of urban planning, however, systematic economic evaluations of it are scant (Bertaud, 2018), especially for de novo planning.

2.2 Urban planning in Dar es Salaam

In Dar es Salaam, under German and later British colonial rule, planning and building standards differed, to facilitate residential segregation: the European core had a grid and strict planning standards with large plots; the Asian parts had lower standards but were still planned; and the African parts were unplanned (Kironde, 1994; MLHHSD, 2018). 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. Formal planning standards were retained in theory, sometimes with new justifications (Kironde, 1994), and a series of masterplans were developed (MLHHSD, 2018). In practice, however, most of the city comprises of informal settlements and, even in formal neighborhoods, plots and building footprints often violate zoning regulations (MLHHSD, 2018).

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, however, recent evidence, shows that the de novo approach resulted in better housing quality and price premia (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

Our study focuses on the "20,000 Plots" project, which the Tanzanian government initiated in late 1990s in response to perceived unmet demand for formal de novo plots (Tiba et al., 2005). This project, which was implemented from 2000-2010, delivered around 36,000 residential plots in 12 neighborhoods spanning a total of 75 square km area. The residential plots took up about half the project area (~38 square km) and were formally surveyed and titled. Around 1,500 additional plots (spanning ~12 square km in total) were designated for non-residential public and commercial uses. The remaining area (~25 square km) was taken up by roads and shoulders, almost exclusively unpaved, and by land deemed hazardous (e.g., in streams or water bodies or adjacent to them) which was left empty.[5]

Figure 1 contains maps of the project areas, which were mostly near the fringes of Dar es Salaam. Like the Sites and Services projects in Dar es Salaam from 40-50 years ago, the expectation is that as the city grows, these locations will no longer be on the fringe. The maps also show the pre-existing main paved roads and the boundary of the Dar es Salaam metropolitan area. The government set 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).[6] The maps also show that compared to the government

---

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

[6] We include in our analysis one area, Malindi, which was developed from 1998 and later integrated into the 20k 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.

sale prices, plot prices appreciated rapidly in all project areas, though not uniformly so.

Of the two aforementioned 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 ~ 30 million USD 2021, or ~ 0.8 per sqm of residential plot. 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 was recouped (Tiba et al., 2005).

But the second limitation of Sites and Services, that they did not cater to the poor, was not addressed. 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).[7] But the speed at which the plots were sold to repay the above-mentioned loan effectively barred most potential buyers.[8] Another impediment to the purchase of de novo plots by poorer Tanzanians was their large size and resulting high cost, as the government price per square meter within each project area was fixed. 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.[9]

The constraints outlined above resulted in the initial allocation of plots to many lucky or connected owners, including (anecdotally) many government officials. Those who bought benefited from rapid price appreciation, and often resold excess plots at market prices to those who actually wanted to build in 20k areas.[10]

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.

---

[7] Part of this cost reduction is thanks to Global Positioning System (GPS) technology (Tiba et al., 2005).

[8] 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 poor's limited access to credit, they found it difficult to purchase plots.

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

[10] In the questionnaire we administered, the majority of 20k owners reported some additional land holdings - not necessarily in planned areas. But a large majority (89 percent) said that their household has no additional 20k plot in their mtaa. This motivates our focus below on owners of single 20k plots for their own use.

We obtained maps covering all the project areas in Dar es Salaam, including town planning drawings and survey maps, which cover eleven of the twelve project areas, and cadaster data that cover the remaining area, as we further discuss in Appendix Section A.1.

We also obtained color satellite images with a resolution of ~ 0.5 meters, which cover the project areas and a buffer of 500 meters around them. We mostly use images from 2019-2021, but to understand initial conditions we also use older images. We paid a Nairobi-based firm, Ramani Geosystems, to digitize information from these images, including the footprints of buildings. We provide more details on these images and how we used them in Appendix A.2.

We measure underlying locational fundamentals using 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 mostly reflect 'first-nature" differences across locations, project development may have altered them slightly (e.g., if some land was flattened). These data are detailed in Appendix section A.3.

We also collected additional data on the 20k projects. First, we held interviews with (i) local experts and (ii) leaders of 34 local administrative areas ('mitaa' in Swahili, singular 'mtaa'), 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 sales dates and prices for individual plots in 20k areas and nearby non-20k areas and (ii) residents in over 3,200 plots 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 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. [11] 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, and are typically separated by roads (not shown on this version of the plan).[12] The plan also shows non-residential insulae

---

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

[12] We use the term insulae (singular - insula) to describe sets of contiguous (planned) plots, following the common usage in Roman residential terminology (Storey, 2004), and avoid the term "blocks", which in Tanzania refers to a number of adjacent 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 (e.g., some may be more elevated). The parcels may be assigned to different treatments by the planners, who may allocate them to residential plots of different sizes or with different proximity to amenities. Our empirical methodology, outlined 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, corresponding to the minimum formal plot (400 square meters). We typically identify each gridcell with its centroid and relate it to the plot and the insula in which this centroid falls. We focus on the approximately 95,000 gridcells whose centroids fall inside residential plots. Figure 4 shows the size distribution of these residential plots, 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 even by developed country standards, exceeding 1000 sq meters (approximately 1/4 acre), despite the low levels of income in Tanzania.

We also define "super-insula" areas as sets of contiguous insulae, whose mean sizes are relatively uniform, corresponding to one of the three size categories above (small, medium, and large).

The main outcomes we study are related to the model in Section 4, including the real price of plots that were unbuilt ("bare land") when they were sold.[13] Such prices are available for 1,446 residential plots (1,122 from the real estate agents and the rest from the residents).

Other important outcomes include measures of housing investment taken from the satellite imagery, and thus available for all the gridcells. First is a reduced-form measure of construction intensity - the share of a gridcell (restricted to residential plots within the gridcell's insula) that is built. Second is a measure of extensive margin housing construction - an indicator for the gridcell's plot containing the centroid of at least one building whose footprint is at least 30 sqm. Finally, to capture housing capital intensity we use two intensive margin measures: (i) the logarithm of the total footprint of (up to) three largest buildings on the gridcell's plot; and (ii) an indicator for multiple buildings in the gridcell's plot.

Another model outcome is the income of plot owners. Our best measure for this is years of schooling from the household questionnaires. We prefer this to current household income, since (i) it better captures lifetime income, as it is not affected by life-cycle and transitory elements and (ii) we had better response rates for the schooling question.[14]

Our main regressors of interest are the logarithm of plot size and measures of amenities - preexisting, planned, and implemented - at the gridcell level. These include the distance to the nearest preexisting main paved road; elevation; ruggedness; indicators for being within 100m of rivers (including streams) or of water (including wetland); Indicators for the gridcell being within 100m of planned amenities: open space, nursery school, religious site, education (primary or secondary school), service trade, housing estate, public buildings, cemetery, or other; and a Z-index of insula characteristics: (homogeneity of plot sizes within the insula; rectangularity, which measures the similarity of its shape to a rectangle; and alignment, which measures how aligned it is with its nearest neighboring residential insula).

Key controls include fixed effects (f.e.) for project areas and for mitaa, which we typically interact with each other. In price regressions we also control for f.e. for the time period of sale interacted with source (real estate agents or residents). Finally, in regression discontinuity (RD) specifications we control for the distance from the gridcell to the boundary between its insula and that of the neighboring insula and sometimes indicators for the nearest boundary segment, as illustrated in Appendix Figure A.2. We similarly define boundaries between "super insulae". The construction of these and the rest of the variables we use is discussed in detail in the appendix. The main variables are also described in Table A.1 and their summary statistics are reported in Table A.2.

---

[13] We inflate historical prices up to the year 2021 using annual inflation rates all in Tanzanian Shillings as detailed in Appendix A.3.

[14] Mincer regressions indicate that years of schooling account for much variation in current income. But years of schooling are still only a rough proxy for lifetime income, so cannot be used to hold income fixed in the regressions.

4 Modeling the development of 20,000 plot areas

De novo planners have considerable discretion in how to allocate greenfield land between private and public uses. The model focuses on two key choices of planners: plot sizes and amenities.[15] In choosing these, planners may try to meet different objectives, such as increasing land values (e.g., Turner et al., 2014), which may in turn affect potential tax bases (e.g., Besley and Persson, 2014)); ensuring that formal plots are built, an issue raised by the Tanzanian Minister for Lands, Housing and Human Settlements Development (Jamal, 2018); or increasing intensive-margin development (e.g., to increase agglomeration benefits, see (e.g., Henderson, Nigmatulina and Kriticos, 2021)); and considering distributional consequences, such as widening access to ownership. The model outlined here considers each of these outcomes: land values; whether plots are built upon and when; the intensive margin of construction – housing capital; and the sorting of owners with different income levels.

At the model's starting point (time 0, which is around 2005), people live in the city (Dar es Salaam) and are offered opportunities to buy plot of different sizes (l) from the initial owners. These initial owners were the lucky (or connected) ones who purchased underpriced plots from the government before time zero (e.g., 2000). We focus on those who buy the 20k plots from their initial owners and move into these 20k plots at a future time τ, when they invest a one-time irreversible k in housing capital on their plot.

The people who consider leaving the city to live in a 20k area have incomes distributed between $[\underline{w}, \bar{w}]$. The 20k plots are scarce, and plots are purchased by the higher income people in equilibrium - those with incomes from $[w_m, \bar{w}]$, where $\underline{w} < w_m$. The limited supply of 20k plots therefore prevents poorer potential owners from buying in the 20k area.

Would-be buyers face fixed supplies of plots of different sizes. Each set of plots of the same size is endowed with a distribution of amenities. To characterize the equilibrium, we begin by outlining some analytical results, after which we discuss a concrete example of an equilibrium in the text, with more examples in the Appendix. In the example, plots differ in both size l and amenity bundle B. There is sorting whereby holding amenities fixed, bigger plots will go to higher income people; and, holding plot size fixed, plots with better amenities will go to higher income people.

We conclude this section by drawing conclusions to guide our empirical investigation.

4.1 The consumer's optimization problem

To rationalize the movement of people out of city, we assume that the city amenity deteriorates over time relative to the 20k amenity (e.g., due to congestion or pollution). Specifically, we assume

---

[15] Some papers, especially on developed countries, report correlations of plot size and outcomes such as land values (e.g., Asabere and Colwell, 1984; Brownstone and De Vany, 1991; Colwell and Sirmans, 1993; Kolbe et al., 2021; Combes et al., 2021)

that at time zero, city amenity (A) is higher than the 20k amenity (B), but while the 20k amenity is fixed over time, the city amenity deteriorates at a rate θ.

In the city, housing offerings are market driven, allowing each person to choose their optimal housing h, for a given unit price. However, in the planned 20k areas, the supply of plots of each size is limited, and each owner builds their own housing by combining the land on their plot (l) with capital (k).

Each owner decides on the time τ to move from the city and sink k by solving the optimization problem:

$$\begin{aligned} \max_{h_1,z_1,k,z_2,\tau} \quad & \int_0^{\tau} [\varphi lnh + \beta lnz_1 + Ae^{-\theta t}]e^{-\rho t}dt + \int_{\tau}^{\infty} [\varphi ln(l^{\alpha}k^{1-\alpha}) + \beta lnz_2 + B]e^{-\rho t}dt \\ & + \omega\left(\int_0^{\infty} we^{-\delta t}dt - \int_0^{\tau}(ph + z_1)e^{-\delta t}dt - \int_{\tau}^{\infty} z_2e^{-\delta t}dt - rke^{-\delta\tau} - R(0)\right), \end{aligned} \tag{1}$$

where z_1 (z_2) denotes all goods other than housing consumed in the city (upon moving to the 20k area); k is the housing capital invested at the time of move τ; and l is the size of the plot that the owner buys for a price $R(0)$ in time 0.

We assume that and non-housing consumption is the numeraire, p is the rental price (or the opportunity cost) of housing in the city, and r is the purchase cost of capital. We make the strong assumption of a perfect capital market, noting we are considering owners who are among the richest in their society. What we call wage (w) is a measure of permanent income, where with a perfect capital market all that matters is W: the present discounted value of lifetime earnings (including any endowment). Finally, ρ is the personal discount rate and δ is the interest rate. We equate ρ and δ, which will imply that consumption ($z_1 = z_2 = z$) 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.

4.2 Interpreting variation of equilibrium choices

Looking ahead to the empirics, our focus is on the effects of the two main planning variables: plot size (l) and amenities, (B). Here we develop an economic framework for interpreting this observational variation. We obtain expressions for our outcomes of interest from the first order conditions for equation 1 (see Appendix B). We examine four types of outcomes: land prices; sorting of into plot ownership of owners with different incomes; the probability that a plot is developed by a certain date, say 15 years after the project began, $Prob(\tau < 15)$, which is inversely related to τ; and the level of capital investment.

Beginning with land prices as an outcome, higher prices must be paid for higher l and B in any Nash equilibrium. For example, if two plots of the same size offer different amenities, they cannot have the same price: a consumer of the worse amenity plot would be willing to pay more for the better amenity plot. Thus, prices rise with B or l according to the equilibrium elasticities:

$$\eta_{R,B} \equiv \frac{\partial R}{\partial B}\frac{B}{R} \geq 0\ ;\ \ \eta_{R,l} \equiv \frac{\partial R}{\partial l}\frac{l}{R} \geq 0$$

Next we turn to dates of development and investment levels.

4.2.1 Results for date of development

In the Appendix, Eq. 7 gives an expression for τ. Differentiating that equation we get:

$$G d\tau = \underbrace{\frac{dw}{w-\delta R}}_{\text{sorting effect}} \underbrace{-\left(\frac{B}{\alpha\varphi} + \frac{\delta R}{w-\delta R}\eta_{R,B}\right) dB/B}_{\text{amenity effect}} \underbrace{-\left(1 + \frac{\delta R}{w-\delta R}\eta_{R,l}\right) dl/l}_{\text{plot size effect}} \tag{2}$$

where $G \equiv \frac{A\theta e^{-\theta\tau}(\beta+\varphi(1-\alpha e^{-\delta\tau}))+\delta e^{-\delta\tau}\alpha^2\varphi^2}{\alpha\varphi(\beta+\varphi(1-\alpha e^{-\delta\tau}))} > 0$.

In 2, **holding owner income constant**, higher amenity and bigger plots are developed sooner (τ is lower). For example for B, the direct effect of a higher amenity 20k plot, $\frac{B}{\alpha\varphi}$, makes the owner move sooner from the city; this is reinforced by the indirect price effect, $\frac{\delta R}{w-\delta R}\eta_{R,B}$, which makes a higher amenity plot more expensive, shifting owner expenditure from the city to 20k, and inducing faster development. A similar argument applies to larger plots, where again the direct effect and the price effect induce faster development.

This picture may change, however, once we consider the sorting of owners into plots with higher B or l. For instance, if raising lot size leads to sorting in of richer individuals, this will dampen the negative effect of lot size on time of development τ. We consider those sorting effects in the specific examples of equilibria discussed below.

4.2.2 Results on investment

The equation for k is Eq. 8 in the Appendix, where k is directly affected by w and τ. Differentiating that equation and substituting in for $d\tau$ from 2, we get:

$$\frac{dk}{k} = \frac{X}{X+Z}\left(\underbrace{\frac{dw}{w-\delta R}}_{\text{sorting effect}} + \underbrace{\left[\frac{B}{\alpha\varphi}\frac{Z}{X} - \frac{\delta R}{w-\delta R}\eta_{R,B}\right] dB/B}_{\text{ambiguous effect of B}} + \underbrace{\left[\frac{Z}{X} - \frac{\delta R}{w-\delta R}\eta_{R,l}\right] dl/l}_{\text{ambiguous effect of l}}\right) \tag{3}$$

where $X \equiv\ A\theta e^{-\theta\tau}(\beta + \varphi(1-\alpha e^{-\delta\tau})) > 0$ and $Z \equiv\ \delta e^{-\delta\tau}\alpha^2\varphi^2 > 0$.

In equation 3, again holding income constant, the direct, complementary effect of an increase in either B or l is to increase k. However, that increase is potentially offset by the indirect price effect, whereas R rises that squeezes the budget for k. The result is that even holding income constant, the effect of an increase in either B or l on k is ambiguous. This ambiguity is pushed towards a positive association by the owner sorting effects, which we discuss below.

4.3 Equilibria: Price and sorting effects

We illustrate an equilibrium with two plot sizes each with 4 sets of amenities. As noted above, in equilibrium, people moving to the 20k area have incomes in the interval interval $[w_m, \bar{w}]$. These consumers divide themselves across plot sizes and amenities. Small plots with the lowest amenities will go to the lowest income segment, defined by when the size of the segment is such that the count of people in that segment equals the number of smallest lowest amenity plots. The next adjacent income segment will be on a larger and better amenity plot type until the supply of those plots is exhausted. In general, as we go up the income scale, we move to either larger plots or higher amenity plots based, from the preference function, on the plot-level index $\phi \alpha ln(l) + B$, that serves the next income interval of consumers until the supply of those plots is exhausted.

In any solution the key to solving sorting and prices involves consumers at margins. The first margin involves the lowest income person in 20k areas. The price of a small, lowest amenity plot is that which leaves the w_m consumer indifferent between remaining in the city and consuming a small, lowest amenity plot. All higher income people in the interval which consume small, lowest amenity plots pay this price and get a surplus from being in the 20k area over the city. Then successive prices are defined by the marginal consumer at the beginning of the next income interval, who are indifferent between being on a smaller or lower amenity plot and moving to the next larger or higher amenity plot. The pricing principle is developed in detail for one case in Appendix B.

To illustrate equilibria, we calibrate the model to fit general patterns in the data and utilizing parameters from the literature. Details of the calibration exercise are in Appendix B.3. We choose an interval of incomes that settle in the 20k area based on the income range in our survey data. Then we choose somewhat arbitrarily eight sub-intervals of incomes that sort into a corresponding set of eight plot types (distinguished by two plot sizes and four amenity levels). Results from this stylized example are plotted in Figure 5. In these figures, blue lines represent small plots and red lines large plots, while darker lines represent higher amenity values.

In part (a) of Figure 5, we show that an equilibrium exists. Each income interval has a different colored curve. The solid parts of the colored curves plot the outer envelope of realized lifetime utilities net of what people would get in the city. In the figure, no person wants to switch from their plot size-B combination at equilibrium prices to a different combination. The person at w_m has zero net gain from being in 20k areas. In this example, the largest amenity small plots sell to a higher income interval than the lowest amenity large plots. If B does not vary much, it is easy to do an example where all small plots are sold to lower income people than large plots. In part (b) we plot equilibrium price per square meter for each plot size-amenity type. First, total plot price rises with amenities and plot size, an essential feature of any equilibrium. Second, as can be seen in the figure, the price per sq meter of land rises with amenity (as lines get darker). In contrast, for a given amenity level, price per sq meter falls with plot size (blue to red lines). This demonstrates that large plots are "over-supplied" in this equilibrium. To equalize prices on equal amenity land we would need to have relatively more smaller plots of each amenity type. That is, in the examples we have solved, that means moving the income cut-off points into larger plots to the right.

In part (c) of Figure 5, we show how τ varies by income for different plot size-amenity 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 (i.e. τ drops noticeably), as illustrated by the horizontal dashed straight lines at the the cross-over level of wealth from the lowest valued plot to the second lowest. Similarly, for a given income and amenity level, increasing the plot size lowers τ noticeably. However, between switch points, as income rises within an interval, so does τ, for the same B, R and l, as shown in Eq. (2).[16] Including sorting, part (c) of Figure 5 hints that, holding l fixed, τ could tend to drop as B rises.

In part (d) of Figure 5, we show how investment changes as plot size and B increase. At any margin of increase, the numbers in the figure tell us that the ambiguous effect in the comparative statics of an increase in l or B here is a tiny negative effect on investment. The increase in R that impinges on the budget barely outweighs the direct positive effect. However, what the figure says most clearly is that investment rises sharply and almost linearly with income and is little related to either l or B. In regressions the effect of the omitted w variable and of sorting dominate, so that investment rises strongly with l and B.

4.4 Summary

Below we estimate several 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 2020 or not and at investment levels if developed, as key outcomes. That involves an econometric implementation of Appendix equations 7 for τ and 8 for k, with impacts defined by equations 2 and 3 respectively. For time of development, we only observe whether a plot is developed by 2020. If we think this is 15 years after the start of development we are then modeling the $Prob(\tau < 15)$. One way to implement this is to do a first order Taylor series expansion of $e^{-\theta\tau}$ and $e^{-\delta\tau}$ in Appendix Eq. 7 and rearrange to yield

$$\tau = C - FB - \alpha\varphi Fln(l) + \alpha\varphi Fln(w - \delta R)$$

---

[16] Within intervals with the same plot price, as incomes rise, they have more after-plot-purchase wealth to spend on market goods when in the center city, so the role of diminishing amenities is relatively less important.

where C is a constant and $F \equiv \frac{\beta+\varphi(1-\alpha)}{A\theta(\beta+\varphi(1-\alpha))+\delta\alpha^2\varphi^2} > 0$. If we rewrite B so that $B = \bar{B}_i + \epsilon_{ij}$ for person j on plot i and we assume a symmetric distribution for ϵ. Then we will be estimating $Prob(M_{ij} < \epsilon_{ij}) = \Phi(-M_{ij})$ where $-M_{ij} \equiv F^{-1}[C - 15 + \bar{B}_i + \alpha\varphi ln(l_i) - \alpha\varphi ln(w_j - \delta R_i)]$ where Φ is the cumulative distribution. Note $\bar{B}_i$ will be parameterized with amenity covariates and coefficients. Below we implement a LPM version of this.

In terms of an equation for k, again we can do a first order Taylor series expansion and linearize equation Appendix Eq. 8. With the strong sorting effects we think with the omitted variable w, k will have a positive association with l and B.

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. When studying the consequences of plot size, we also use 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:

$$y_i = \beta_1 \text{Plot\_size}_i + \mathbf{Program\_area}_i' \gamma_1 + \mathbf{Controls}_i' \delta_1 + \epsilon_{1i}, \tag{4}$$

where y_i is the outcome (e.g., logarithm of plot price or plot price per square meter, some measure of housing investment, and whether people have built by 2020) in the gridcell or its plot and Plot_size_i is a measure of the size of the plot in which gridcell i's centroid falls (e.g., logarithm of size in sq m or indicators for medium and large plots). $\mathbf{Program\_area}_i'$ 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; often we further interact these with f.e. for mitaa, which are small administrative units. The standard controls, $\mathbf{Controls}_i$, which we refer to as "amenities" (though some are disamenities), include predetermined and planned features. The predetermined features are: distance in km to the nearest major paved road; elevation; ruggedness; and indicators for being within 100 m of a river or a stream and of water or wetland. The planned include a three-way Z-index of insula characteristics (rectangularity, alignment, and homogeneity, and dummies) and indicators for being within 100m of a 20k area edge and each of the following planned non-residential land uses: recreation, nursery school, education, religious site, service trade, housing estate, public building, cemetery, and any other. In price regressions, time period interactions by source of data (real estate agents or residents) are also 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, or even the 12 project areas.

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

$$y_i = \beta_2 \text{Own\_larger}_i + \textbf{Program\_area}_i' \gamma_2 + \textbf{Dist}_i' \delta_2 + \textbf{Boundary}_i' \rho_2 + \textbf{Controls}_i' \kappa_2 + \epsilon_{2i}, \qquad (5)$$

where Own_larger_i is indicator for gridcell i's insula having a larger mean plot size than the insula with which it shares a boundary segment; $\textbf{Dist}_i$ is the distance in meters to the insula boundary segment interacted with indicators for the "large" and "small" side of each boundary; $\textbf{Boundary}_i$ is vector of boundary segment f.e.; and ϵ_{1i} is an error term, again clustered by insula. The RD regressions we estimate are typically semi-parametric RD, where we restrict the analysis to gridcells within 100m from their insulae boundary, which includes the majority of gridcells.

This RD 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 spatial units, in analyzing a greenfields setting, and in focusing on spatial discontinuities within small administrative units, which allows us to cleanly address threats to identification.

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 logarithm of the plot size difference between the gridcell's insula's mean plot size and that of its adjacent insula's mean. This final specification offers a direct comparison to the OLS estimates of log plot size effects.

5.2 Empirical findings

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. Table 1 shows that the logarithm of real land prices increased in all project areas, with a mean increase of about two, corresponding to a rise of over 600%. There was considerable dispersion in price increases across project areas, although areas that were initially more expensive did not systematically do better or worse.[17]

Not all the price appreciation above was necessarily due to the 20k project. To learn more, Table 2 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 20k land prices are about twice those of non-20k informal plots (since $e^{-0.7} \approx 0.5$). In between the two is the relatively small share of non-20k plots that were formalized, reflecting more secure property rights without the planning benefits of 20k. The results are similar when we control for (nearest area) fixed effects. These estimates 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 on average about twice higher than in informal non-20 plots within their mtaa. If informal plots near 20k appreciated due to the 20k development, these large estimates may even understate the gains from 20k.

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?". The 31 (of 34) leaders who answered this question emphasized two factors: property rights and access. First, among the 31 responders, 24 mentioned property rights (of which 21 mentioned land titles explicitly), saying that they reduced 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, and of these 20 mentioned roads specifically. The leaders suggested that non-20k areas tend to clog up over time, and some said that inadequate access outside 20k made local service provision harder. Explanations other secure property rights and access were much less common in the mtaa leaders' interviews.

5.2.2 Implications of own plot size

While we are interested in the elasticity of price of bare land per square meter with respect to plot size, we begin with regressions of specification (4) where the outcome is the logarithm of (overall) plot price, which avoids potential concerns about division bias (if a noisy measure of plot size enters both sides of the regression). Panel A of Table 3 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. This is indeed what columns (1) and (2) of Panel B of the same table show when the logarithm of plot price per square meter is the outcome. Columns (3) and (4) use the official size cutoffs for plot sizes, showing that price per sqm decreases in size. Panel A of Figure A.3 shows the same negative relationship between non-parametrically, using 100-meter bins as regressors. Appendix Table A.3 shows that

---

[17] Coefficients on period time effects (not shown) indicate that prices initially rose even more (by about 20-30% from 2000-2010), before before declining a bit and stabilizing at the high levels reported in Panel Table 1.

outside 20k, where plot supply sizes are not fixed, the price elasticity is lower, at least for surveyed plots.[18]

Next, we examine the implications of plot size for housing outcomes. Table 4, which reports estimates from specification (4), suggests that the share of the gridcell that is built declines in plot size; the probability that a plot is built circa 2020 is unaffected by plot size, at least once we control for amenities; and that for built plots, the elasticity of built area with respect to plot size is around 0.11-0.14. In other words, large plots have much more open space but also, as the final column shows, a higher likelihood of housing multiple buildings. This could reflect more outbuildings or some backyarding in larger plots (Brueckner et al., 2019), or both. Panels B-E of Appendix Figure A.3 show the same relationships non-parametrically, and at least for well-populated bins the relationships are mostly monotonic.

While the estimates reported above include some with controls for small area fixed effects and many amenities, a residual concern may be that plot size is correlated with unobserved amenities. To address this, we estimate RD specifications of specification (5), which we report in Table 5. Panel A shows estimates for all insula pairs, irrespective of the mean plot size difference between them. Here, as in the OLS estimates above, plot size reduces the price and the share built, and increases the probability of multiple buildings; the only qualitative difference from OLS is that here there is no significant effect on log building size. Panel B restricts the analysis to insula pairs with large (>400 sqm) difference in mean plot size, and the precisely estimated coefficients increase about two-threefold. Panel C studies discontinuities with small size gaps (<100 sqm) and only the coefficient on share built is significant, and its size is halved compared to Panel A. To further support our identification strategy, Appendix Table A.4 reports balancing of the three RD specifications on first-nature locational fundamentals, and we see no significant differences in the Z-index that combines these amenities. And Appendix Table A.5 reports a different robustness check for both OLS and RD, where we drop gridcells with plots that initially contained buildings, and the results are largely unchanged.

While OLS and RD specifications give robust and qualitatively similar results, we are interested in comparing them quantitatively. In Table 6, we report estimates of a specification where own larger is interacted with the gap in log mean plot size between neighboring insulae plots. A comparison of Panels A and B shows that the plot size effects are remarkably similar between OLS and the interacted RD, even though the RD estimates are not as precise since RD uses only variation around the insulae boundaries. Taken together, our estimates provide consistent evidence that plot size decreases land values and share built, does not significantly change the probability that a plot is built, and increases the log size of buildings and the probability of multiple buildings

---

[18] Panel A of Appendix Figure A.4 shows that for unsurveyed non-20k plots, which are smaller, and the price elasticity is also lower from around 200-700 sqm.

(our measures of k). Our estimated elasticity of land value with respect to plot size, of around -0.5, strongly suggests an oversupply of large plots, which is inefficient.[19] 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. Given the administrative and technological constraints faced by the governments, this planning rigidity may provide owners confidence that their property rights are protected, supporting 20k land values as discussed above.

To complete our discussion of plot size effects, we consider their implications for population density. Appendix Table A.6 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. Here we again use specification (5), but this time for boundaries between super-insulae rather than insulae and controlling for the log size of the own plot.[20] At a border between super-insulae of different types, residents experience a mixed neighborhood (e.g., 50%-50% small and large plots). As we move into the super-insula interior, residents are increasingly exposed to neighboring plots with a similar size to theirs. Since the number of super-insulae boundaries is much smaller than the number of insulae boundaries, we focus here on the housing outcomes and not on prices, which are available for only a small fraction of the sample. Table 7 shows that as we move into neighborhoods with smaller plots and homogeneity increases, we see a higher likelihood that a plot is built and a larger share of the plot that is built. As we move away from the boundary on the larger plot side, however, the estimated coefficients are all insignificant. So perhaps surprisingly, it is small plot

---

[19] For concreteness, 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 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.

[20] We restrict the sample to boundaries that are no more than 30 meters apart, losing about 2% of the sample.

owners who value uniformity of the neighborhood, not large plot owners.

We further explored this result in two ways. First, in Appendix Table A.7, we distinguish between three types of borders: small-medium, small-large and medium-large. The results indicate that moving into small (vs medium) super-insulae increases the share built and the probability that a plot is built; moving into medium (vs large) super-insulae increases the probability that a plot is built, as does moving into large (vs small) super-insulae. All these support valuation of homogeneity, again perhaps more so for super-insulae with small plots.

Finally, we used a different way of measuring the locality, by calculating the share of each gridcell's neighboring gridcells that are "small" (smaller than 800 sqm) within 100 meters, restricting the calculation to residential plots and excluding the own plot's gridcells. We then estimate specification (4), adding as covariates the share small and the share small interacted with own plot size. This formulation allows us to use price as well as housing outcomes. The results reported in Appendix Table A.8 are mostly imprecise, but the coefficient on share neighbors small is mostly positive and the coefficient of share small interacted with plot size is typically 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 suggests that small plot owners value being surrounded by small neighbor plots. This positive effect declines to 0 at plot size of about 1900 sq m.

Taken together, our regressions suggest that owners of small plots in particular may benefit modestly from having similarly-sized neighboring plots.

5.2.4 Valuation of amenities

We next turn to the valuation of amenities more generally. Table 8 reports estimates of specification (4) for all our baseline controls. The most clearly valued amenity is access. Increasing the distance to a main paved road by 1km reduces land value by almost 15% and entails large and precisely estimated declines in all the housing outcomes. Next are the natural ("first nature") features. While none seems to move prices significantly, elevation seems beneficial with three positive and significant housing outcomes, while ruggedness is likely a disamenity, with mostly negative coefficients on housing, as do proximity to rivers (or streams) and water (or wetland). 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.

Next come the planned amenities. Higher Z-index of insula regularity seems to increase the share built and the probability that a plot is built.[21] Proximity to the edge of a 20k area does

---

[21] Appendix Table A.9 further suggests that rectangularity and possibly insula alignment (but not insula homogeneity) are valued.

not seem to make much difference. Finally, proximity to the nine types of non-residential insulae seems to be ignored, as almost all the coefficients are small and imprecise. However, this does not mean that owners do not value these amenities; instead, as we show, the problem is with the implementation rates.

Evidence that implemented non-residential amenities are valued is reported in Table 9. Here we again estimate specification (4), but use implemented non-residential amenities instead of planned ones.[22] The regressions, while not causal, show that some amenities (especially services trade and housing estates, but also others) either attract housing, or follow it, or some mix of the two.

While implemented non-residential amenities are valued, Figure 6 shows that implementation significantly lags the plans. Panel A presents compares the number of non-residential plots planned and implemented for different planned uses. Of the eight main planning categories, implementation lags planning somewhat for three categories (cemeteries and religious and educational uses), and much more for the remaining categories (recreation, public buildings, nursery, service trade, and housing estate). At the same time, plots designated as residential are often misused as residential or for farming or have uses that are unknown (to local leaders) or no use at all, where some unused plots are kept and others are not. Panel B focuses on the non-residential plots intended for the eight main planning categories and asks how each is used. Around half the plots intended for cemeteries and just below 40% of those intended for educational or religious uses are implemented as planned, while the implementation rates for the other five categories are much lower. 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.

Even though implementation rates for non-residential amenities are low, we examine whether conditional on implementation, the planners' designated locations are generally followed. To test this, Appendix Table A.10 reports actual 'compliance ratios', which we define as: P(implemented as j — planned as j)/P(implemented as j). The table shows that the actual rates in the first column are not far from those that would have resulted from perfect implementation (second column) and much higher than those that would have resulted from random implementation. In other words, conditional on implementing non-residential uses, the planners' intent was largely.

5.2.5 Sorting of owners and other residents

We conclude our discussion of the results with evidence on the sorting of residents by years of schooling, 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. Figure 7 shows two aspects of the sorting of

---

[22] Since implementation rates are low and the price sample is small, we focus here on the housing outcomes.

plot owners, which reside in around half of the plots we surveyed. First, mean years of schooling (around 13-14) 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.[23] 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.

The sorting of owners by education is also evident when we consider individual plot characteristics. As shown in Table 10, owners of larger plots, of plots whose land is more expensive, and of plots which are currently more valuable (including housing), as reported by the owners, are more educated. Surprisingly, this is not the case for owners of plots closer to main paved roads, and within 20k project more educated owners even live further from these roads. But the sorting by education within area on plot size, land value, and overall plot value is again positive. Appendix Table A.11 relates sorting to settlement dates for owners, and shows that in line with the model's predictions, conditional on owners' years of schooling (our proxy for income), plots that are larger or with better amenity (closer to a main road) were occupied earlier (column 2); the estimates are similar when we consider as outcome the year construction began (column 4).

6 Concluding Remarks

De novo urban planning provides a key policy option for developing country cities faced with large and rapidly growing informal areas. This paper provides, to the best of our knowledge, the first systematic quantitative analysis of the planning decisions that de novo entails, which we study in the context of the 20k plot project in Dar es Salaam.

Our findings show that providing the 20k plots was cheap (around 1 USD per sqm of residential plot), and this cost was recouped through the buyers' payments. This cost recoupment resolves a major concern with earlier de novo projects, such as the World Bank's "Sites and Services" projects. The 20k plots also increased land values, perhaps as much as doubling them compared to nearby unplanned areas.

We also uncover evidence that two key mechanisms are central for the 20k project's success: first is the protection of owners' property rights and second is the preservation of access through local unpaved roads, which connect to main paved ones.

Nevertheless, de novo neighborhoods also have limitations, which can be improved. Key among these is an oversupply of large plots, which command lower land values, a small share of built space, and much lower population density. An important likely cause for this over-provision is the persistence of colonial-era rules. Our evidence suggests that offering relatively more smaller plots and a sales process that is less rushes may make de novo projects both more valuable and more inclusive, allowing more people to benefit from affordably priced formal plots.

---

[23] In 2014, heads of households in Dar es Salaam had, on average, 8.7 years (World Bank, 2019).

Our findings also indicate that non-residential amenities are largely ignored, generating neither land value appreciation nor more built activity. We show that this is most likely due to low implementation rates.

Finally, we find that despite their scarcity, only half the residential plots in our setting are built upon. We show that slow plot development is in part due to plot characteristics (e.g., plots that are smaller and with worse amenities), and in part due to higher income people delaying their move into the de novo plots. Still, an important question which we explore in follow-up work is whether there are other important factors that can shape the dynamics of de novo ploy settlement.

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'.

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

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.

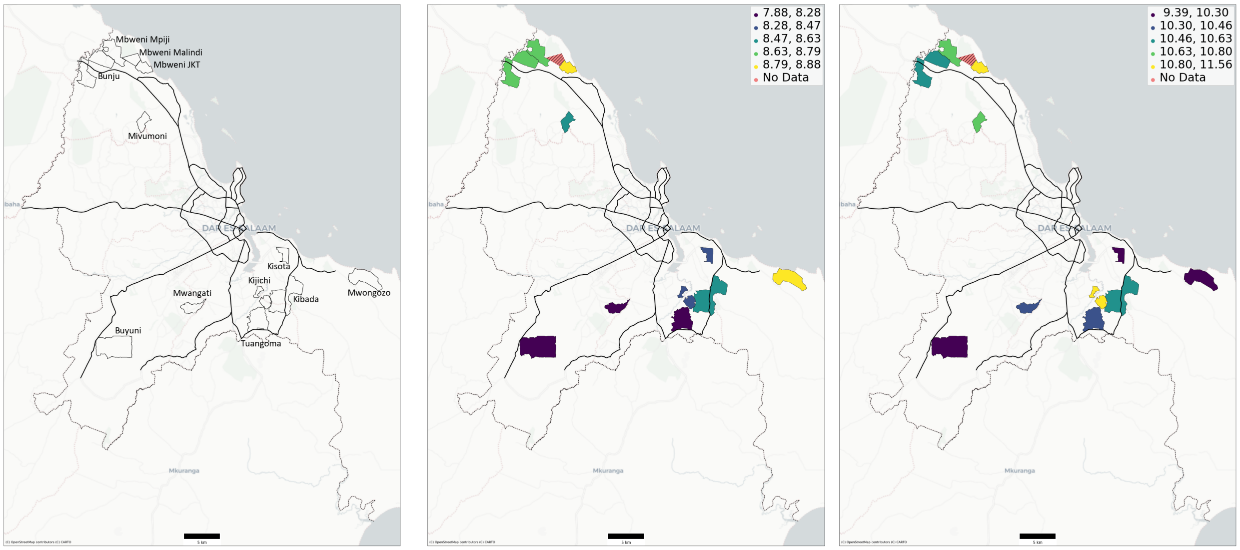
World Bank (1987), Tanzania: Second national sites and services project, Technical Report 6828.

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

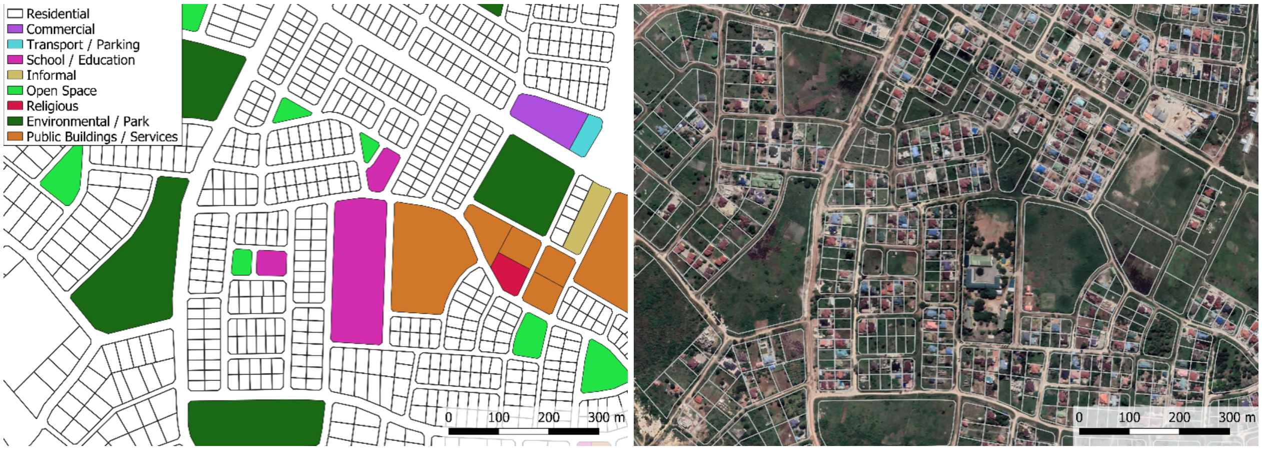
Figure 1: Map of 20k project areas in Dar es Salaam



(a) 20k project area names (b) Initial ln price per sqm (c) Current ln price per sqm

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. Panel (a) shows the names of each 20k project area. In Panel (b), each area is colored by its initial government-charged ln price per sqm (in 2021 TZS). In Panel (c), each area is colored by its predicted current transaction ln price per sqm (in 2021 TZS).

Figure 2: Example of land uses in Mbweni Mpiji



(a) Planned plot boundaries and non-residential uses (b) Planned plot boundaries and satellite imagery

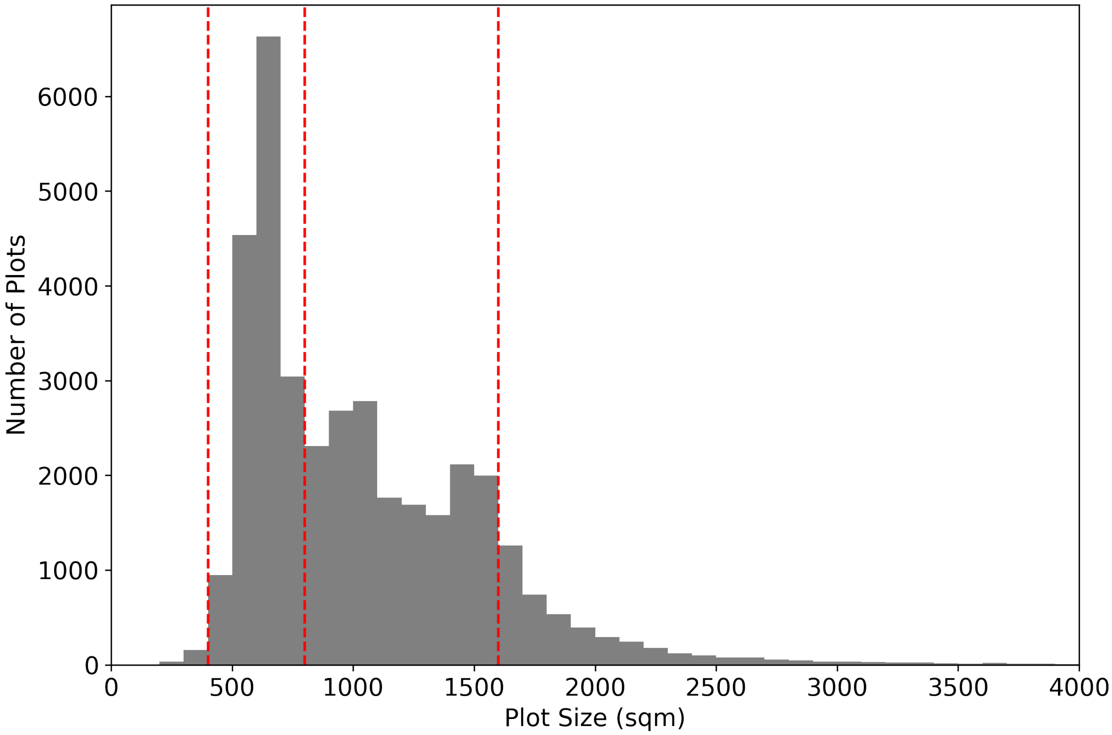
Notes: This figure shows an example of planned plot boundaries in Mbweni Mpiji. In Panel A, each plot is colored by its planned use. In Panel B, satellite imagery is displayed in the background.

Figure 3: Example of 20k boundary in Tuangoma

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

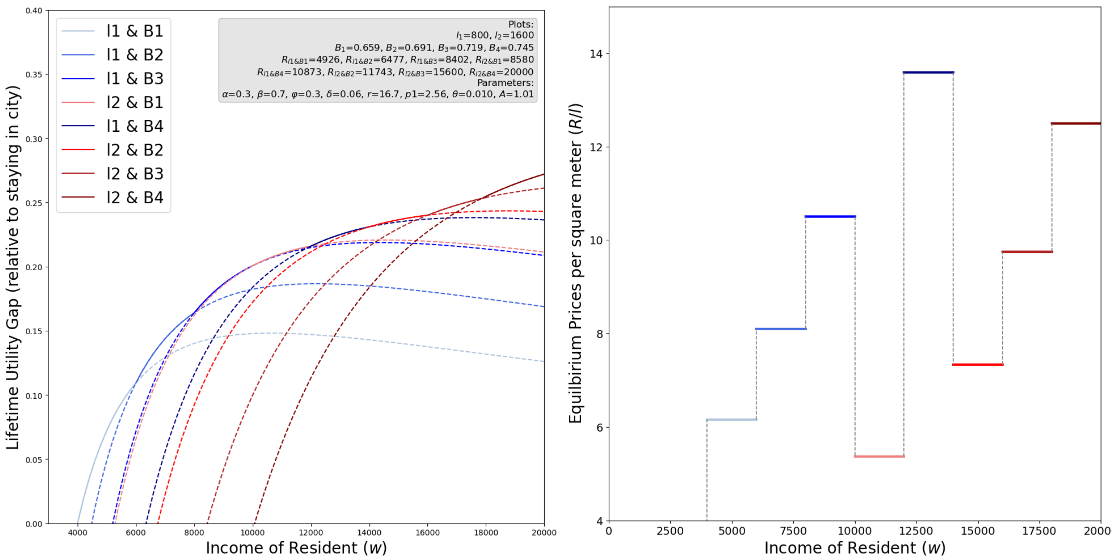
Notes: This figure shows an example of a 20k project boundary in Tuangoma. In Panel A, background satellite imagery is from 2001 (pre-implementation). In Panel B, background satellite imagery is from 2021 (post-implementation).

Figure 4: Histogram of residential plots sizes



Notes: This figure shows the histogram of residential plot sizes in our sample, using 100-sqm bins. 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 with varying amenities

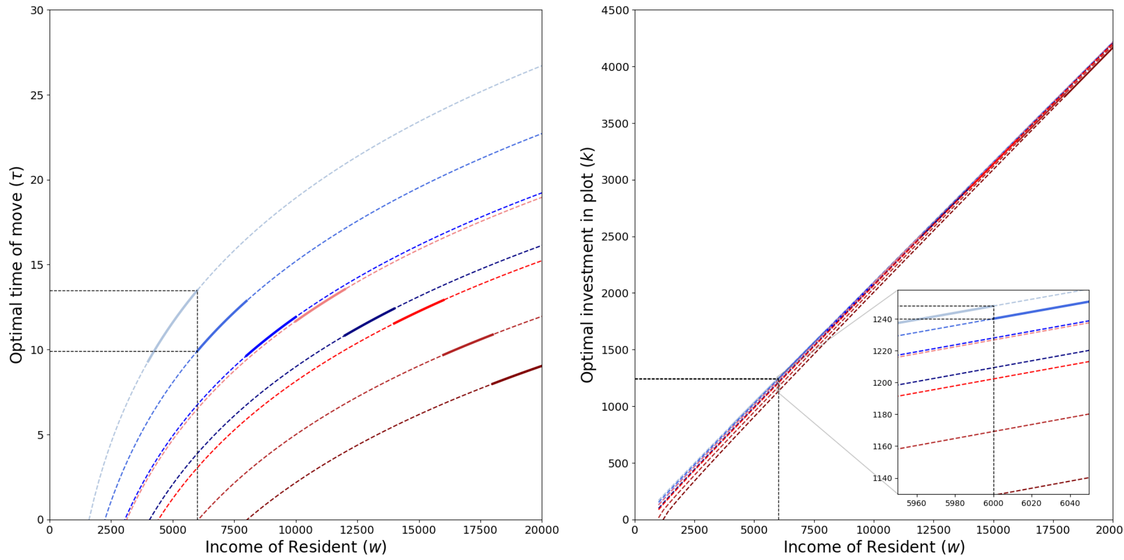


(a) Lifetime utilities

(b) Equilibrium price per sqm (R/l)

(c) Optimal τ in equilibrium

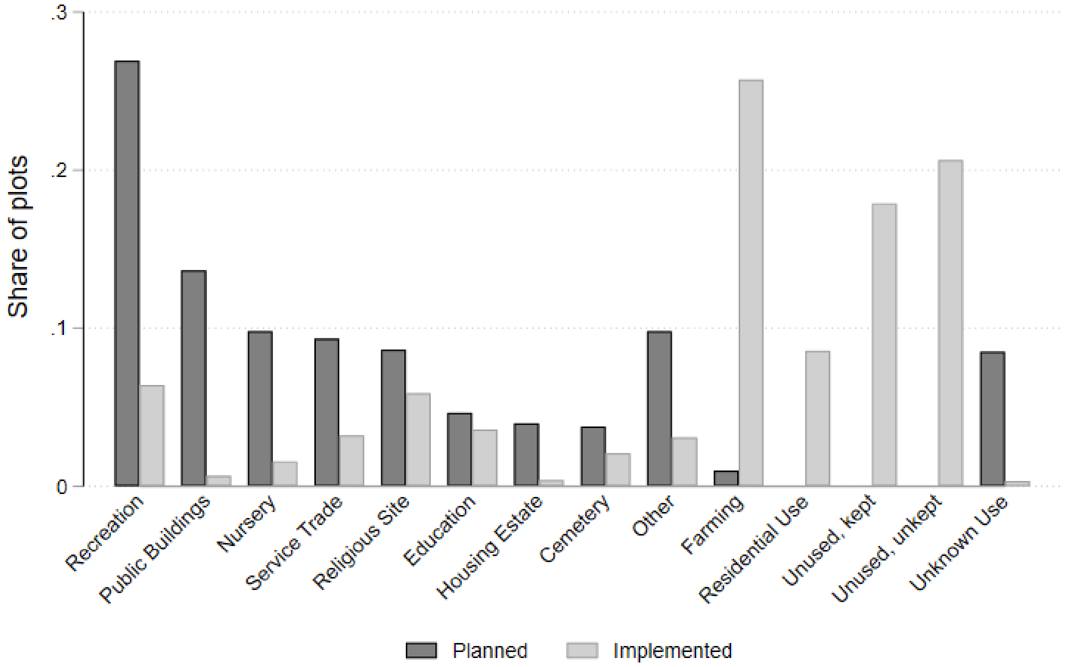
(d) Optimal k in equilibrium



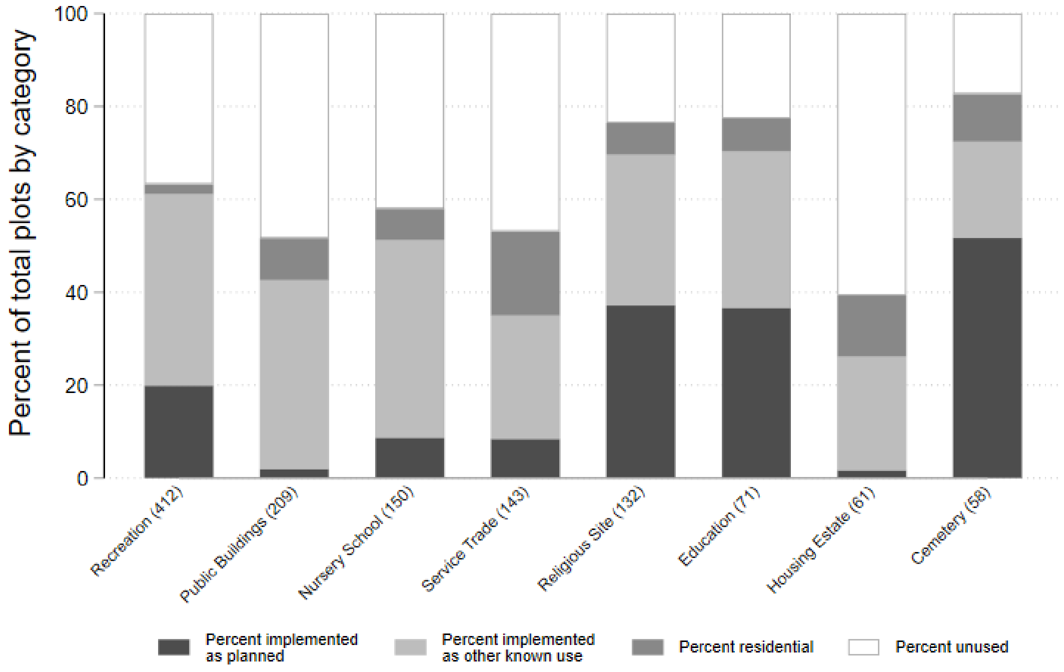
Notes: This figure shows equilibrium model outcomes vs resident income for different plot sizes and amenity values. Small plots are represented by blue lines and large plots are represented by red lines. Higher amenity plots are represented by darker colored lines. In panels (c) and (d), the vertical dashed black line denotes the cross-over level of wealth from the lowest valued plot to the second lowest, and the horizontal lines depict the optimal choice for each type of plot at this wealth level Panel (a) plots the lifetime utilities as outcomes. 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-amenity combination. In panel (b) the outcome is the equilibrium plot price per square meter. Each line corresponds to the price of an amenity-size level over the range of incomes purchasing that plot type in equilibrium. In panel (c), the outcome is the optimal time of move. 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 in equilibrium, while the dashed lines show out-of-equilibrium choices of τ. In panel (d), the outcomes is the optimal capital investment. Again solid parts of the line depict in-equilibrium, and dashed lines out-of-equilibrium, choices.

Figure 6: Non-residential plots by planned and implemented uses

(a) Planned and implemented land use on non-residential plots

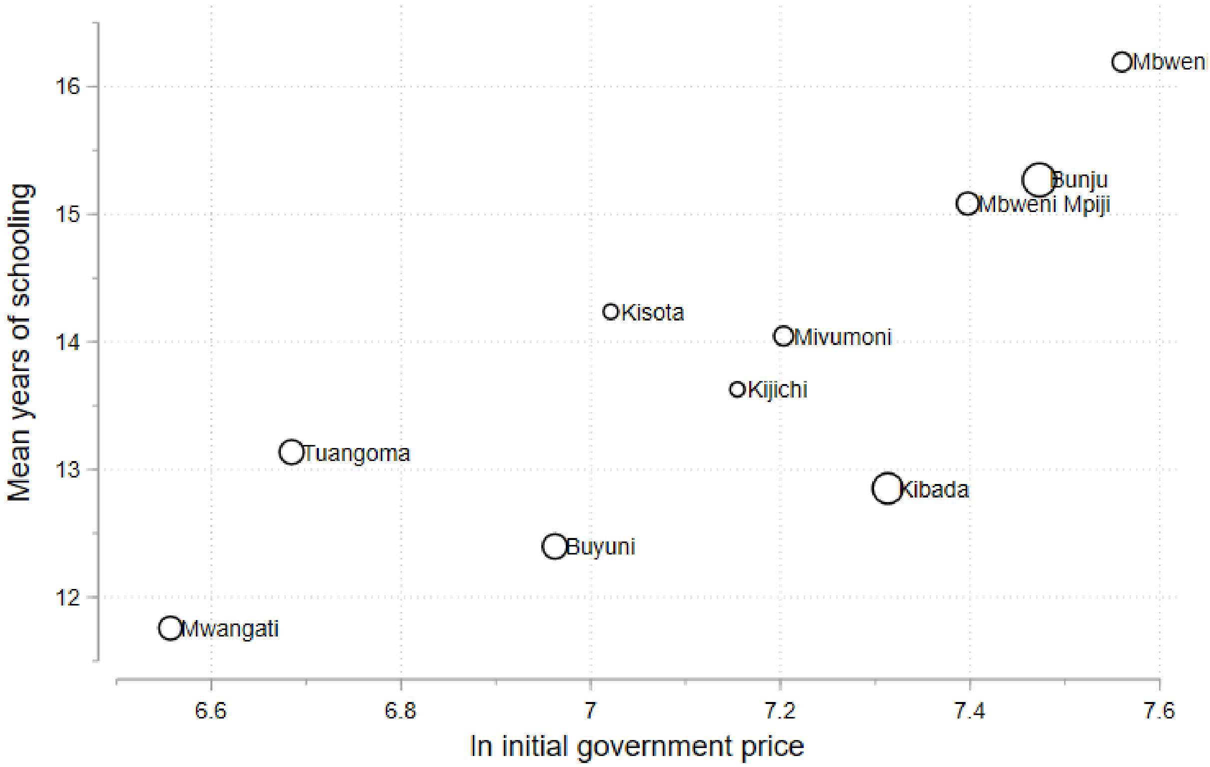


(b) Planned non-residential broken into implemented uses



Notes: this figure describes planned and implemented uses on non-residential plots. In Panel (a), the share of planned non-residential plots by planned use (dark grey) and current/implemented use (light grey). In Panel (b), the breakdown of each planned use into its current/implemented uses.

Figure 7: Sorting by education across 20k areas



Notes: this figure plots mean years of schooling by initial ln price per sqm across 20k areas. The sample is restricted to owner residents. Note 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: Land price appreciation in 20k areas

	Initial Price (2000 TZS)	Ln Initial Price (2021 TZS)	Δ Ln Price (2021 TZS)
Bunju	1760	8.79	1.83
Buyuni	1056	8.28	1.12
Kibada	1500	8.63	1.90
Kisota	1120	8.34	1.88
Mbweni JKT	1920	8.88	2.59
Mbweni Mpiji	1632	8.72	2.07
Mivumoni	1344	8.52	2.13
Mwangati	704	7.88	2.44
Mwongozo	1920	8.88	0.99
Tuangoma	800	8.00	2.46
Kijichi	1280	8.47	2.66
Mean	1367	8.49	2.01

Note: This table shows initial government fees and price appreciation for land price per square meter in 20k areas. Initial prices are based on Mwiga (2011) Table 6.4 which was sourced from the Tanzanian Ministry of Lands in 2010. Notably this source does not contain information for Malindi and so it does not appear here. In the second column, we take logs and inflate the initial prices to 2021 using Tanzanian inflation rates from Statista (2022). In the third column we take the difference in log land prices from initial to prices predicted for 2021 using our transactions data. Current (2021) prices are 20K area FE estimates + the constant from a regression of log price per square meter on 20K area fixed effects and period dummies (sold in; 2023, 2022, 2019-20, 2016-18, 2011-15, and pre-2010) interacted with a dummy for dalali vs. occupier survey (with dalali survey and sold in 2021 as base). The price data are from bare land transaction prices from both the Dalali and occupier surveys.

Table 2: Land price inside and nearby 20k areas

	(1) Ln Price	(2) Ln Price
Ln plot size	0.71	0.69
	(0.054)	(0.041)
Non-20K Surveyed	-0.23	-0.27
	(0.16)	(0.12)
Non-20K Unsurveyed	-0.70	-0.71
	(0.099)	(0.079)
Mean Outcome	17	17
20K or Nearest FE		✓
N	2074	2074

Note: This table uses data from the dalali survey for 20k and non-20k plots. It reports regressions of ln price on ln plot size and indicators for non-20k surveyed and non-20k unsurveyed plots. Each observation is a transaction which took place inside or nearby 20k areas. The sample consists of 1246 transactions inside 20K areas; 266 transactions of surveyed plots outside 20K; and 562 transactions of unsurveyed plots outside 20K. Controls include fixed effects for Municipality (Ilala, Temeke, Kigamboni, and Kinondoni) and transaction time period (2023, 2022, 2021 2019-20, 2016-18, 2011-15, and pre-2010). Note that there are no 20k areas in Ubungo, the fifth Municipality in Dar es Salaam. Column 2 additionally includes fixed effects for the the nearest 20k area (that is, the own 20k area for transactions inside 20k areas and the nearest 20k area for transactions outside 20k areas). Standard errors in parentheses are clustered by 20K area.

Table 3: Prices and Plot Sizes in 20k areas (OLS)

	(1)	(2)	(3)	
Panel A: Ln plot price				
Ln plot size	0.55	0.46	0.49	
	(0.071)	(0.053)	(0.060)	
Mean Outcome	17	17	17	
20k*Mtaa FE		✓	✓	
Amenities			✓	
N (gridcells)	4074	4074	4074	
N (plots)	1446	1446	1446	
Panel B: Ln plot price per square meter				
Ln plot size	-0.54	-0.51		
	(0.053)	(0.060)		
Medium			-0.33	-0.28
			(0.043)	(0.045)
Large			-0.61	-0.56
			(0.063)	(0.066)
Mean Outcome	10	10	10	10
20k*Mtaa FE	✓	✓	✓	✓
Amenities		✓		✓
N (gridcells)	4074	4074	4074	4074
N (plots)	1446	1446	1446	1446

Note: This table presents OLS regressions of plot price and size relationships. In Panel A the outcome is always log plot price, and in Panel B the outcome is always log plot price per square meter. Prices combine bare land transactions from the dalali and occupier surveys. In Panel A the regressor is always log plot size, and in Panel B the regressor is log plot size (cols 1-2) or indicators for medium or large plot size (cols 3-4). We always control for transaction period by source (dalali or occupier survey) fixed effects. Otherwise controls vary across columns as denoted in the bottom rows: 20K*Mtaa Area FEs (panel A cols 2-3 and panel B cols 1-4), and amenities (panel A col 3 and panel B col 2 and 4). Amenity controls include distance to major paved road, average elevation and ruggedness, a three-way Z-index of insula characteristics (rectangularity, alignment, and homogeneity), and dummies for within 100m of a 20k area edge, river, wetland, and each of the planned non-residential land uses: recreation, nursery school, education, religious site, service trade, housing estate, public building, cemetery, and any other. Note that in these specifications, the dummy for wetland within 100m is collinear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table 4: Built Outcomes and Plot Sizes in 20k areas (OLS)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Multiple buildings on plot
Panel A: 20k*Mtaa FE controls				
Ln plot size	-0.087	-0.031	0.11	0.18
	(0.0025)	(0.0091)	(0.017)	(0.011)
Mean Outcome	0.11	0.49	5.3	0.38
N (gridcells)	94789	94789	46465	46465
N (plots)	36215	36215	17822	17822
Panel B: 20k*Mtaa FE + Amenity controls				
Ln plot size	-0.078	-0.00040	0.14	0.19
	(0.0026)	(0.0094)	(0.018)	(0.012)
Mean Outcome	0.11	0.49	5.3	0.38
N (gridcells)	94789	94789	46465	46465
N (plots)	36215	36215	17822	17822

Note: This table presents regressions of five quantity outcomes on log plot size. In column 1 the outcome is the share of the gridcell area that is built. In column 2 it is an indicator for whether the plot is built [has at least one building above 30sqm]. In columns 3-4 observations are restricted to built upon plots only, and the outcomes are: log total area of the three largest buildings on the plot (col 3), and an indicator for multiple buildings on the plot (col 4). Controls vary across panels: panel A controls for 20K*Mtaa Area FEs and panel B adds amenities. Amenities are the same as described in Table 3. Standard errors in parentheses are clustered by insula.

Table 5: Prices, Built Outcomes and Plot Sizes in 20k areas (RD)

	(1) Ln Price per sqm	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Panel A: all insula pairs					
Own Larger	-0.17	-0.017	0.011	-0.018	0.026
	(0.055)	(0.0025)	(0.0080)	(0.018)	(0.013)
Mean Outcome	9.9	0.11	0.49	5.2	0.39
N (gridcells)	3511	87569	87569	42715	42715
N (plots)	1228	33613	33613	16474	16474
Panel B: gap≥400sqm					
Own Larger	-0.45	-0.036	0.0082	-0.0078	0.080
	(0.14)	(0.0050)	(0.016)	(0.042)	(0.030)
Mean Outcome	9.8	0.093	0.47	5.3	0.41
N (gridcells)	1003	22483	22483	10526	10526
N (plots)	335	9066	9066	4219	4219
Panel C: gap<100sqm					
Own Larger	-0.097	-0.0094	-0.0030	-0.011	-0.015
	(0.066)	(0.0042)	(0.013)	(0.027)	(0.019)
Mean Outcome	10	0.12	0.50	5.2	0.36
N (gridcells)	1016	30460	30460	15079	15079
N (plots)	472	15157	15157	7469	7469

Note: This table presents RD regressions across neighbouring insula boundaries. All panels restrict the sample to within 100m of the insula-pair boundary. Panel B further restricts to insula pairs with at least 400sqm gap in mean plot size, and Panel C to those insula pairs with no more than 100sqm gap. 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. In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*Mtaa Area, insula-segment FEs, and amenities. Amenities are the same as described in Table 3. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table 6: Prices, Built Outcomes and Plot Sizes in 20k areas (RD with size gap interaction)

	(1) Ln Price per sqm	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Panel A: RD across insulae with interaction for log mean difference					
Own Larger ×Δ ln mean size	-0.50	-0.078	-0.025	0.13	0.23
	(0.28)	(0.0052)	(0.017)	(0.046)	(0.029)
Own Larger	-0.073	0.0013	0.017	-0.046	-0.023
	(0.070)	(0.0027)	(0.0089)	(0.020)	(0.014)
Mean Outcome	9.9	0.11	0.49	5.2	0.39
N (gridcells)	3511	87569	87569	42715	42715
N (plots)	1228	33613	33613	16474	16474
Panel B: OLS with RD sample from panel A					
Ln plot size	-0.51	-0.082	-0.0081	0.12	0.19
	(0.057)	(0.0026)	(0.0094)	(0.019)	(0.012)
Mean Outcome	9.9	0.11	0.49	5.2	0.39
N (gridcells)	3511	87569	87569	42715	42715
N (plots)	1228	33613	33613	16474	16474

Note: This table presents RD and OLS regressions of both price and quantity outcomes on log plot size. Panel A runs RD regressions across neighbouring insula boundaries with the sample restricted 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 it's interaction with the difference in log mean plot size across insulae. This specification always controls for linear distance to the boundary between insula pairs on each side of the boundary, and insula-segment FEs. Panel B runs OLS regressions of outcomes on log plot size restricting to the same sample in Panel A. In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*Mtaa Area FEs, and amenities. Amenities are the same as described in Table 3. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table 7: Built Outcomes and Plot Sizes in 20k areas (super-insula RD)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Multiple buildings on plot
Own Larger	-0.0013	-0.00053	-0.0031	0.023
	(0.0026)	(0.011)	(0.021)	(0.015)
Own Smaller × Dist. (km)	0.053	0.21	-0.051	-0.010
	(0.017)	(0.066)	(0.12)	(0.083)
Own Larger × Dist. (km)	-0.029	0.030	0.094	-0.048
	(0.018)	(0.070)	(0.13)	(0.088)
Ln plot size	-0.066	0.026	0.18	0.21
	(0.0032)	(0.013)	(0.027)	(0.018)
Mean Outcome	0.11	0.49	5.3	0.38
N (gridcells)	92753	92753	45559	45559
N (plots)	35525	35525	17474	17474

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 76m, median 52m, 75th percentile 109m, and 95th percentile 226m. In columns 1-4 the outcomes are the same built outcomes as described in Table 4 notes. Controls always include 20K*Mtaa FEs, super-insula-segment FEs, and amenities. Amenities are the same as described in Table 3. Standard errors in parentheses are clustered by insula.

Table 8: Prices and Built Outcomes in 20k areas (OLS with amenities and planned uses)

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Ln plot size	0.49	-0.078	-0.00040	0.14	0.19
	(0.060)	(0.0026)	(0.0094)	(0.018)	(0.012)
Dist (km) paved major road	-0.15	-0.015	-0.041	-0.063	-0.040
	(0.031)	(0.0016)	(0.0071)	(0.012)	(0.0088)
Elevation (m)	0.0024	0.00089	0.0028	0.0031	0.00036
	(0.0024)	(0.000098)	(0.00043)	(0.00067)	(0.00049)
Ruggedness	-0.0089	-0.0058	-0.016	-0.011	-0.0095
	(0.022)	(0.00098)	(0.0039)	(0.0090)	(0.0052)
River/stream 100m	0.00060	-0.027	-0.11	-0.061	-0.040
	(0.17)	(0.0052)	(0.022)	(0.058)	(0.048)
Water/wetland 100m		0.0073	-0.070	-0.081	0.050
		(0.0088)	(0.031)	(0.16)	(0.22)
Z-index: 3 Ins. Characteristics	0.046	0.0031	0.017	0.0091	0.0075
	(0.027)	(0.0013)	(0.0058)	(0.010)	(0.0070)
20k edge in 100m	0.024	-0.0045	-0.012	-0.030	0.013
	(0.043)	(0.0023)	(0.0096)	(0.016)	(0.011)
Pln. recreation in 100m	-0.012	-0.00094	-0.0088	-0.011	-0.0065
	(0.040)	(0.0019)	(0.0071)	(0.012)	(0.0089)
Pln. nursery school in 100m	0.071	0.0061	0.017	0.029	0.0049
	(0.043)	(0.0026)	(0.0097)	(0.017)	(0.013)
Pln. religious site in 100m	0.037	0.0020	0.016	-0.0075	-0.0076
	(0.055)	(0.0030)	(0.012)	(0.020)	(0.015)
Pln. education in 100m	0.15	-0.0049	-0.0090	-0.026	-0.0027
	(0.074)	(0.0030)	(0.011)	(0.021)	(0.014)
Pln. service trade in 100m	-0.058	-0.0014	-0.0031	-0.0051	-0.011
	(0.092)	(0.0043)	(0.016)	(0.030)	(0.021)
Pln. housing estate in 100m	-0.13	0.0016	0.0097	0.0092	-0.036
	(0.098)	(0.0075)	(0.031)	(0.048)	(0.033)
Pln. public building in 100m	-0.0042	-0.0053	-0.010	-0.043	-0.028
	(0.085)	(0.0044)	(0.016)	(0.029)	(0.020)
Pln. cemetery in 100m	0.044	0.0042	0.038	-0.046	0.0025
	(0.13)	(0.0051)	(0.019)	(0.034)	(0.023)
Pln. any other non-res in 100m	0.14	-0.0025	-0.021	0.0053	0.022
	(0.073)	(0.0031)	(0.012)	(0.023)	(0.015)
Mean Outcome	17	0.11	0.49	5.3	0.38
N (gridcells)	4074	94789	94789	46465	46465
N (plots)	1446	36215	36215	17822	17822

Note: This table presents OLS regressions of both price and quantity outcomes on log plot size. In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*Mtaa FEs and amenities. Amenities are the same as described in Table 3. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Note that in the col 1 specification, the dummy for wetland within 100m is collinear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table 9: Built Outcomes in 20k areas (OLS with amenities and implemented uses)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Multiple buildings on plot
Ln plot size	-0.077	0.0015	0.14	0.19
	(0.0025)	(0.0093)	(0.018)	(0.012)
Dist (km) paved major road	-0.015	-0.041	-0.063	-0.039
	(0.0016)	(0.0072)	(0.012)	(0.0087)
Elevation (m)	0.00089	0.0028	0.0032	0.00033
	(0.000099)	(0.00043)	(0.00067)	(0.00048)
Ruggedness	-0.0054	-0.015	-0.0096	-0.0089
	(0.00098)	(0.0039)	(0.0090)	(0.0052)
River/stream 100m	-0.028	-0.12	-0.067	-0.047
	(0.0052)	(0.023)	(0.058)	(0.048)
Water/wetland 100m	0.0050	-0.078	-0.081	0.052
	(0.0084)	(0.031)	(0.16)	(0.23)
Z-index: 3 Ins. Characteristics	0.0032	0.017	0.0092	0.0075
	(0.0013)	(0.0057)	(0.010)	(0.0069)
20k edge in 100m	-0.0037	-0.0091	-0.030	0.011
	(0.0023)	(0.0095)	(0.016)	(0.011)
Impl. recreation in 100m	0.0064	0.0024	0.033	0.0043
	(0.0031)	(0.011)	(0.019)	(0.013)
Impl. nursery school in 100m	0.0072	0.037	0.0016	0.019
	(0.0072)	(0.028)	(0.039)	(0.026)
Impl. religious site in 100m	0.011	0.029	0.013	0.013
	(0.0043)	(0.016)	(0.027)	(0.018)
Impl. education in 100m	0.000043	0.0016	-0.027	-0.0015
	(0.0044)	(0.015)	(0.027)	(0.017)
Impl. service trade in 100m	0.016	0.050	0.052	0.057
	(0.0053)	(0.019)	(0.029)	(0.023)
Impl. housing estate in 100m	0.051	0.20	0.17	0.19
	(0.011)	(0.062)	(0.094)	(0.099)
Impl. public building in 100m	-0.0062	-0.034	-0.013	0.083
	(0.010)	(0.040)	(0.060)	(0.045)
Impl. cemetery in 100m	0.0028	0.021	-0.021	0.031
	(0.0054)	(0.018)	(0.034)	(0.024)
Impl. other non-res in 100m	-0.0052	-0.033	0.0054	0.0086
	(0.0076)	(0.025)	(0.042)	(0.025)
Impl. farming in 100m	-0.00023	0.0067	-0.011	-0.0034
	(0.0023)	(0.0087)	(0.015)	(0.010)
Impl. as residential in 100m	-0.0034	-0.0017	-0.013	-0.0049
	(0.0033)	(0.014)	(0.023)	(0.017)
Unused, kept in 100m	0.0032	0.00017	-0.021	-0.0022
	(0.0026)	(0.010)	(0.016)	(0.012)
Unused, unkept in 100m	-0.0084	-0.035	-0.042	-0.023
	(0.0027)	(0.010)	(0.017)	(0.012)
Impl. unknown non-res in 100m	-0.0017	0.13	-0.14	-0.26
	(0.060)	(0.22)	(0.041)	(0.058)
Mean Outcome	0.11	0.49	5.3	0.38
N (gridcells)	94789	94789	46465	46465
N (plots)	36215	36215	17822	17822

Note: This table presents OLS regressions of built outcomes and log plot size. In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*Mtaa FEs and amenities. Amenities include distance to major paved road, average elevation and ruggedness, a three-way Z-index of insula characteristics (rectangularity, regularity, and alignment), and dummies for within 100m of a 20k area edge, river, wetland, and each of the following implemented non-residential categories: farming, recreation, religious site, education, cemetery, service trade, nursery school, public building, housing estate, other use, unknown use, unused but kept, unused and unkept, and residential. We select these non-residential uses as controls as they have at least 100 gridcells within 100m. Standard errors in parentheses are clustered by insula.

Table 10: sorting by years of schooling within 20k areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln plot size	1.27			1.26	0.93			0.92
	(0.35)			(0.35)	(0.33)			(0.33)
Ln property value estimate		1.36				1.19		
		(0.17)				(0.18)		
Ln price			1.15				0.73	
			(0.23)				(0.27)	
Dist (km) paved major road				0.08				0.63
				(0.14)				(0.19)
Mean Outcome	14	14	14	14	14	14	14	14
Period*Source FE			✓				✓	
20K*Mtaa FE					✓	✓	✓	✓
N (gridcells)	5019	4086	1027	5019	5018	4085	1027	5018
N (plots)	1649	1352	339	1649	1648	1351	339	1648

Note: This table presents regressions of years of schooling of household heads on plot characteristics. The sample is restricted to landowning households. The outcome is always the number of schooling years completed by the head of household from the occupier survey. Columns 5-8 controls include 20K*Mtaa FEs. Columns 3 and 7 (prices) additionally control for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

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.[24].

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.

We draw on the TPDs for two purposes. First, we use them to update the planned use of

---

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

non-residential plots in our plot boundary data where the SMs are missing this information. [25] Second, we digitize planned road reserves and their widths by manually tracing the georeferenced TPDs. [26]

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. [27] 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.

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

---

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

[26] 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).

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

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.[28]

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 1.

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.

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.

---

[28] 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.4.1 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.[29]

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.[30] 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

---

[29] 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.

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

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 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

We conducted two rounds of data collection with local real estate agents. The first round took place over two months from November to December 2021. The second round took place over two months from October to November 2023.

Sampling frame

Each round of real estate agent data collection was carried out in two phases: a phone questionnaire (phase one) and a field questionnaire (phase two). First, we contacted 48 (round one) and 38 (round two) 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).[31] 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 (round one) and 29 (round two) real estate agents were targeted for the field questionnaire. However, only 12 (round one) and 4 (round two) 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 (round one) and 21 (round two) additional real estate agents meeting our criteria. This gave us a final pool of 18 (round one) and 25 (round two) real estate agents respondents.

Questionnaire details and protocol

After establishing reliable real estate agents, 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 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.

Sample Selection

In total, we collected information on 2,588 (1,326 in round one and 1,262 in round 2) transactions, including: 1,666 (1,126 and 540) on 20K plots, 311 (110 and 201) on non-20K formal plots, and 611 (90 and 521) 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

---

[31] 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

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 2,404 transactions, including: 1,507 (967 and 540) on 20K plots, 297 (96 and 201) on non-20K formal plots, and 600 (79 and 521) on non-20K informal plots.

We next match the 1,507 bareland transactions in 20k areas uniquely to 20k plots. First, we discard any transactions inside 20k areas that do not match to a planned 20k plot (these could be sub-divisions or formal or informal plots added later on), leaving us with 1,370 (890 and 480) bareland transactions. 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. This leaves us with 1,319 (843 and 476) plots with a transaction.

Finally, to get the set of plots used in our within-20k analysis, we discard plots that are non-residential. We also drop any plots where we only know the 'on sale' listing price (i.e. where no transaction had yet occured). That leaves 1,122 (664 and 458) 20k plots with transaction prices. We add 324 plots with prices recorded from our resident questionnaire to get a total sample size of 1,446 plots with a measure of transacted price.

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.[32]

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).[33] 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.[34]

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 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.

---

[32] 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.

[33] 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.

[34] 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.

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.[35]

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. [36] 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 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 prox-

---

[35] 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.

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

ies.

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. 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

$$\begin{aligned}\max_{h,z_1,k,z_2,\tau} & \int_0^\tau [\varphi lnh_1 + \beta lnz_1 + Ae^{-\theta t}]e^{-\rho t}dt + \int_\tau^\infty [\varphi ln(l^\alpha k^{1-\alpha}) + \beta lnz_2 + B]e^{-\rho t}dt \\ & + \omega\left(\int_0^\infty we^{-\delta t}dt - \int_0^\tau (ph + z_1)e^{-\delta t}dt - \int_\tau^\infty z_2 e^{-\delta t}dt - rke^{-\delta \tau} - R(0)\right), \qquad (6)\end{aligned}$$

where h 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. $h = \frac{\varphi}{\omega p}, \rightarrow h = \frac{\varphi z}{\beta p}$
3. $k = \frac{\varphi(1-\alpha)}{\omega r \delta}$
4. $\frac{w-z}{\delta} - \frac{ph}{\delta}(1 - e^{-\delta\tau}) = R(B, l_2) + rke^{-\delta\tau}$
5. $\varphi lnh + Ae^{-\theta\tau} - \varphi ln(l_2^\alpha k^{1-\alpha}) - B + \omega[\delta rk - ph] = 0$

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:

$$\begin{aligned}ln(w - \delta R) = & -\frac{1}{\alpha\varphi}[Ae^{-\theta\tau} - B - \varphi\alpha ln(\beta + \varphi(1 - \alpha e^{-\delta\tau})) \\ & + \varphi ln(\frac{\varphi}{p}) - \varphi\alpha lnl - \varphi(1-\alpha)ln(\frac{\varphi(1-\alpha)}{r\delta}) - \alpha\varphi] \qquad (7)\end{aligned}$$

$$k = \frac{\varphi(1-\alpha)(w - \delta R)}{r\delta(\beta + \varphi(1 - \alpha e^{-\delta\tau}))} \qquad (8)$$

These correspond to the conditions used to derive the comparative statics in the text.

B.2 Solving for an equilibrium

We detail a simple case where B is the same for all lots and there are lots of 2 sizes. Then we discuss generalizations.

There are a set of margins defining prices. One margin concerns the "last" person to enter, in our example the lowest income person (w_m) in 20k areas. For that person, we equate utility from staying in the city forever (consuming the same h_1 and z_1 forever) to the utility from moving. The indirect utility from staying forever is defined as $U_1 = (\varphi ln\frac{\varphi w_m}{p(\beta+\varphi)} + \beta ln\frac{\beta w_m}{\beta+\varphi})/\rho + \frac{A}{\rho+\theta}$. The utility from moving (eventually) to 20k areas comes substituting the FOC's into the expression for the present value of utility in 6 including conditions for τ and the expression for ω. This yields $U_2 = Z/\rho + (1 - e^{-(\theta+\rho)\tau})\frac{A}{\theta+\rho} - ln(\beta + \varphi(1 - \alpha e^{-\delta\tau}))\frac{\beta+\varphi(1-\alpha e^{-\delta\tau})}{\rho} + ln(w_m - \delta R_1)\frac{\beta+\varphi(1-\alpha e^{-\delta\tau})}{\rho}$. Equating the utility from staying to that for leaving gives

$$\begin{aligned} ln(w - \delta R_1) = & \frac{1}{\beta + \varphi(1 - \alpha e^{-\delta\tau})}\left[\varphi ln\frac{\varphi w}{p(\beta+\varphi)} + \beta ln\frac{\beta w}{\beta+\varphi} + \frac{A\rho}{\rho+\theta}e^{-(\theta+\rho)\tau} - Z\right] \\ & + ln(\beta + \varphi(1 - \alpha e^{-\delta\tau})) \end{aligned} \tag{9}$$

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 person w_m moving to the 20k area, we then have two equations (7, 9) in 2 unknowns, R_1 and τ. A solution is depicted in the figures in the text for the lowest income entrant. We are in essence finding the maximal R consistent with the consumer condition for τ, based on the gap between U_2 and U_1, given U_1 is exogenous to our problem.

Then there a a set of margins that define R for the marginal consumer indifferent between a smaller and larger lot and/or a higher or lower B. This is done in sequence (starting with w_m), so that the price on the smaller or lower B plot is already solved for. Here we illustrate the margin between a smaller and larger lot. If $w(N_2)$ defines the lowest income person on a larger 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 the equivalent of Eqs. 7 and 9, or by these if B remains unchanged and there are only two plot sizes. The utility from a small plot is $U(l_1, R_1, w(N_2)) = [\beta ln(\beta) + \varphi(1 - e^{-\delta\tau_1})ln(\varphi/p) + \varphi e^{-\delta\tau_1}(\alpha ln(l_1) + (1-\alpha)ln(\varphi(1-\alpha)/r\delta)) + e^{-\delta\tau}B]/\rho + ln\left(\frac{w(N_2)-\delta R_1}{\beta+\varphi(1-\alpha e^{-\delta\tau_1})}\right)[\beta + \varphi(1 - \alpha e^{-\delta\tau_1})]/\rho + \frac{A}{\rho+\theta}(1 - e^{-(\theta+\rho)\tau_1})$. 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. 7 to the relevant plot size optimization problem. Equating and simplifying we get

$$
\begin{aligned}
0 =& [\alpha[e^{-\delta\tau_2} ln(l_2) - e^{-\delta\tau_1} ln(l_1)] + [e^{-\delta\tau_2} - e^{-\delta\tau_1}][(1-\alpha) ln(\varphi(1-\alpha)/r\delta) - ln(\varphi/p)]]\varphi/\rho \\
&+ [e^{-\delta\tau_2} - e^{-\delta\tau_1}]B/\rho + [ln\left(\frac{w(N_2) - \delta R_2}{\beta + \varphi(1-\alpha e^{-\delta\tau_2})}\right)(\beta + \varphi(1-\alpha e^{-\delta\tau_2})) \\
&- ln\left(\frac{w(N_2) - \delta R_1}{\beta + \varphi(1-\alpha e^{-\delta\tau_1})}\right)(\beta + \varphi(1-\alpha e^{-\delta\tau_1}))]/\rho + \frac{A}{\rho + \theta}(e^{-(\theta+\rho)\tau_1} - e^{-(\theta+\rho)\tau_2})
\end{aligned} \tag{10}
$$

These margins that come from solving equations like 10 and the corresponding 7 to give relevant R's and τ's are depicted in the figures in the text.

B.3 Calibrating parameters for stylized equilibrium example

To illustrate equilibria, we calibrate the model to fit general patterns in the data. First we specify some parameters. We make the housing consumption share, φ, and the land share in housing production, α, both 0.3 based on the literature; and we make the 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 our equilibrium example about 0.4. Again this is consistent with urban land rent gradients (e.g., Henderson, Regan and Venables (2021)). In the example in the text, 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. 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 the lowest amenity level $B = 0.691$ and calibrate a corresponding A consistent with our data on incomes and typical time of moving. [37] That value is $A = 1.01$. That means that, at time 0, center city amenities at 1.01 are higher than those in the 20k area at 0.691.

B.4 Consumer heterogeneity

Single crossing property models like the one just presented have consumers only differing by incomes. With the advent of empirical implementation of structural models, having a stochastic component is essential. We could introduce differences in "tastes" by specifying $B = \bar{B} + \epsilon_i$. Holding income constant, people with a taste for 20k area amenities will move sooner and get more lifetime utility from living in a 20k area. While introducing ϵ_i has a simple effect on comparative statics, it does affect any characterization of the equilibrium. As in Epple and Sieg (1999), at the margin of entry

---

[37] To calibrate A we take the average household income from our survey of plot owners at \$10,000 (USD 2021). We treat them as moving at a typical time $\tau = 15$ (roughly half of our plots are developed by this time) and onto a typical plot (we take 800sqm, roughly the mean), which in our data sells for an average \$9,500 (USD 2021). Calibration involves solving FOC 1-6 in the Appendix section above for the A which satisfies these conditions given the θ and B.

to the 20k plot area there is a locus of combinations of w and ϵ's, where ϵ 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.

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. [38] 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).

---

[38] 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

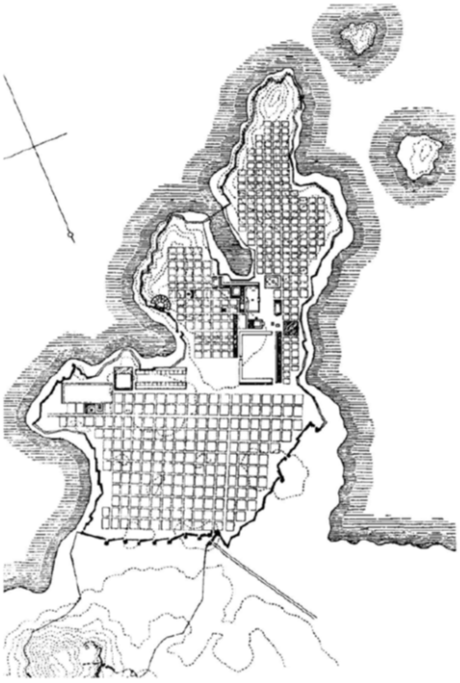
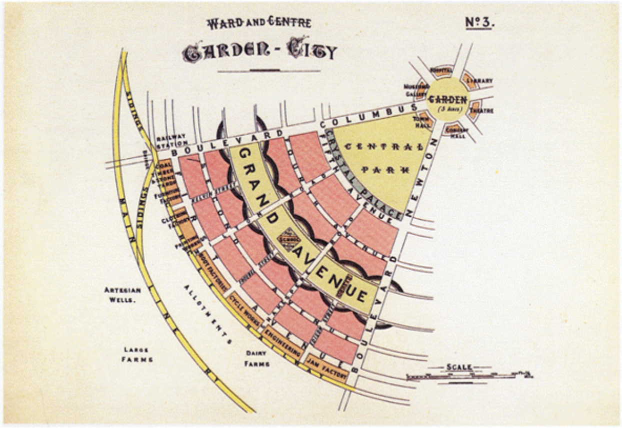


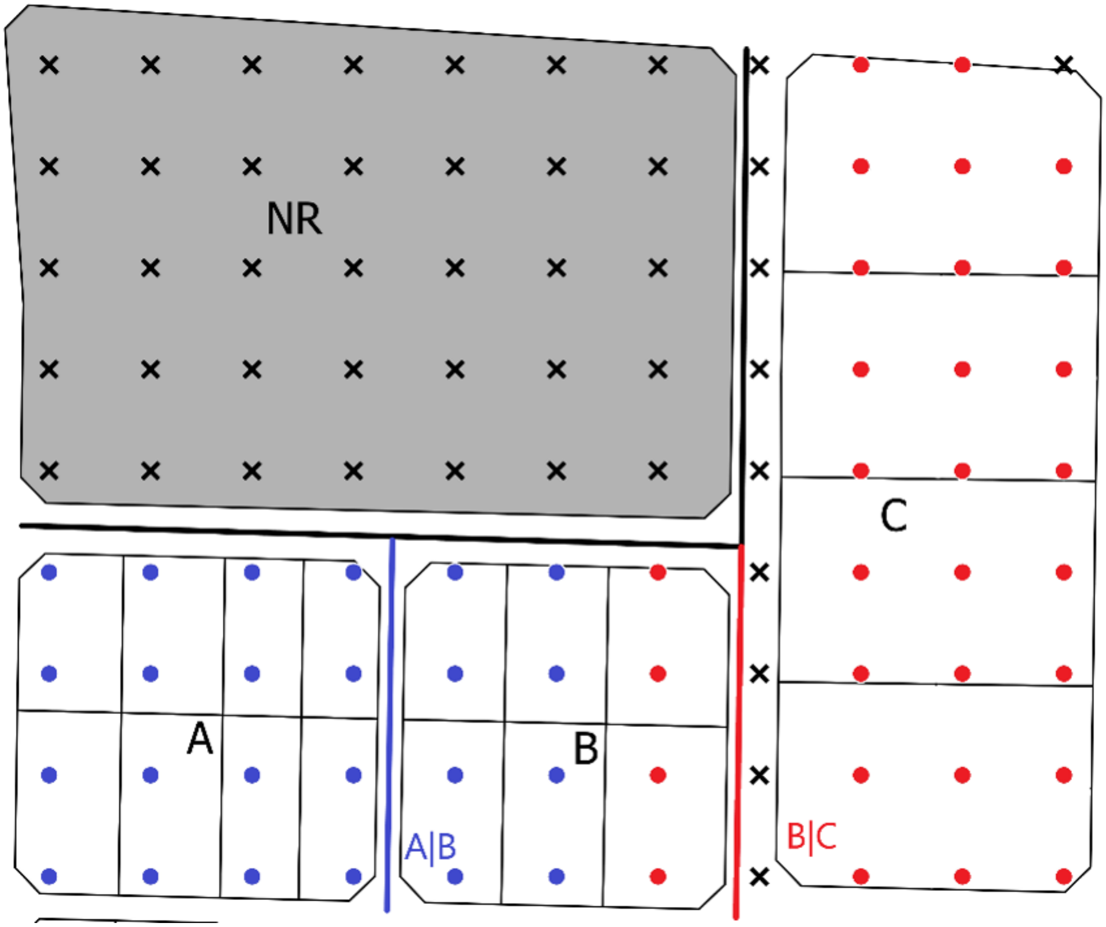
Figure 1. *Miletus.*[21]

(b) Diagram from Howard (1902) “Garden Cities of To-Morrow”



Notes: This figure plots historical examples of urban planning.

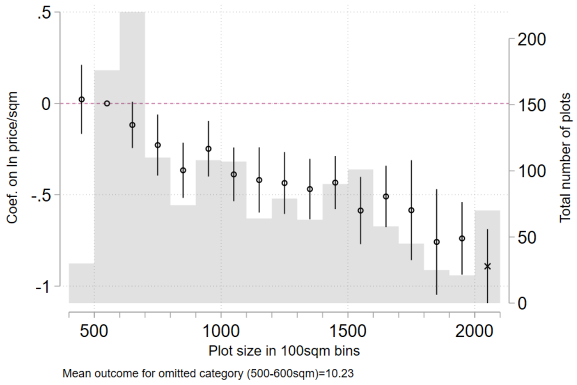
Figure A.2: Diagram of insula, plots, gridcells, and boundaries



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).

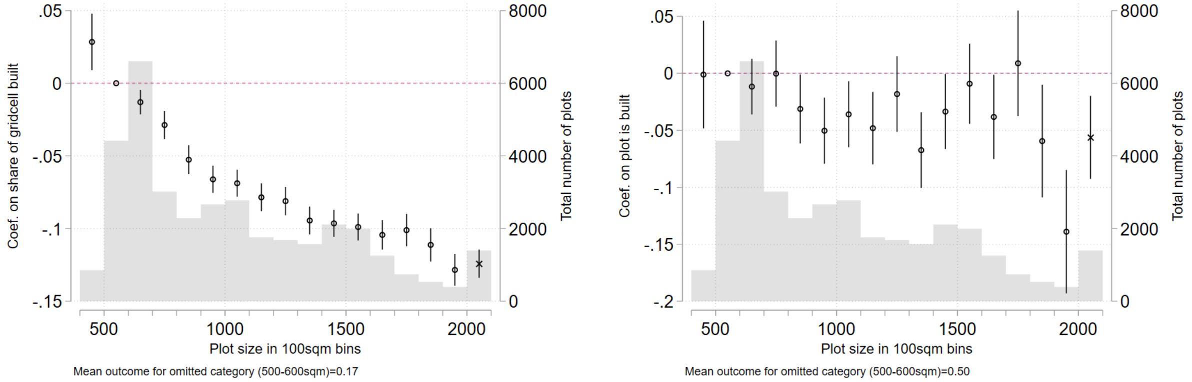
Figure A.3: Prices, built quantities, and plot size bins (OLS)

(a) ln price per sqm



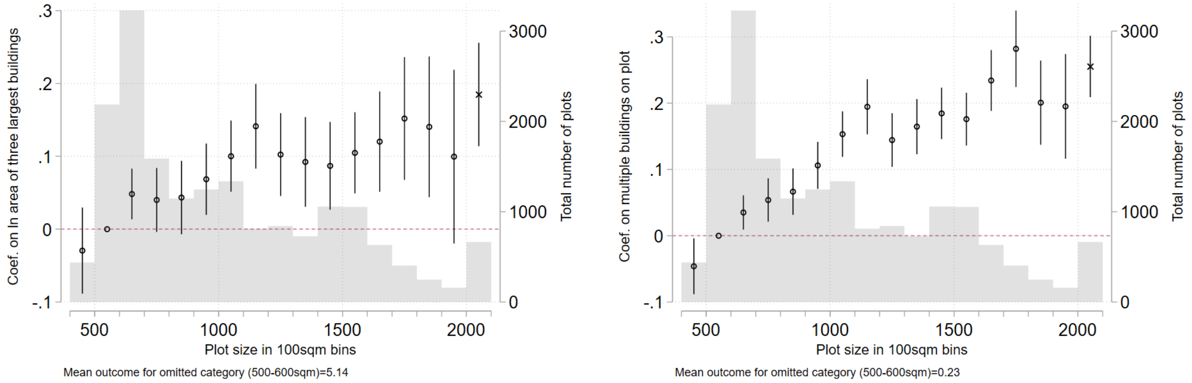
(b) Share of gridcell built

(c) Plot has a building



(d) Log area of 3 largest buildings

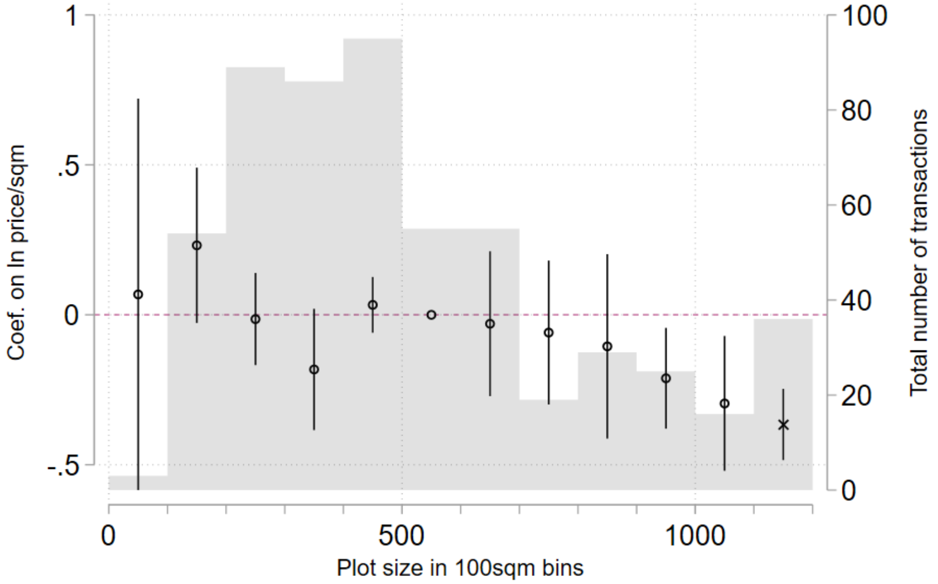
(e) Plot has multiple buildings



Notes: This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter and building quantity measures on 100sqm plot size bins (plots with size above 2000sqm are pooled into one bin, marked by an 'x'). The omitted bin is 500-600sqm. Outcomes vary by panel. In panel (a) the outcome is the log price per square meter. OUtcomes in panels (b-e) are the same as those from Table 4. Controls always include MTAA by 20K area FEs. Panel a also controls for period by source FEs. Observations are gridcells and standard errors are clustered by insula. Coefficients below 400sqm are not displayed, but are included in the regression.

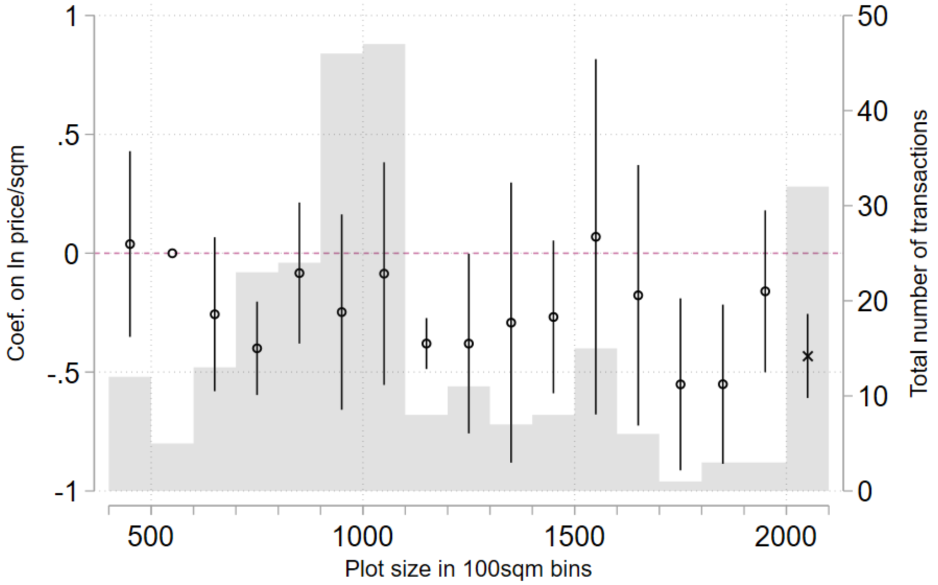
Figure A.4: Prices, built quantities, and plot size bins (OLS) in non-20k areas

(a) ln price per sqm (non-20k unsurveyed)



Mean outcome for omitted category (500-600sqm)=9.98

(b) ln price per sqm (non-20k surveyed)



Mean outcome for omitted category (500-600sqm)=10.70

Notes: This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter in non-20k areas. The omitted bin is 500-600sqm. Samples vary by panel: unsurveyed plots in non-20k areas (a), and surveyed plots in non-20k areas (b). Controls include fixed effects of transaction period, municipality, and nearest 20k area. Observations are bareland transactions and standard errors are clustered by insula.

E Appendix Tables

Table A.1: Variable Descriptions

Variable Label	Description
Ln price	Log price (2021 TZS) of plot at time of last bareland transaction. Combines real estate agent and resident questionnaire responses.
Ln price per sqm	Log price per square meter of plot.
Ln plot size	Log size of plot in square meters.
Share gridcell built	Share of the gridcell area that is built.
Plot is built	Indicator for whether the plot is built upon (contains the centroid of at least one building above 30sqm).
Log area of buildings on plot	Log total area in square meters of the three largest buildings on the plot. Constrained to built upon plots only.
Multiple buildings on plot	Indicator for multiple building centroids on the plot. Constrained to built upon plots only.
Dist (km) paved major road	Distance from gridcell to nearest major paved road in kilometers.
Elevation (m)	Gridcell average elevation in meters above sea level.
Ruggedness	Gridcell elevation ruggedness.
River/stream 100m	Gridcell is within 100m of a river or a stream.
Water/wetland 100m	Gridcell is within 100m of a water body or wetland.
Z-index: 3 Ins. Characteristics	Z index of three insula characteristics.
Z1: Rectangularity	Insula rectangularity; ratio of size of insula to size of minimum containing rectangle.
Z2: Insula alignment	Insula alignment; relative angle (modulo 90 degrees) of insula's minimum containing rectangle to the nearest other insula's.
Z3: Homogeneity	Insula homogeneity; 1 - plot size coefficient of variation within same insula.
20k boundary in 100m	Gridcell is within 100m of a 20k boundary.
Pln. recreation in 100m	Gridcell is within 100m of a planned recreation plot.
Pln. nursery school in 100m	Gridcell is within 100m of a planned nursery school plot.
Pln. religious site in 100m	Gridcell is within 100m of a planned religious (e.g., church or mosque) plot.
Pln. education in 100m	Gridcell is within 100m of a planned education (e.g., primary or secondary school) plot.
Pln. service trade in 100m	Gridcell is within 100m of a planned service or trade plot.
Pln. housing estate in 100m	Gridcell is within 100m of a planned public housing plot.
Pln. public building in 100m	Gridcell is within 100m of a planned public building plot.
Pln. cemetery in 100m	Gridcell is within 100m of a planned cemetery plot.
Pln. any other non-res in 100m	Gridcell is within 100m of a planned non-residential plot with any other use.
Impl. farming in 100m	Gridcell is within 100m of a non-residential plot implemented as farming.
Impl. recreation in 100m	Gridcell is within 100m of a non-residential plot implemented as recreation.
Impl. religious site in 100m	Gridcell is within 100m of a non-residential plot implemented as religious.
Impl. education in 100m	Gridcell is within 100m of a non-residential plot implemented as education.
Impl. cemetery in 100m	Gridcell is within 100m of a non-residential plot implemented as cemetary.
Impl. service trade in 100m	Gridcell is within 100m of a non-residential plot implemented as service or trade.
Impl. nursery school in 100m	Gridcell is within 100m of a non-residential plot implemented as nursery school.
Impl. other non-res in 100m	Gridcell is within 100m of a non-residential plot implemented as some other use.
Impl. public building in 100m	Gridcell is within 100m of a non-residential plot implemented as public building
Impl. housing estate in 100m	Gridcell is within 100m of a non-residential plot implemented as public housing.
Impl. unknown non-res in 100m	Gridcell is within 100m of a non-residential plot implemented as an unkown use.
Unused, kept in 100m	Gridcell is within 100m of a non-residential plot left empty, but well kept.
Unused, unkept in 100m	Gridcell is within 100m of a non-residential plot left empty, but not well kept.
Impl. as residential in 100m	Gridcell is within 100m of a non-residential plot implemented as residential.

Note: This table describes variables.

Table A.2: Summary Statistics

	Full sample (94789 gridcells, 36215 plots)				Price sample (4074 gridcells, 1446 plots)			
	mean	sd	min	max	mean	sd	min	max
Ln price					17.07	0.89	12.23	19.29
Ln price per sqm					9.96	0.88	5.26	12.25
Ln plot size	7.04	0.45	5.45	8.29	7.10	0.44	5.46	8.25
Share gridcell built	0.11	0.17	0.00	1.00	0.11	0.18	0.00	0.94
Plot is built	0.49	0.50	0.00	1.00	0.48	0.50	0.00	1.00
Log area of buildings on plot (if built)	5.26	0.68	3.40	8.12	5.43	0.68	3.42	6.92
Multiple buildings on plot (if built)	0.38	0.49	0.00	1.00	0.45	0.50	0.00	1.00
Dist (km) paved major road	2.05	1.30	0.01	6.90	2.04	1.29	0.02	6.36
Elevation (m)	48.96	28.07	2.50	111.00	56.70	27.20	7.00	108.50
Ruggedness	0.37	0.96	0.00	18.00	0.34	0.90	0.00	12.50
River/stream 100m	0.02	0.15	0.00	1.00	0.01	0.11	0.00	1.00
Water/wetland 100m	0.00	0.04	0.00	1.00	0.00	0.00	0.00	0.00
Z-index: 3 Ins. Characteristics	0.00	0.79	-2.77	1.08	0.01	0.79	-2.76	1.07
Z1: Rectangularity	0.00	1.00	-3.64	1.06	-0.01	1.02	-3.47	1.05
Z2: Insula alignment	-0.00	1.00	-3.07	0.78	0.06	0.95	-3.07	0.78
Z3: Homogeneity	-0.00	1.00	-6.58	1.42	-0.02	1.06	-6.58	1.42
20k boundary in 100m	0.22	0.42	0.00	1.00	0.23	0.42	0.00	1.00
Pln. recreation in 100m	0.37	0.48	0.00	1.00	0.34	0.47	0.00	1.00
Pln. nursery school in 100m	0.12	0.33	0.00	1.00	0.14	0.35	0.00	1.00
Pln. religious site in 100m	0.09	0.29	0.00	1.00	0.10	0.30	0.00	1.00
Pln. education in 100m	0.09	0.28	0.00	1.00	0.08	0.27	0.00	1.00
Pln. service trade in 100m	0.06	0.23	0.00	1.00	0.06	0.24	0.00	1.00
Pln. housing estate in 100m	0.03	0.16	0.00	1.00	0.05	0.22	0.00	1.00
Pln. public building in 100m	0.06	0.25	0.00	1.00	0.06	0.24	0.00	1.00
Pln. cemetery in 100m	0.05	0.21	0.00	1.00	0.04	0.19	0.00	1.00
Pln. any other non-res in 100m	0.25	0.43	0.00	1.00	0.26	0.44	0.00	1.00
Impl. farming in 100m	0.23	0.42	0.00	1.00	0.26	0.44	0.00	1.00
Impl. recreation in 100m	0.09	0.29	0.00	1.00	0.10	0.30	0.00	1.00
Impl. religious site in 100m	0.05	0.22	0.00	1.00	0.05	0.21	0.00	1.00
Impl. education in 100m	0.05	0.21	0.00	1.00	0.06	0.24	0.00	1.00
Impl. cemetery in 100m	0.03	0.16	0.00	1.00	0.02	0.15	0.00	1.00
Impl. service trade in 100m	0.03	0.18	0.00	1.00	0.04	0.19	0.00	1.00
Impl. nursery school in 100m	0.02	0.13	0.00	1.00	0.01	0.12	0.00	1.00
Impl. other non-res in 100m	0.02	0.14	0.00	1.00	0.02	0.15	0.00	1.00
Impl. public building in 100m	0.01	0.09	0.00	1.00	0.01	0.08	0.00	1.00
Impl. housing estate in 100m	0.00	0.05	0.00	1.00	0.00	0.04	0.00	1.00
Impl. unknown non-res in 100m	0.00	0.03	0.00	1.00	0.00	0.00	0.00	0.00
Unused, kept in 100m	0.15	0.36	0.00	1.00	0.16	0.37	0.00	1.00
Unused, unkept in 100m	0.15	0.35	0.00	1.00	0.13	0.34	0.00	1.00
Impl. as residential in 100m	0.07	0.25	0.00	1.00	0.06	0.24	0.00	1.00

Table A.3: Land price size interactions inside and nearby 20k areas

	(1) Ln Price	(2) Ln Price	(3) Ln Price	(4) Ln Price
Ln plot size	0.71	0.69	0.64	0.63
	(0.054)	(0.041)	(0.048)	(0.036)
Non-20K Surveyed	-0.23	-0.27	-2.53	-2.44
	(0.16)	(0.12)	(0.67)	(0.46)
Non-20K Unsurveyed	-0.70	-0.71	-1.12	-0.98
	(0.099)	(0.079)	(0.81)	(0.78)
Non-20K Surveyed × Ln plot size			0.33	0.31
			(0.081)	(0.054)
Non-20K Unsurveyed × Ln plot size			0.062	0.037
			(0.12)	(0.11)
Mean Outcome	17	17	17	17
20K or Nearest FE		✓		✓
N	2074	2074	2074	2074

Note: This table presents regressions of log price on log plot size (Dalali estimates) and planned/surveyed status. The outcome is always the log price of a bare land transaction from the Dalali survey. Each observation is a transaction which took place inside or nearby 20k areas. The sample is made of 1246 transactions inside 20K areas, 266 outside 20K areas and surveyed, and 562 outside 20K areas and unsurveyed. Controls include fixed effects for Municipality (Ilala, Temeke, Kigamboni, Kinondoni) and transaction time period (2023, 2022, 2021 2019-20, 2016-18, 2011-15, and pre-2010). Note that there are no 20k areas in Ubungo, the fifth Municipality in Dar es Salaam. Columns 2 and 4 additionally include fixed effects for the the nearest 20k area (own 20k area for transactions inside 20k areas and the nearest 20k area for transactions in non-20k areas). Columns 3 and 4 include interaction terms between plot size and planned/surveyed status. Standard errors in parentheses are clustered by 20K area.

Table A.4: Balance on natural amenities (RD)

	(1)	(2)	(3)	(4)	(5)
	Elevat.	Rugged.	River or stream in 100m	Water or wetland in 100m	Z-index c. 1-4
Panel A: all insula pairs					
Own Larger	-0.0066	-0.033	-0.00037	0.0015	0.00058
	(0.054)	(0.016)	(0.0019)	(0.0012)	(0.0087)
Mean Outcome	50	0.35	0.022	0.0019	0.015
N (gridcells)	87569	87569	87569	87569	87569
N (plots)	33613	33613	33613	33613	33613
Panel B: gap≥400sqm					
Own Larger	-0.12	-0.047	0.012	0.00017	-0.010
	(0.13)	(0.040)	(0.0050)	(0.00061)	(0.015)
Mean Outcome	50	0.44	0.029	0.0022	-0.0167
N (gridcells)	22483	22483	22483	22483	22483
N (plots)	9066	9066	9066	9066	9066
Panel C: gap<100sqm					
Own Larger	-0.018	-0.0039	-0.0054	0.0012	0.0030
	(0.075)	(0.019)	(0.0017)	(0.00095)	(0.0079)
Mean Outcome	50	0.29	0.015	0.0016	0.049
N (gridcells)	30460	30460	30460	30460	30460
N (plots)	15157	15157	15157	15157	15157

Note: This table presents RD regressions across neighbouring insula boundaries testing for balance on natural amenities. All panels restrict the sample to within 100m of the insula-pair boundary. Panel B further restricts to insula pairs with at least 400sqm gap in mean plot size, and Panel C to those insula pairs with no more than 100sqm gap. 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. In columns 1-4 the outcomes are natural amenities or disamenities: elevation in metres, ruggedness, an indicator for a river or stream within 100m, and an indicator for water or wetland within 100m. Column 5 is a z-index of the four outcomes in columns 1-4, where elevation (an amenity) is positive and the other three (disamenities) are negative. Controls include 20K*Mtaa FEs and insula-segment FEs. Standard errors in parentheses are clustered by insula.

Table A.5: Robustness: exclude ‘pre-built’ plots for RD and OLS

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Panel A: OLS, drop pre-built gridcells					
Ln plot size	0.48	-0.082	-0.010	0.14	0.19
	(0.064)	(0.0029)	(0.010)	(0.020)	(0.013)
Mean Outcome	17	0.11	0.50	5.3	0.38
N (gridcells)	3323	76666	76666	38042	38042
N (plots)	1182	29370	29370	14779	14779
Panel B: OLS full sample					
Ln plot size	0.49	-0.078	-0.00040	0.14	0.19
	(0.060)	(0.0026)	(0.0094)	(0.018)	(0.012)
Mean Outcome	17	0.11	0.49	5.3	0.38
N (gridcells)	4074	94789	94789	46465	46465
N (plots)	1446	36215	36215	17822	17822
Panel C: RD, drop pre-built gridcells					
Own Larger	-0.16	-0.018	0.0053	-0.0043	0.031
	(0.060)	(0.0028)	(0.0091)	(0.020)	(0.014)
Mean Outcome	10	0.11	0.50	5.3	0.38
N (gridcells)	2854	69868	69868	34656	34656
N (plots)	996	26858	26858	13522	13522
Panel D: RD full sample					
Own Larger	-0.17	-0.017	0.011	-0.018	0.026
	(0.055)	(0.0025)	(0.0080)	(0.018)	(0.013)
Mean Outcome	9.9	0.11	0.49	5.2	0.39
N (gridcells)	3511	87569	87569	42715	42715
N (plots)	1228	33613	33613	16474	16474

Note: This table presents robustness to dropping gridcells that had a pre-treatment building. Panels A and C drop all gridcells belonging to plots that contain the centroid of a pre-treatment building. Panels A and B follow the same specifications as Table 4 Panel B. Panels C and D follow the same specifications as Table 5 Panel A.

Table A.6: Population density by plot size

	Mean pop. per built res. plot	Share of Plots Built	Mean pop. per res. plot	Mean plot size (sqm)	Pop. dens residential (ppl/sqkm)	Pop. dens (ppl/sqkm)
Small Plots (≤800sqm)	5.3	0.50	2.6	629	4166	2083
Medium Plots (800-1600sqm)	5.4	0.49	2.6	1179	2232	1116
Large Plots (≥1600sqm)	5.6	0.49	2.7	1961	1392	696
All Plots	5.4	0.49	2.7	1040	2552	1276

Note: This table presents population statistics by plot size 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 our household questionnaire captures a representative sample of 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.7: Built Outcomes and Plot Sizes in 20k areas (super-insula RD by sizes)

	(1) Share gridcell built	(2) Plot is built	(3) Log area of buildings	(4) Multiple buildings on plot
Sml-Med × Own Larger	-0.010	-0.012	0.019	0.034
	(0.0039)	(0.013)	(0.025)	(0.018)
Sml-Lrg × Own Larger	-0.027	-0.14	-0.11	0.038
	(0.011)	(0.036)	(0.093)	(0.049)
Med-Lrg × Own Larger	0.0053	0.015	-0.018	0.021
	(0.0033)	(0.017)	(0.033)	(0.024)
Sml-Med × Own Larger × Dist. (km)	-0.019	0.025	0.050	-0.013
	(0.020)	(0.072)	(0.13)	(0.091)
Sml-Lrg × Own Larger × Dist. (km)	0.17	1.82	0.90	-0.47
	(0.22)	(0.58)	(1.08)	(0.58)
Med-Lrg × Own Larger × Dist. (km)	0.031	-0.10	0.18	-0.25
	(0.041)	(0.23)	(0.51)	(0.34)
Sml-Med × Own Smaller × Dist. (km)	0.13	0.27	0.21	0.16
	(0.033)	(0.11)	(0.14)	(0.11)
Sml-Lrg × Own Smaller × Dist. (km)	0.057	0.23	0.21	-0.30
	(0.074)	(0.23)	(0.31)	(0.25)
Med-Lrg × Own Smaller × Dist. (km)	0.026	0.21	-0.19	-0.082
	(0.020)	(0.086)	(0.17)	(0.12)
Ln plot size	-0.059	0.041	0.19	0.20
	(0.0034)	(0.014)	(0.029)	(0.019)
Mean Outcome	0.11	0.49	5.3	0.38
N (gridcells)	92753	92753	45559	45559
N (plots)	35525	35525	17474	17474

Note: This table presents RD regressions across neighbouring super-insula boundaries broken down by super-insula-pair size categories (small-medium, medium-large, and small-large). 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 76m, median 52m, 75th percentile 109m, and 95th percentile 226m. In columns 1-4 the outcomes are the same built outcomes as described in Table 4 notes. Controls always include 20K*MTAA FEs, super-insula-segment FEs, and amenities. Amenities are the same as described in Table 3. Standard errors in parentheses are clustered by insula.

Table A.8: Effects of ln plot size and share neighbors small (gridcell level)

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Ln plot size	0.42	-0.052	0.028	0.16	0.19
	(0.097)	(0.0037)	(0.016)	(0.032)	(0.022)
Shr. Neighbs Small	0.34	0.28	0.37	0.33	-0.17
	(1.25)	(0.052)	(0.21)	(0.47)	(0.27)
Ln plot size × Shr. Neighbs Small	-0.077	-0.037	-0.051	-0.045	0.025
	(0.19)	(0.0077)	(0.030)	(0.071)	(0.040)
Mean Outcome	17	0.11	0.49	5.3	0.38
N (gridcells)	4074	94785	94785	46464	46464
N (plots)	1446	36215	36215	17822	17822

Note: This table presents OLS regressions of both price and quantity outcomes on log plot size and the fraction of small neighbouring plots (within 100m). In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*MTAA FEs and amenities. Amenities are the same as described in Table 3. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Note that in the col 1 specification, the dummy for wetland within 100m is collinear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table A.9: Prices and Built Outcomes in 20k areas (OLS with insula z-index broken out)

	(1) Ln Price	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Ln plot size	0.48	-0.083	-0.015	0.12	0.19
	(0.059)	(0.0026)	(0.0095)	(0.018)	(0.012)
Insula rectangularity (standardized)	0.0060	0.0059	0.025	0.012	0.0012
	(0.027)	(0.0014)	(0.0057)	(0.011)	(0.0067)
Insula alignment (standardized)	0.020	0.0019	0.0057	0.013	0.0099
	(0.022)	(0.0011)	(0.0043)	(0.0084)	(0.0055)
Insula regularity (standardized)	0.029	-0.0028	-0.0080	-0.011	-0.0024
	(0.027)	(0.0012)	(0.0049)	(0.0095)	(0.0059)
Mean Outcome	17	0.11	0.49	5.3	0.38
N (gridcells)	4074	94789	94789	46465	46465
N (plots)	1446	36215	36215	17822	17822

Note: This table presents OLS regressions of both price and quantity outcomes on log plot size, breaking out the insula z-index control into it's three respective parts. In column 1 the outcome is log price per square metre on the plot, and columns 2-5 are the same built outcomes as described in Table 4 notes. Controls always include 20K*MTAA FEs and amenities. Amenities are the same as described in Table 3, except that the z-index for insula characteristics are broken out into their three component parts. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Note that in the col 1 specification, the dummy for wetland within 100m is collinear with other controls, and so dropped from the regression. Standard errors in parentheses are clustered by insula.

Table A.10: Non-residential plots implementation vs plan ratios and simulations

	(1) Observed Ratio	(2) Perfect Ratio	(3) Random Implementation Median	(4) 95-pct	(5) 99-pct	(6) N Plots Plan	(7) Impl.
recreation	2.9	3.4	.99	1.2	1.3	411	96
nursery school	5.6	9.4	.86	2.1	2.6	148	22
religious	6.1	11	.99	1.6	1.9	131	86
education	10	20	.77	1.9	2.7	71	51
service trade	2.6	9.8	.87	1.7	2.2	143	45
housing estate	11	23	0	11	11	61	2
public buildings	2.7	6.7	.67	2	2.7	209	10
cemetery	24	24	.79	2.4	3.2	57	31
Weighted average	5.3	10	.96	1.5	1.8	·	·
Total	·	·	·	·	·	1,231	343

Note: This table presents estimates for each planned use of the ratio of probabilities: P(implemented as use j given planned as use j)/P(implemented as use j). The sample is non-residential plots, and any plot with unknown planned use or unknown implemented use is dropped. Column 1 gives the ratio based on observed shares. Column 2 is based the counterfactual of all plots implemented exactly as planned [here the ratio simplifies to 1/P(planned as j)]. This is equivalent to the counterfactual where implementation only occurs where planned, even if not fully [i.e. P(implemented and planned as use j)=P(implemented as use j)]. Columns 3-5 are based on 10000 random draws of plot implementation, holding the aggregate implementation rates at their observed rates. For these draws we report the median (col. 3), the 95th percentile (col. 4), and the 99th percentile (col. 5) of the ratio. Columns 6 and 7 give the number of plots planned and implemented in the data. Each of the top eight rows represent a specific landuse and the bottom two rows represent the average ratio weighted by the proportion of plots planned in use j, and the total plot counts.

Table A.11: Settlement dates, plot sizes, and amenities

	(1) Year Building Occupied	(2) Year Building Occupied	(3) Year Construction Started	(4) Year Construction Started
Ln plot size	-0.71	-0.91	-0.55	-0.69
	(0.43)	(0.43)	(0.38)	(0.39)
Years of schooling		0.24		0.14
		(0.037)		(0.036)
Dist (km) paved major road	0.83	0.68	0.64	0.59
	(0.29)	(0.28)	(0.29)	(0.28)
Mean Outcome	2,014	2,014	2,010	2,010
N (gridcells)	4907	4893	4237	4230
N (plots)	1611	1606	1383	1381

Note: This table presents OLS regressions where the outcome is the year in which a plot was built. The sample is restricted to landowning households. In columns 1-2 the outcome is the year the plot was first residentially occupied, and columns 3-4 is the year construction first started on the plot. Controls always include 20K*Mtaa FEs and amenities. Amenities are the same as described in Table 3 (though only the coefficient on distance to major paved road is displayed), except that the dummy for water or wetland within 100m is perfectly collinear with the other controls, and so dropped from all the regressions. Columns 2 and 4 additionally control for the years of schooling of the household head. Standard errors in parentheses are clustered by insula.