# Economics of Greenfield Urban Planning[*]

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

October 6, 2025

**Abstract**

Urban planning has shaped cities for millennia, demarcating property rights and mitigating coordination failures, but its rigidities often conflict with market-driven development. Although planning is common in high-income countries, rapidly growing cities in the developing world are characterized by urban informality. Greenfield urban planning is a key option, but we lack economic theory and evidence to evaluate planners' choices. This paper presents a dynamic model to evaluate the effects of plot sizes and amenities on consumer outcomes. This framework is applied to a flagship project in Dar es Salaam that subdivided peri-urban land into more than 36,000 formal plots, which people purchased and built homes on. We assemble a novel dataset using administrative records, satellite imagery, and primary surveys. Informed by the model, we study the effects of planning choices using within-neighborhood variation and spatial regression discontinuities. We find that by securing property rights and local road access, the project doubled land values relative to nearby unplanned areas. Connectivity to the city is prized, as evidenced by price appreciation and construction rate differences between and within areas. The price elasticity of bare land to plot size is -0.5, suggesting an oversupply of large plots despite the sorting of highly educated owners into the project and its larger plots. In contrast to connectivity and plot size, other planning choices, such as intended non-residential land uses and plot configurations, matter less. Counterfactual analysis using the estimated structural model shows that while land value maximization provides larger plots, welfare maximization provides smaller plots serving more low-income people.

KEYWORDS: Urban Planning, Economic Development, Africa.

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

[*]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, Maisy Wong, Jenny Schuetz, Victoria Wenxin Xie, and Tom Cui, and many colleagues, especially Shlomo Angel, Joseph Kironde, Gabriel Kreindler, Clemence Mero, David Msangi, and Tony Yezer, and seminar and conference participants at the Universities of Chicago, Galway, Kyoto, Peking, Pennsylvania (Wharton), Syracuse, and Tokyo, HEC Lausanne, HKUST, Sorbonne, UQAM, USC, Bank of Italy, the Online Spatial and Urban Seminar, CEPR Paris Symposium, Harvard Cities and Development workshop, 16th Conference on Regional and Urban Economics, 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 shapes land use in developed country cities, where typically about half of the land is in public use and private land is regulated by zoning (e.g., Bertaud, 2018; American Planning Association, 1950). In contrast, in developing countries, and especially in Africa's large and growing cities, planning is often ineffective (e.g., Castells-Quintana, 2017; Henderson et al., 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 urban neighborhoods in developing countries are crucial.

A common approach is 'de-novo' urban planning, where greenfield agricultural land on urban fringes is purchased and partitioned into formally surveyed and titled plots with roads and occasionally some utilities and services. People can then buy plots and build on them. This de-novo approach was pursued in many developing countries under the World Bank "Sites and Services" agenda during the 1970s and 1980s, and several African governments have since implemented similar strategies (Centre for Affordable Housing Finance, 2024). Evidence from Tanzania indicates that the de-novo approach was cost effective in the *long-run*, promoting higher quality housing and land values compared to neighboring unplanned settlements and slums that were later upgraded (Michaels et al., 2021).

De-novo planners face many possible choices in allocating greenfield land to plots of different uses, sizes, and configurations. While there is a related literature in economics, which we review below, what is missing is a comprehensive approach for evaluating the effects of planners' choices within de-novo areas. This gap in our ability to assess de-novo planning is glaring, since urban planning has shaped cities for millennia and is currently being taught in hundreds of universities worldwide (Symonds, 2023). This paper starts to fill this gap. We combine theory, novel data, reduced-form evidence, structural estimates, and counterfactuals to study a flagship de-novo project implemented in Tanzania in the early 2000s. Our analysis highlights the importance of property rights, access, and plot size, as well as the consequences of owner sorting. In contrast, other planning decisions, including those about planned amenities and plot configurations, seem less consequential.

We contribute to the literature in four ways. The first is a framework for evaluating planners' decisions on the sizes and amenities of plots within de-novo areas. This model accounts for the sorting of owners with different incomes between and within areas, the speed of neighborhood development, land values, and consumer welfare. Second, we assemble new data, including detailed project maps showing the locations of residential and non-residential plots, roads, and planned amenities; primary surveys with local leaders about their areas, estate agents about individual market sales of bare land, and residents about their bare land acquisitions, education, and investments in their plots; and high-resolution satellite imagery, which we use to examine building footprints. Third, we estimate causal effects of planning treatments using OLS and spatial regression discontinuity (RD) designs, focusing on variation within neighborhoods. Finally, since our findings suggest significant misallocation of the plot size distribution by the planners, we estimate a structural version of our model of the development of de-novo areas to construct different counterfactual distributions that maximize first welfare and then land values.

The context of our investigation is the "20,000 Plots" project (that we refer to as "20k"), which the Tanzanian government implemented around 2000-2005. This project delivered more than 36,000 residential de-novo plots (almost twice the number originally intended) in 12 project areas on the fringes of Dar es Salaam. We find that the government's de-novo investment was rapidly recouped using the purchasers' payments. The real value of bare land in 20k then increased sharply and is now roughly *twice* as high as in nearby informal areas (for plots of the same size). Our findings indicate that the 20k plots' price premium reflects more secure individual property rights and better local road access, compared to unplanned informal areas. Thus, by defining and protecting property rights and access, planning resolves costly coordination failures that afflict informal areas.

Our model predicts that 20k areas attract higher-income owners and those owners sort into larger and higher-amenity plots, and our empirical evidence confirms this. Such sorting would have occurred even if the plot sizes and amenities had been initially randomly assigned across space. As we discuss in later sections, this shapes the interpretation of our reduced-form estimates, which capture the total marginal effects of planning decisions. For example, reallocating a tiny land parcel from a small plot to a geographically proximate large plot affects parcel-level outcomes due to differences by plot size in (i) owner valuation (holding income constant) and (ii) owner income. Such total marginal effects are informative for incremental changes that planners or individuals may consider, for assessing which planning decisions matter, and for evaluating efficiency. Our setting is ideal for identifying these effects due to its greenfield nature and our ability to make comparisons within small localities.

Intuitively, our spatial RD estimates of plot size effects compare different-sized plots on opposite sides of roads, keeping amenities constant. We show that the resulting estimates are very similar to those using OLS with small-area controls. For example, both methods give a price elasticity per sqm with respect to plot size of -0.5. This implies that land in larger plots is less valuable, *despite* belonging to more educated (and richer) owners. As we discuss below, this is consistent with an over-provision of large plots, likely in part due to persistent colonial-era planning standards and norms.

Our analysis of amenities in 20k areas reveals that those which relate to access are the most important. The remotest project areas saw the slowest rates of construction and land value appreciation. And our within-area regressions similarly show that plots that are further from preexisting main paved roads have significantly lower land prices and lower rates of construction.

Some other amenities also matter, but less than access. We find that a more gridded layout, higher elevation, and greater distance from water (in a flood-prone city) lead to higher construction rates. Planned non-residential amenities are generally ignored, due to low rates of actual provision; however, actual provision, where it does take place, correlates positively with construction rates. Finally, our results provide little evidence of plot-size externalities. Smaller plots (though not larger ones) that are bunched together with other small plots are more likely to be built on and built on more intensively, although the magnitudes of such effects are modest.

To assess how planning could have been improved, we use our data to estimate a structural version of our model. We note that ensuring private property rights and access, while valuable, entails a rigidity, where plot sizes and layout cannot be readily reconfigured ex post to accommodate market needs. Therefore, it is important to assess the effects of changing the initial distribution of formal plot sizes on sorting, prices, housing investment, and plot development rate. Counterfactual analysis using the structural model shows that land value maximization entails increasing plot sizes even further, but welfare maximization entails providing more small plots to accommodate many more people, extending ownership of formal plots to more lower-income people.

The model offers a sufficient statistic for evaluating greenfield urban planning efficiency in contexts like ours: land price per sqm should increase (modestly) in plot size. Otherwise, splitting increases welfare. This statistic would allow planners to spot inefficient plot size distributions that may arise from applications of simple rules-of-thumb.

Our paper relates to longstanding debates on the respective roles of planners and markets in determining the allocation of land. 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. Early economics work recognized the role of planning in accounting for externalities (Davis and Whinston, 1962, 1964) and allowing space for roads (Solow and Vickrey, 1971; Dixit, 1973).[1] Recent work (e.g. 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 and distributional issues. Although urban planners and economists could learn from each other how to improve city design, such mutual learning is limited (Bertaud, 2018). But the stakes are high, as cities concentrate a large and growing share of the world's population and play outsized roles in the global economy (Glaeser, 2012; Moretti, 2012).

We contribute to a literature on land use regulation and zoning in cities – primarily in high-income countries (Glaeser and Ward, 2009; Turner et al., 2014; Gyourko and Krimmel, 2021; Chiumenti et al., 2022) but also in low-income countries (Anagol et al., 2021; Nagpal and Gandhi,

[1] 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).

2024). We extend this literature by studying many different planning decisions and isolating the effects of plot sizes from other factors with which they are often bundled. Our de-novo setting is important, as previous research has demonstrated that zoning itself is influenced by prior development (Shertzer et al., 2018). A related literature studies the costs of overly segmented plots near large city centers (Harari and Wong, 2024; Yamasaki et al., 2023), but we focus on a suburban setting, where large tracts of land are abundant. Consistent with research on suburban areas in high-income countries (e.g., Combes et al., 2021; Larson and Shui, 2022), we find that unit land prices decrease with plot size. By studying a greenfield setting at a fine granular spatial scale, we provide tighter evidence of causal effects of plot size on prices and construction.

We connect to the literature on institutions and development. This includes studies of colonial institutions' impact on current economic development (Acemoglu et al., 2001; Baruah et al., 2021), which we add to by studying the effects of residential planning regulations carried over from British colonial rule, which emphasized large plot sizes. Related is a descriptive literature on the prevalence of large minimum plot sizes in Africa (e.g., Gulyani and Connors, 2002; Collier and Venables, 2014; Tipple, 2015), to which we contribute by studying plot size effects in a quantitative economic framework. Our quantitative work also adds to the descriptive literature on de-novo planning in Tanzania (Tiba et al., 2005; Mwiga, 2011; Kironde, 2015). Important contributions have emphasized the value of combining planning with property right protection (De Soto, 1989; Libecap and Lueck, 2011; Angel, 2012), and case studies in Tanzania have explored the role of property titles and plot demarcations in securing tenure (Manara and Regan, 2024, 2025). We look inside formal areas and unpack the effects of specific planning decisions from those due to property rights.

Finally, aspects of our paper are influenced by the literature on the valuation of local neighborhood amenities and the implications of sorting (Epple and Sieg, 1999; Bayer et al., 2007; Gechter and Tsivanidis, 2023; Almagro and Dominguez-Iino, 2024). A related literature studies residential patterns and local access to public service provision in developing country cities (Adukia et al., 2024; Harari, 2024). We shed light on sorting that follows de-novo planning and we examine planned amenities of different types.

The remainder of our paper is organized as follows. Section 2 discusses the institutional background; Section 3 discusses the data; Section 4 presents a framework for evaluating reduced-form empirical results and Section 5 presents our reduced-form estimation results; Section 6 estimates the model and conducts counterfactuals; and Section 7 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 (Smith, 2007), as did some ancient cities in Mesopotamia, Assyria, and Egypt (Paden, 2001). Ancient Greek cities initially developed organically, but in the fifth century BCE Hippodamus was credited with designing Miletus and Piraeus (Athens's port) with gridded layouts (Paden, 2001). Miletus (panel A of Figure A.1) had grids of two sizes and public spaces with several public buildings. The use of planned gridded cities spread across the Ancient World under the empires of Alexander the Great and Rome. More than two millennia later, Howard (1902) set out de-novo plans for "garden cities" (panel B of Figure A.1), which influenced suburban planning in many countries (Hall and Tewdwr-Jones, 2019). Today, urban planning is studied in graduate programs worldwide (Symonds, 2023), which combine planning theory, policy debates, planning standards, and case studies. However, systematic economic evaluation of planning standards and schemes is rare (Bertaud, 2018), especially for de-novo planning.

## 2.2 De-novo planning in Tanzania and elsewhere

Kironde (1994) and MLHHSD (2018) provide an overview of urban planning in Dar es Salaam. Under German and later British colonial rule, the European core had a grid, strict planning standards, and large plots; the Asian parts were planned, but with lower standards; and the African parts were unplanned. All of this promoted segregation. After Tanzania's independence in 1961, Dar es Salaam's population grew from less than 280,000 in 1967 to about 5.4 million in 2022 (United Republic of Tanzania, 2022). Formal planning standards were retained in theory, sometimes with new justifications, and a series of masterplans were developed. In practice, however, most of the city comprises informal settlements.

From the 1970s, some de-novo planned neighborhoods were developed, notably through "Sites and Services" in collaboration between the Tanzanian government and the World Bank. Similar projects were developed in Indonesia, Vietnam, Myanmar, Uganda, Kenya, Nigeria, Ethiopia, Egypt, India, and Latin America (Grimes, 1976; Bolton, 2020). 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). However, as noted above, later evidence showed that the de-novo approach resulted in better housing quality and price premia (Michaels et al., 2021), and in recent years, de-novo planning has been implemented by several African governments (Centre for Affordable Housing Finance, 2024).

### 2.3 The "20,000 Plots" Project

Our study focuses on the "20,000 Plots" project, which the Tanzanian government initiated in the late 1990s in response to perceived unmet demand for formal de-novo plots (Tiba et al., 2005). Such plots secure owners' property rights against outright or partial expropriation and against nearby changes to the 20k layout that could reduce plot value (e.g., plot splitting, as discussed in Section 6.3 or blocking road access). Formal property rights are secured by making the cadaster rigid, so it is difficult and costly to change to ex-post, especially when state capacity is limited, as in many lower income countries.[2] We note that in the last few years Tanzania has worked to create formal plots in about 40 additional cities.[3]

The 20k project, which was implemented around 2000-2005, delivered over 36,000 residential plots in twelve project areas (neighborhoods) spanning a total area of 75 square km (sqkm). The residential plots, which take up roughly half the area (~38 sqkm), were formally surveyed and titled. Approximately 1,500 additional plots (~12 sqkm in total) were designated for non-residential public and commercial uses. The remaining area (~25 sqkm) was taken up by roads and shoulders (almost all unpaved) and hazardous land (e.g., near streams or water bodies) which was left empty.[4]

Figure 1 shows that the 20k project locations were mostly near the fringes of Dar es Salaam. Like the Sites and Services projects in Tanzania ~50 years ago, the expectation is that as cities grow, these locations will no longer be on the fringe. The maps also show the preexisting main paved roads and the boundary of the Dar es Salaam metropolitan area. The government charged a fixed price per square meter within each project area to cover its costs; variation in prices between project areas likely resulted from higher initial land costs in areas with better access to the city center and the coast (Mwiga, 2011).[5] The maps also show that compared to the government sale prices, plot prices appreciated rapidly in all project areas, though not uniformly.

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 was ~ 33 million USD 2021 (~ $1 per sqm of residential plot).[6] The initial phase was financed by an internal 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 project cost was recouped (Tiba et al., 2005).

However, the second limitation of Sites and Services, that they did not cater to the poor, was not addressed in the 20k project. Tanzania's income has risen in recent decades, and the mean

[2] Most of Dar es Salaam is made up of informal areas where it is possible to make changes such as plot splitting. These changes are monitored by local leaders but not recorded in a cadaster.

[3] Part of our fieldwork involved meeting with officials from these localities to share insights on the 20k project.

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

[5] We include in our analysis one area, Malindi, which was developed from 1998 and later integrated into the 20k project, but we do not have the initial government-set price for this area. The 20k project also provided a few thousand plots in other cities in Tanzania, but we do not study those since we have no data on their precise locations.

[6] Details on the project cost figures are in Appendix .

price per square meter of the 20k residential plots sold by the government was much lower than in the earlier Sites and Services project (Michaels et al., 2021). Nevertheless, the poor were largely excluded from 20k, likely due to a combination of three reasons. First, the plots were sold quickly to repay the loan mentioned above, which was an obstacle to many of the poor.[7] Second, the resale of plots at a premium over the initial government price likely meant that many plots eventually ended up with richer buyers. Finally, plots' large size and resulting higher cost might have crowded out poorer potential buyers. As discussed above, large minimum plot sizes were common in British colonies, and in Tanzania they were retained after independence (Kironde, 2006). We note that when the 20k project was implemented, formal plot sizes in Tanzania ranged from 400-4,000 square meters, although the minimum has since been reduced to 300 sqm (MLHHSD, 2018).

These considerations – especially the first – resulted in the initial allocation of plots to lucky or connected owners, including (anecdotally) many government officials. Those who bought benefited from rapid price appreciation and often resold plots at market prices to those who actually wanted to build in 20k areas. Our analysis focuses on understanding who these eventual buyers were.[8]

# 3 Data

## 3.1 Data sources

This paper uses many 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 A. We have project plot and neighborhood mappings for all 20k areas, which we use to measure planning treatments. We also obtained satellite images with a resolution of ~ 0.5 meters, which cover the project areas and a buffer of 500 meters around them. Ramani Geosystems in Nairobi digitized information from these images, including the footprints of buildings in the end period (typically 2019-2021) and earlier periods (see Appendix A.2). We measure underlying locational fundamentals, such as elevation and ruggedness, using a digital elevation model (United States Geological Survey, 2000). Open Street Map (OpenStreetMap contributors, 2017) is used to determine the locations of rivers or streams and water or wetland (see Appendix section A.3).[9]

Additional data gathered include the following primary sources (see Appendix section A.4).

[7] 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 given to applicants 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 paying within 14 days. Failure to complete the payment and finalize the transaction within 60 days resulted in the reallocation of the plot to another potential buyer.

[8] In the questionnaire we administered, the majority of 20k owners reported some additional land holdings - not necessarily in planned areas. However, a large majority (89 percent) said that their households do not own additional 20k plots in their local area ("mtaa" for singular and "mitaa" for plural). This motivates our focus below on owners of single 20k plots for their own use.

[9] While these mostly reflect "first-nature" differences across locations, project development may have altered them slightly (e.g., if some land was leveled).

First, we interviewed (i) local experts and (ii) leaders of 34 local administrative areas ('mitaa'), whose jurisdictions span almost all 20k plot areas and adjacent non-20k areas. Second, we administered questionnaires to (i) local real estate agents ('madalali'), who provided sales dates and prices for individual plots in 20k areas and nearby non-20k areas that were sold in market transactions 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.

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

Using the project maps, 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 (the formal maximum size at the time of the 20k project). We classify the remaining plots as non-residential; these serve both private and commercial uses, as described in Appendix A. Figure 2 offers a concrete example of an area on 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 unpaved road, as planned.[10] The plan also shows non-residential insulae with a variety of intended uses. Finally, the figure gives an example of a super-insula, a contiguous set of insulae with similar size plots, as defined later to be either small, medium, or large plots. Panel B includes an image of the same area, showing that the buildings mostly conform to the planned plot outlines, although some residential plots and many non-residential ones are unbuilt.

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 roughly 20 years later, in 2021. Within the planned area, buildings are regularly spaced, with roads between the insulae, in accordance with the plan. In contrast, in the informal (unplanned) area, building sizes are less uniform and typically smaller, especially away from roads; some are bunched together irregularly, and many seem inaccessible via roads. This example highlights some of the benefits of de-novo planning.

### 3.3 Dataset construction

In our main dataset, the units of analysis are small square land parcels ("gridcells"). These gridcells may be "treated" by the planners, who may assign them to residential plots of different sizes or vary their proximity to planned amenities, and our empirical methodology in Section 5 disentangles

[10]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.

the relative effects of these treatments. Each gridcell is a 20 x 20-meter square, corresponding to the minimum formal plot size. 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 ~ $95,000$ gridcells whose centroids fall within residential plots. The project defined official minimum thresholds for small (400 sqm), medium (800 sqm), and large (1600 sqm) plots, with 15043 small, 16853 medium, and 4319 large plots in the 20k areas.

The outcomes we study in Section 5 are mostly related to the model outlined in Section 4 and estimated in Section 6. First, we measure the real market price of plots that were unbuilt ("bare land") when they were sold privately; these prices are available for 1,446 residential plots (1,122 from the real estate agents and the rest from the resident questionnaires).[11] Second, we measure housing investments using the satellite imagery, which cover all the gridcells: (i) a measure of construction intensity - the share of each gridcell that is built and (ii) a measure of the rate of uptake, or extensive margin of housing construction, whether a plot is built upon by 2020 - 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.

Our regressors of interest are the logarithm of plot size and measures of amenities - preexisting, planned, and implemented - at the gridcell level. The foremost of these, the general "attractiveness" of each 20k area such as its access to the Dar es Salaam city center are swept up in area fixed effects. Other controls include, within 20k areas, the distance to the nearest preexisting main paved road and a variety of geographic characteristics listed later, indicators for the gridcell being within 100m of different planned amenities listed below, and a Z-index of three insula characteristics: rectangularity, alignment, and homogeneity defined later. In price regressions, we control for the time period of sale interacted with source (real estate agents or residents). The main variables are also described in Table A.1 and their summary statistics are reported in Table A.2.

## 4 Modeling the development of 20k plot areas

Planners of de-novo neighborhoods choose large tracts of greenfield land on city outskirts, which they partition into residential plots of different sizes and other designated uses. Planners' objectives may include increasing social welfare; raising land values (e.g., Turner et al., 2014); ensuring that formal plots are built, an issue raised by the Tanzanian Minister for Lands, Housing, and Human Settlements Development (Jamal, 2018); increasing intensive-margin development and distributional considerations, such as widening access to formal ownership. The framework outlined

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

here considers outcomes relevant for all of these objectives.

In this section, we discuss the consumers' optimization problem for those who ultimately build in 20k areas, analyzing their choices so as to help interpret our reduced-form analysis. We leave the discussion of the estimation of the full model including who builds in 20k areas and the model counterfactuals to Section 6.

## 4.1 Modeling assumptions and the consumers' problem

We focus on buyers who at time 0 (the year 2005) live in the city (Dar es Salaam) but ultimately move to 20k areas.[12] We assume that these people live infinitely, have a time discount rate $\rho$ and face an interest rate $\delta$, and derive instantaneous utility $\varphi lnh + \beta lnz_1 + Ae^{-\theta t}$ when in the city and $\varphi ln(l^{\alpha}k^{1-\alpha}) + \beta lnz_2 + B$ once they move to 20k, where and $z_1$ and $z_2$ denote non-housing consumption in both locations. To rationalize the movement of people from the city to 20k areas, we assume that the city's amenity, $A$, which we normalize to 1, deteriorates at a rate $\theta$ relative to the fixed amenity of the 20k areas, $B$ (which can vary across plots). We think of the formulation of the city amenity as representing relative deterioration of traffic conditions, air quality, safety and the like over time. We also assume that in the mostly informal markets in the city, each person chooses their optimal housing $h$ for a unit price, while in 20k areas plot supplies are limited. [13] We assume that a person who wants to move to 20k buys a plot of size $l$ at time 0 and moves to it at a time of their choosing, $\tau$, when they irreversibly invest $k$ in housing capital on their plot. Buyers' period incomes, $w$, are distributed between $[w_1, \bar{w}]$, as discussed later.

At time 0, these future 20k residents buy from "initial owners", who had previously purchased the underpriced plots from the government. If some initial owners do not sell, we assume that they are willing to pay the market price since this is the opportunity cost of keeping the plot. The 20k plots are scarce and in equilibrium every plot is purchased at time 0 for a price $R_{B,l}$, equating demand and supply for those who choose to move to 20k.

We assume that 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 assume that would-be owners are not constrained by capital market imperfections, noting that we mostly focus on owners who are relatively rich. What we call a period income ($w$) is a measure of permanent income, since with a perfect capital market, all that matters is $W$, the present discounted value of lifetime earnings (including any endowment). For simplicity, we equate $\rho$ and $\delta$, so optimized non-housing consumption ($z_1 = z_2 = z$) is constant over the lifetime. The perfect capital market assumption and the equating of $\rho$ and $\delta$ are simplifications that do not affect the generality of the principles we

[12] We focus on higher-income owners, who likely lived in the city before moving to 20k, rather than poorer immigrants from rural areas.

[13] Recent estimates suggest that around 80% of the buildings in the city, on which more than 80% of the residents live, are on unplanned and unsurveyed land, and so lack formal plot boundaries (Manara and Pani, 2023).

develop. Finally, we assume that owners cannot rent out their 20k plots, which is consistent with the small share of renters that we observe, perhaps because it is difficult to evict renters even when plots' ownership is undisputed.[14]

Conditional on choosing to build in 20k areas, each plot buyer decides on the time $\tau$ to move from the city and sink $k$ by solving the optimization problem:

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

## 4.2 Interpreting reduced-form empirical results

Here we focus on the effects of the key planning variable, plot size ($l$) and then for general interpretation, amenities ($B$). We obtain expressions from the first-order conditions for Eq.1 (see Appendix B.1), for the three outcomes of interest with and without sorting: land prices ($R_{l,B}$); the probability that a plot is developed 15 years after the project began, $Prob(\tau < 15)$, which is inversely related to $\tau$; and the housing capital investment ($k$).

**Plot prices.** Higher prices must be paid for higher $l$ and $B$ in any Nash equilibrium. For example, two plots with the same amenities but of different sizes cannot have the same price: a consumer on the smaller plot would be willing to pay more for the bigger one. Thus, prices rise with $B$ or $l$ according to the equilibrium elasticities:

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

**Date of development.** In Appendix B.1, Eq. 6 gives an expression for $\tau$. Differentiating that equation we get:

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

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 Eq. 2, holding owner income constant, bigger plots as well as higher- amenity ones are developed sooner ($\tau$ is lower). The direct effect of a bigger plot, $l$, makes the owner move sooner from the city; this is reinforced by the indirect price effect, $\frac{\delta R}{w-\delta R}\eta_{R,l}$, which makes a bigger plot more expensive, shifting owner expenditure from the city to 20k, and inducing faster development. But in equilibrium with sorting as emphasized in Section 6,

[14] For tractability, we do not model commuting to work from 20k to the city. But in Section 6 we explore the role of access from 20k to the city as an important amenity.

otherwise identical bigger plots are purchased by higher-income people, and the impact of higher income in Eq. 2 is to raise $\tau$ and delay development. Thus, without a control for owners' different lifetime incomes, the total effect of plot size on $\tau$ becomes ambiguous. A similar argument applies to the effect of amenities, $B$.

**Investment.** The equation for $k$ is Eq. 7 in Appendix B.1. Differentiating and substituting in for $d\tau$ from Eq. 2, we get:

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

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 Eq. 3, holding income constant, the direct, complementary effect of an increase in $l$ or $B$ is to increase $k$. However, countervailing this is the indirect effect of the increase in price, $R$, which squeezes the budget and reduces $k$. Thus, even holding income constant, the effect of an increase in $l$ or $B$ on $k$ is ambiguous. Given that, with positive sorting effects, in the reduced form empirics and in the structural estimation in Section 6.2, we will generally see that a higher $B$ or $l$ is associated with higher investment.

**Summary for reduced-form empirics.** The variables that we construct and their relationship to the model are noted in Section 3.3. In the equations we estimate below, we expect the following relationships. First, we expect richer owners to sort into larger and higher-amenity plots. Second, plot price (though not necessarily the price per sqm) should rise with plot size. Third, plot price and the price per sqm should increase in amenities. Fourth, our measure of plot development is an indicator of whether each plot is built in 2020, which is an indicator for $\tau < 15$. The expression for this term is derived in Appendix Eqs. 6 for $\tau$ and 7 for $k$, with impacts defined by Eqs. 2 and 3 respectively. Conditional on amenities and owner income, larger plots are more likely to be developed, but in our reduced-form regressions where we cannot control for owners' lifetime income, we may have an ambiguous effect on the rate of development. Fifth, controlling for plot size and owner income, higher-amenity plots are more likely to be developed, but again owner income may matter. Finally, the effects of plot size and amenity on investment levels ($k$) are theoretically ambiguous when controlling for income, and without a control for income, the sorting effects generally dominate so that investments increase with plot size and amenities.

# 5 Reduced-form effects of planning choices

## 5.1 Methodology

Our reduced-form analysis takes as given the 20k plan and the resulting sorting of owners across plots, as discussed in Section 4. From this starting point, this section considers the effects of marginal (hypothetical) planning changes. Quantitative analysis of the effects of larger changes in the planned layout on outcomes and ownership, including general equilibrium (GE) effects, requires stronger assumptions, and is deferred to Section 6.

As an example, consider a marginal planning decision to treat a tiny land parcel differently by reassigning it between two geographically proximate plots of different sizes, without affecting its amenity. This tiny parcel itself is affected due to a "total" combination of a plot size effect and a sorting effect, as discussed in Section 4. Such total effects are relevant for assessing the impacts of incremental changes that planners may have considered. They are also informative for considering small changes that individuals may wish to undertake, such as splitting a single plot and reselling it, although in our setting such splitting is prohibitively expensive, as discussed in Section 2.3. Finally, they inform us about which planning decisions are important and shed light on whether the planned layout is efficient.

In practice, we consider effects on small gridcells, as discussed in Section 3. Our analysis is aided by the fact that the project areas were largely agricultural (greenfields) circa 2000, which limits preexisting differences in amenities or sorting. We begin with OLS regressions, which control for area fixed effects and observable physical characteristics, to mitigate the potential for confounding factors within project areas. When studying the consequences of plot size, we also use spatial regression discontinuity (RD) designs, which compare gridcells that are very close to each other (e.g., they on either side of an unpaved road) but differ in whether they are part of a smaller or a larger plot.

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

$$y_i = \beta_1 \text{Plot\_size}_i + \textbf{Program\_area}_i' \gamma_1 + \textbf{Controls}_i' \lambda_1 + \epsilon_{1i}, \tag{4}$$

where $y_i$ is an outcome in the gridcell or its plot. $\text{Plot\_size}_i$ is a measure of the size of the plot in which gridcell $i$'s centroid falls. $\textbf{Program\_area}_i'$ is a vector of twelve 20k project area fixed effects [FE] (within which the initial government-set price per sqm was identical); usually we further interact these with FEs for the 34 mitaa (small administrative units), which focuses the analysis within small areas, further reducing the potential role of any unobserved amenities within the greenfield areas. Our standard controls, $\textbf{Controls}_i$, which we refer to as "amenities", include predetermined and planned features. Predetermined features are: distance in km to the nearest major paved road; elevation; ruggedness; and indicators for being within 100m of a river or a stream and of water or wetland. Planned features, on which more details are provided below, are a Z-index of three insula characteristics (rectangularity, alignment with neighboring insula, and homogeneity of plot sizes within the insula) and indicators for being within 100m of an edge of a 20k area and each of nine planned non-residential land uses (such as schools or religious sites). In price regressions, time period interactions by source of data (real estate agents or residents) are also included. $\epsilon_{1i}$ is an error term, and 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.[15]

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

$$y_i = \beta_2 \text{Own\_larger}_i + \mathbf{Program\_area}_i' \gamma_2 + \mathbf{Dist}_i' \lambda_2 + \mathbf{Boundary}_i' \rho_2 + \mathbf{Controls}_i' \kappa_2 + \epsilon_{2i}, \quad (5)$$

where $\text{Own\_larger}_i$ is indicator for gridcell $i$'s insula having a larger mean plot size than the insula with which it shares a boundary segment; $\mathbf{Dist}_i$ is the distance in meters to the boundary segment and its interaction with $\text{Own\_larger}_i$ indicator; $\mathbf{Boundary}_i$ is vector of boundary segment FEs; and $\epsilon_{2i}$ is an error term, again clustered by insula. The RD regressions we estimate are typically semiparametric, where we restrict the analysis to gridcells within 100m from their insulae boundaries, which includes most gridcells, but we have experimented with other bandwidths and found few differences. Our identification strategy assumes that amenities vary smoothly at the boundary and the specification allows us to control for them.

To compare the RD and OLS estimates of plot size effects, we follow Calonico et al. (2025) and add terms to Eq. 5 for the absolute difference in log mean plot size across the insula-pair boundary and its interaction with $\text{Own\_larger}_i$ and all other independent variables in the specification. We explore this further by estimating treatment effect heterogeneity in our RD model, using the semi-parametric smooth varying-coefficient model (Rios-Avila, 2020). The spatial RD strategies we use relate to the literature (e.g., Dell, 2010; Turner et al., 2014; Michaels et al., 2021), but we differ from most existing studies in our analysis of a greenfield setting, in examining nonlinearities, and in focusing on spatial discontinuities within smaller administrative units, which strengthen identification.

## 5.2 Main empirical findings

### 5.2.1 Aggregate land value gains from planning

We begin with evidence about the appreciation of land values (prices) in the 20k project. Table A.3 shows that compared to the government-set prices, which reflect the cost of delivering the plots in

[15]To justify insulae clustering, we note that insulae fixed effects have high R-squared - typically around 0.8 - in explaining plot size variation within project areas. Our s.e. estimates are, however, broadly similar when we cluster on smaller plot identifiers (of which there are roughly 36,000) or larger units, such as 189 interactions of program areas with enumeration areas in the 2012 census, 34 mitaa, or even the 12 project areas.

each area, the logarithm of real land prices increased in all project areas, with a mean increase of about two, corresponding to a rise of more than 600%.[16] Areas that were initially more expensive did not see differential price appreciation, but price appreciation was slowest in the remote areas of Mwongozo and Buyuni (see Figure 1). In Section 5.2.4 we argue that poor access in these two areas is a key disamenity.

To see how planning is valued, Table 1 compares bare land values in 20k to those in nearby non-20k areas, which were sold by the same set of real estate agents. Columns 1 and 2 show that *conditional* on plot size, 20k land prices are about twice those of non-20k informal plots (since $e^{-0.7} \approx 0.5$), reflecting the benefits of both formalization and planning. Priced between the two is the smaller set of non-20k plots that were formalized, reflecting more secure property rights. Formal plots that are in 20k areas sell for ~30% more, reflecting the planning benefits of 20k. Column 3 of Table 1, however, shows that if we do not control for plot size, the *unconditional* price premium of 20k areas over informal non-20k areas is 50% rather than 100%; this reflects the larger mean size of 20k plots and the lower price of large plots, which we discuss below.

We compare the conditional price estimates with separate estimates (not reported here) based on interviews of 34 mitaa leaders, each of whom estimated the price of bare-land plots of different sizes – both in the 20k areas and in informal non-20k areas – within their mtaa. Those estimates similarly suggest that the 20k premium, conditional on plot size, is roughly twofold. We also asked the 34 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 leaders who answered this question emphasized two factors. First, 24 leaders mentioned property rights (21 of whom explicitly mentioned land titles), saying that they reduced boundary conflicts, increased tenure security, and improved access to financial credit. Second, 23 leaders mentioned better access under planning in 20k areas (20 of whom specifically mentioned roads). The leaders noted that non-20k areas tended to become crowded over time and inadequate access made local service provision harder.[17]

#### 5.2.2 Sorting

The model predicts that high-income owners sort into 20k areas, and the data confirm this. Our survey shows that the mean years of schooling among owners is 13.8, compared to 8.7 years for heads of households in Dar es Salaam as a whole (National Bureau of Statistics, 2019). As expected, our survey data also show a strong positive correlation between years of schooling and income.

[16] Coefficients on time period effects (not shown) indicate that the market quickly anticipated the value of 20k, since the roughly sixfold appreciation relative to the government prices happened as early as 2000-2010, before declining by about 20-30% and stabilizing at the high levels reported in Table A.3.

[17] Michaels et al. (2021) discuss the possibility that water mains were also an important feature of de-novo, but piped water was only provided to part of one 20k project area; most 20k residents obtain water in other ways.

The model also predicts sorting within 20k areas. To study this, Table A.4 uses survey data for "Near" areas, as discussed in Section 5.2.4 below.[18] The table shows that mean years of schooling (a strong predictor of income) of owners of built plots. Schooling increased over time within plot size group and across plot size groups in the same time period. In Section 6 we show that these patterns are consistent with the calibrated model.

#### 5.2.3 Plot size effects on land price, plot development, and housing investment

**OLS estimates.** We begin our investigation of plot size effects by using OLS to estimate the elasticity of the price of bare land per square meter with respect to plot size, using specification (4). In columns 1-3 of Table 2, we show that this elasticity is ~ $-0.5$ with different sets of controls. In column 4, 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), and the elasticity of ~ 0.5 is consistent with the estimates in the other columns. Panel A of Figure A.3 shows non-parametrically that the negative relationship holds throughout the plot size range, using 100-sqm bins as regressors. Table A.5 shows that outside 20k areas, where plot sizes are not fixed by planning, the elasticity of price per sqm with respect to plot size is much lower (about 0), at least for surveyed plots, suggesting arbitrage where subdivision is less costly.[19]

Next, we examine the implications of plot size for housing outcomes. Table 3, which reports estimates from Eq. 4, suggests that the share of the gridcell that is built declines in plot size, so large plots have a larger share of open space. The probability that a plot is built is unrelated to plot size, consistent with the ambiguity in the comparative statics when sorting is accounted for in Eq. 2. For built plots, the elasticity of built area with respect to plot size is around 0.11-0.14, so that with sorting bigger plots have more investment, but with a relatively low size elasticity. The final column shows that large plots also have a higher likelihood of housing multiple buildings.[20] For both measures of investment, as expected from Eq. 3, sorting effects dominate any direct effects on investment of changes in plot size. Panels B-E of Figure A.3 show the same relationships non-parametrically, and at least for well-populated bins the relationships are mostly monotonic.

**RD estimates.** The estimates reported above control for small area fixed effects and many amenities, but a residual concern is that plot sizes may be correlated with unobserved amenities. To address this, Table 4 reports RD estimates using Eq. 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

[18] The sample for far areas is too small since we only surveyed one of the far areas.

[19] Panel A of 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.

[20] Partly, this may reflect backyarding on larger plots due to the high costs of subdividing them, as discussed in the context of South Africa by Brueckner et al. (2019).

buildings; the only qualitative difference from OLS is that here there is no significant effect on log building size. In panel B, we restrict the sample to gridcells where the mean plot size gap across insulae is large (> 400 sqm), and the statistically significant effects are roughly two-three times larger. Panel C focuses on discontinuities with small size gaps (< 100 sqm); all estimates are small and statistically insignificant. We conducted a number of robustness tests.[21]

The OLS and RD estimates are qualitatively similar, but we also want to compare their magnitudes. In panel A of Table 5, we report estimates of an RD specification where the gap in log mean plot size across insulae within each segment is interacted with $\text{Own\_Larger}_i$ and all control variables (Calonico et al., 2025), and in panel B we report OLS results for the same sample as the RD. The estimates show that, where the OLS and the corresponding interacted RD estimates are both significant (columns 1, 2, and 5), they are quantitatively similar. In Figure A.6, we use a semi-parametric approach to estimate treatment effect heterogeneity of the plot-size effect across insulae pairs with varying gaps in mean plot size (Rios-Avila, 2020). For price, share built, and multiple buildings, we find effects that are approximately linear in log gap size, consistent with the results in columns 1, 2, and 5 of Table 5. For plot built there is heterogeneity; the effect is not statistically different from zero for a gap size near zero, then becomes slightly positive for moderate changes in gap size, and finally becomes negative for more substantial changes in plot size. This is consistent with column 3 of Table 5 that estimates a positive intercept and negative slope when constrained to be linear. In short, Figure A.6 suggest that the effect on plot built is somewhat non-linear. Through the lens of the model, this can be interpreted as the plot size effect in Eq. 2 dominating for small plot size gaps (lowering $\tau$, or raising likelihood of plot built), but as the gap starts to grow the sorting effect offsets and then dominates (raising $\tau$, or lowering likelihood of plot built).

A key result in this section is that the estimated elasticity of land value with respect to plot size, of around -0.5, suggests a potential oversupply of large plots because if a large plot were subdivided its total value would increase. However, to determine whether splitting plots indeed represents misallocation from a social planner's perspective, we need to consider the equilibrium effects of changing supply, which we turn to in Section 6.

#### 5.2.4 Access as a key amenity

Our findings show that even though all 20k plots benefit from local road access, connectivity is an important amenity, and most likely the most important one in our setting. As we discuss here, this is evidenced by across areas and within them. To highlight the cross-area comparison, we

[21] First, we verified that they are balanced on first-nature fundamentals in Table A.6. Second, we select optimal bandwidths (Calonico et al., 2014) for each outcome, in Table A.7 all are more narrow than our preferred 100m bandwidth and give very similar and even more precise results. In addition (not reported), both RD and OLS are robust to dropping the few gridcells whose plots contained buildings before 2005 (e.g. farm buildings).

partition the 20k project areas into two groups: "Near" and "Far", which reflect proximity to the Central Business District (CBD) of Dar es Salaam. We classify as "Far" two project areas, Buyuni and Mwongozo, which stand out for: (i) their distance from the CBD (Figure 1); (ii) the lowest land price appreciation relative to the government-set prices; (iii) the lowest current land price (Table A.3); and (iv) being the only project areas where in 2010 fewer than 10% of the plots were built, when the mean across the other areas was ~25%. For within-area comparisons, we focus on proximity to preexisting main paved roads. Table 6, which reports estimates of specification (4) using all baseline controls, shows that the most prized amenity is proximity to preexisting paved main roads. Increasing the distance to such roads by 1km reduces land value by almost 15%. The likelihood that a plot is built and the intensive-margin investment levels also decline in distance to main paved roads.

#### 5.2.5 Other Amenities

Having discussed access, the amenities we consider in this section are mostly other predetermined and planned ones. Recall from the model, holding plot size fixed there should be unambiguous positive effects of better amenities on prices, but our sample for prices is comparatively small. For the likelihood of being built upon by 2020, controlling for plot size, we hypothesized potentially ambiguous timing effects from Eq. 2, although we will see that, for amenities, the plot size effect appears to dominate: higher-amenity plots are developed sooner. For levels of investment, we hypothesized strong sorting effects such that better amenities would be associated with more investment in housing.

Table 6 considers as amenities the natural ("first nature") features. We cannot detect significant effects of these on prices given the limited price sample and variation within that sample, noting that, overall, tiny fractions of plots are near a river or wetland. But for housing outcomes, where we have more statistical power, we have results that follow the pattern for distance to main paved roads for probability of being built and investment levels. Elevation seems beneficial with three positive and significant coefficients, while ruggedness looks like a disamenity, significantly reducing the share built and likelihood of being built upon. Proximity to rivers (or streams) also significantly reduces the share built and the likelihood of being built upon, and being near wetlands significantly reduces the likelihood of being built upon. 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.

In the second part of Table 6, we consider two aspects inherent in planned communities. First concerns insula features and second what happens as one approaches the border of 20k areas. For features, we have a Z-index of three insula characteristics: rectangularity, alignment with neighboring insula, and homogeneity of plot sizes within the insula.[22] We find that a more "regular" layout increases the share built and the probability that a plot is built. In Table A.9 we unpack the estimated effects of the Z-index into its three components, and find that insula rectangularity and alignment appear to be valued, but homogeneity is not. For the second variable in the middle panel, we find that proximity to the edges of 20k areas, where there are more informal settlements, may marginally and modestly detract from the share built and the built area.

In the bottom part of Table 6, we estimate effects of proximity to the nine types of non-residential plots. We have indicators for being within 100m of 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. Almost all coefficients on these planned non-residential uses are small and statistically insignificant. However, this does not mean that owners do not value these amenities; instead, the problem is that implementation rates are low. In Table A.11 we see that even *implemented* amenities have a limited association with housing outcomes. Only service trade and housing estates (which take up a tiny fraction of observations) have consistently positive coefficients, and there is some suggestive evidence that religious sites matter. But even those positive correlations could in part reflect implementation following housing construction. Moreover, due to non-implementation, most planned non-residential plots are vacant. These vacant areas may be "maintained" by the local population, but many are unkept (have wild growth, garbage, etc.). Table A.11 suggests that being next to an unkept non-residential use is associated with significantly lower housing outcomes, though again plot maintenance (like implementation) is endogenous.

Figure 4 shows that the implementation of all the planned non-residential significantly lags the plans. The lag is smaller for three categories (cemeteries and religious and educational uses), and larger for the remaining five (recreation, public buildings, nursery, service trade, and housing estate). At the same time, most plots designated as non-residential are either misused (~9% are residential and ~25% are farmed) or unused (~40% are vacant, split between kept and unkept). Related, Figure A.5 focuses on the non-residential plots intended for eight main planning categories and asks how each is used. Approximately half of the plots intended for cemeteries and ~ 40% of those intended for educational or religious uses are implemented as planned, but implementation rates for the other five categories are much lower. All this suggests that planners were overoptimistic when prescribing non-residential uses, which have yet to materialize about 15 years after the onset of the project.[23]

[22]Rectangularity is the size of the insula divided by the size of the minimum bounding rectangle. Alignment of the nearest bordering insula to the own is 1-tan(angle between the two sides), where tan(0)=0 and the maximum angle is 45 (for tan(45)=1). Homogeneity is 1 - the coefficient of variation of plots sizes within the insula. Thus, the best values for the raw measures are 1 for perfect rectangularity, prefect alignment, and no variation in plot size. Each of these measures is standardized to a Z-score and the three Z-scores are averaged to get the Z-index.

[23]Even though implementation rates for non-residential amenities are low, we find that conditional on implementation, the planners' guidelines on landuse locations were followed, compared to a benchmark where implemented use locations were randomly selected among the non-residential plots. To show this, Table A.10 reports the 'compliance

We next investigate local neighbourhood composition effects. In particular, we examine whether people value having neighboring plots that are bigger vs. smaller. Hypothetically, this valuation may depend on expectations of sorting by socio-economic status, differential timing of local development, or amenity effects of plot size per se. While the plot size composition approach allows for a clean identification strategy, we are unable to disentangle these various mechanisms with the data we have. We repeat the specification in Eq. 5, but this time for boundaries between super-insulae rather than insulae, controlling for the log size of the own plot.[24] At a border between super-insulae of different types, residents experience a mix of neighboring plots of different sizes. As we move into the super-insula interior, residents are increasingly exposed to neighboring plots with a similar size to theirs. Since there are far fewer super-insulae boundaries than insulae boundaries, we focus here on the housing outcomes that are available for the entire sample, and not on prices, which are available for only a small subsample. Table 7 shows that as we move into super-insulae with smaller plots, the likelihood that a plot is built and the share built both increase modestly: moving 100m deeper into a super-insula with smaller plots than its neighbor raises the mean share built by 0.56 percentage points and the likelihood that the plot is built by 2.3 percentage points. As we move away from the boundary on the larger plot side, however, the estimated coefficients are all insignificant.[25] Although the effects are small, they suggest that owners of small – rather than large – plot owners may value neighborhood uniformity. However, the small magnitude of effects motivates our model abstraction from externalities.

# 6 Model estimation and counterfactual results

## 6.1 Estimation of the model

We estimate the parameters of a simplified model based on the equilibrium observed in the data, focusing on our findings that land prices per square meter decline sharply with plot size despite the sorting of highly educated households. The model suggests that the project's residential plot sizes were too large, which excluded poorer potential owners and thus did not maximize welfare. We discuss the estimated equilibrium in Section 6.2 and compare it with two alternative counterfactual plot size allocations in Section 6.3.

In our estimation, we make three simplifications noted above. First, we focus only on the reallocation of the current fixed amount of residential land across plot sizes. In Section 5.2.5, as

ratio' for each use $j$, which we define as: $P(\text{implemented as } j \mid \text{planned as } j)/P(\text{implemented as } j)$. The observed ratios in the first column are fairly similar to those that would have resulted from perfect implementation (second column) and much higher than those that would have resulted from random choice of non-residential plots. In other words, conditional on implementing non-residential uses, the planners' intent mattered.

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

[25] We found similar results using specifications that pair types of bordering neighborhoods (e.g., small vs. medium or small vs. large).

just discussed, because of low implementation rates, we cannot assess how consumers value or sort around planned non-residential uses, so we leave that allocation fixed. And since all insulae front on roads, so we leave road allocations unchanged, noting that plot size allocations effectively carve up existing insulae into fewer or more plots. Second, as discussed in Section 5.2.4, we partition the 20k project areas into two groups as noted earlier: "Near" with amenity $B_F$ and "Far" with amenity $B_N$. Attempts to estimate the model with further differentiation of $B$'s yielded very similar $B$'s within the near group. So we end up with three locations: Near, Far, and the city. For amenities connected with configurations of plots sizes in neighboring super-insulae, if having similar plot neighbors matters, one can think of the planner configuring our model allocations of plots so as to group similar types together. Finally, and consistent with the planners' characterization of 20k, we discretize plot sizes, allowing for small (600sqm) $l_S$, medium (1200sqm) $l_M$, and large (2000 sqm) $l_L$ sized plots within each of the two project area groups. In sum, there are six different types of plots (two types of amenities by three sizes) which we index by $m$. Details for solving and estimating the model are in Appendix B. Here we give an overview.

**Plot owner optimization problem.** The plot owner residents face a nested choice problem where they first choose to live in the city permanently vs. move to 20k, then in the second branch, choose which plot type to buy and finally how much to invest in their plot and when to move (i.e., the solution to Eq. 1). This problem can be solved by backward induction as follows. In the final stage, conditional on choosing 20k and a particular plot type $m$, residents receive indirect utility from optimization in Eq. 1 defined as $U_m(w, R_m; \Theta)$ where individuals take income $w$ and the land price $R_m$ as given, and $\Theta$ is a collection of model parameters.[26] In the penultimate stage, residents choose type of plot that gives them the highest indirect utility; $m^*$ where $U_{m^*}(w, R_{m^*}; \Theta) \geq U_m(w, R_m; \Theta) \ \forall \ m \neq m^*$.

In the first stage, residents choose whether to ever move to 20k areas or stay in the city permanently. Before choosing, each draws a preference shock from a gamma distribution $\mu \sim \Gamma(\mu^{shape}, \mu^{scale})$ which additively enters their utility as $ln(\mu) \equiv \tilde{\mu}$, which represents their idiosyncratic preference for moving to 20k areas, as opposed to staying in the city. The indirect utility of staying permanently in the city ($\tau = \infty$) has a closed form solution $U_0(w; \Theta) = (\varphi ln \frac{\varphi w}{p(\beta+\varphi)} + \beta ln \frac{\beta w}{\beta+\varphi})/\rho + \frac{A}{\rho+\theta}$ from Appendix B.1. Therefore, residents will choose to move to 20k areas on their preferred plot type rather than staying in the city forever if $U_{m^*}(w, R_{m^*}; \Theta) + \tilde{\mu} \geq U_0(w; \Theta)$.

**Solving for the equilibrium.** The equilibrium is defined by a set of allocations that equate demand and supply of 20k plots, and a set of prices that ensure no resident could make themselves better off. Details on solving the equilibrium are provided in Appendix B.2.

The supply of plots is the observed number of plots of each type $N_m$, with $N = 36,215$ in total. We assume that 800,000 families in Dar es Salaam in 2005 each potentially demand one

[26] As explained in Appendix B.1, $U_m$ does not have a closed form solution in general, so we solve it computationally.

20k plot. Their annual incomes are distributed between $[w_1, \bar{w}]$. As detailed in the Appendix, we set $\bar{w}$ to \$17,291, the 99.9th percentile of predicted household incomes from a gamma distribution, $w \sim \Gamma(w^{shape}, w^{scale})$, fitted to the observed household incomes in Dar es Salaam in a 2015 survey.[27] The empirical estimates of $w^{shape}$ and $w^{scale}$ are discussed below. The lowest income of those who can afford to move to 20k areas, $w_1$, is estimated in the moment-matching exercise described below.

In equilibrium, the owners sort between plots of different types. The smallest low-amenity plots are purchased by those with the lowest incomes, equating supply and demand. As we move up the income scale, following the plot-level index $\tilde{m} \equiv \phi\alpha ln(l) + B$ from the preference function, the purchased plots are larger, or have better amenities, or both. We use the index $m \in \{1, 6\}$, where $m = 1$ denotes the plot type with the lowest $\tilde{m}$, $m = 2$ the second lowest, etc.

The equilibrium prices equate utility for marginal "entrants" to each plot type, but residents are differentiated by both income and their idiosyncratic $\mu$'s. So, there is a continuum of incomes for which people are indifferent between 20k and the city, and we define these marginal entrants by the locus $\tilde{\mu}(w)$, which is the union of the loci for each segment, $\tilde{\mu}_m(w)$. The price of land in a plot of type $m = 1$ is determined by the poorest individual on the locus (i.e. with scalar $\tilde{\mu}_1(w_1)$), for whom $U_1(w_1, R_1; \Theta) + \tilde{\mu}_1(w_1) = U_0(w_1; \Theta)$. The land price for each plot of each type $m > 1$ is determined by the income $w_m$ of the individual indifferent between $m$ and $m-1$, noting that $\mu$ does not influence the choice of plot type conditional on choosing 20k areas: $U_m(w_m, R_m; \Theta) = U_{m-1}(w_m, R_{m-1}; \Theta)$. In Appendix B.2 we provide details to show that, given $w_1$, we can solve for $\tilde{\mu}(w)$, $\{w_m\}_{m=2}^6$ and $\{R_m\}_{m=1}^6$ that equate demand and supply for each $m$ and overall.

**Model estimation.** To estimate the model, we first take values for the parameters $(\alpha, \beta, \delta, \rho, \phi, p)$ from the literature, as reported in panel A of Table 8 and normalize $r$ and $A$ to be 1. We conducted robustness checks around these values and results are not sensitive to plausible variations. Panel B of the table reports estimates of $w^{shape}$ and $w^{scale}$, discussed above.

This leaves six parameters $(\theta, B_F, B_N, \mu^{shape}, \mu^{scale}, w_1)$ to estimate, which we do by minimizing a loss function that matches twelve moments in the data with moments in the model. We use two moments for each of the six plot types: (1) the average price of land $R_m$ and (2) the fraction built by 2020 $S_m$. The loss function is the relative absolute deviation of prices and shares built both weighted by the relative share of plots of that type, i.e., $\sum_{m=1}^{6} \frac{N_m}{N} \left( \frac{|R_m^{Model} - R_m^{Data}|}{R_m^{Model} + R_m^{Data}} + \frac{|S_m^{Model} - S_m^{Data}|}{S_m^{Model} + S_m^{Data}} \right)$. The values of the parameters estimated by moment matching are reported in panel C of Table 8, and more details on parameter selection and estimation are provided in Appendix B.3.

While the characteristics of the equilibrium are discussed next, we make two observations about the parameters. First, $\theta = 0.0671$ implies that the city amenity equals the near 20k amenity ($B_N = 0.146$) after about 28 years. Second, the lowest income owner in 20k has $w_1 = \$579$,

[27] These data are from Balboni et al. (2020), and were kindly provided by the authors. We use total net household income as reported by each household head. The gamma distribution fits the income data better than log-normal or pareto distributions.

equivalent to the 15th percentile of household income in Dar es Salaam, and an extreme taste draw where $\tilde{\mu}_1(w_1)$ is at the 99.4 percentile. We also experimented with exogenously setting the minimum income in 20k at 1200 - roughly equal to the 10th percentile of the income in our data on 20k owners and the 32nd percentile of income in Dar es Salaam. This doubles the estimated loss function, but the other parameter estimates are similar.

### 6.2 Characteristics of an equilibrium

Figure 5 illustrates the equilibrium allocations, prices, and choices. In panel A, each plot type (amenity and size) has a different colored curve. Sorting is paramount as higher-income people sort into larger and higher-amenity plots, as in the data. The solid sections of the curves show the outer envelope of equilibrium realized lifetime utilities net of $U_0(w;\Theta)$ for a common value of $\tilde{\mu}$, which we set to $\tilde{\mu}_1(w_1)$. In the figure, no one with a common $\tilde{\mu}$ wants to switch from their plot type $m$ to a plot with different $m'$. The person with $w_1$ is indifferent between being in 20k and the city. Other owners with higher $w$ and the same $\tilde{\mu} = \tilde{\mu}_1(w_1)$ strictly prefer 20k, noting the locus of people indifferent between 20k and the city has lower $\tilde{\mu}$'s, i.e. $\tilde{\mu}_1(w) < \tilde{\mu}_1(w_1), \forall \in (w_1, \bar{w}]$.[28] Panel B of Figure 5, shows that the land price per square meter is higher in the high-amenity 'Near' areas ($B_N$), and for the same amenity, the price per square meter decreases in plot size, as expected.

In panel C of Figure 5, the solid segments of the curves show how the choice of $\tau$ varies by plot type. Holding income and amenity fixed, switching to a larger plot reduces $\tau$, as illustrated by the vertical dashed line between $l_S$ and $l_M$, which demonstrates the plot-size effect in equation (2). However, between switching points, $\tau$ increases in income for the same $B$, $R$ and $l$, illustrating the sorting effect in equation (2). This occurs because the relative deterioration of the city amenity is less important to richer people than the unrestricted housing choice that the city offers. The model and the data. In Table A.12 we show how well we match on prices and share built in the Near and Far areas. The model and data moments are remarkably close in most cases. One exception is medium-size plots in Near where prices and share built are off by 25% to 30%. For share built, we are noticeably off for the small fraction (2.6%) of plots that are large and Far (noting the loss function is weighted). For everything else, numbers are close. Second, as discussed in Section 5.2.2, in the data mean education of owners of built plots increases over time for the same plot size, and across plots sizes in the same time period, both consistent with panel C of Figure 5. Third, in terms of our empirics and the model, panel C of Figure 5 shows how, holding $l$ fixed, $\tau$ may drop as $B$ rises; this pattern, where better-amenity plots are built earlier, is suggested by the reduced-form empirics. In assessing this, we note that below the horizontal line at $\tau = 15$, the share of the grey line is smaller than that of the purple line. But holding amenities constant, larger plots need not be built earlier or later than smaller plots, consistent with our empirical findings.

[28] $\tilde{\mu}(w)$ locus is the scalar $\tilde{\mu}_1(w_1)$ minus the outer loop of utilities of Figure 5. It is generally downward sloping.

Panel D of Figure 5 shows that for the same income, the ambiguous effect in equation (3) of an increase in $l$ or $B$ very slightly decreases investment. However, what the figure shows most clearly is that investment increases sharply with income and is little related to either $l$ or $B$, consistent with the reduced-form empirics where sorting effects dominate the ambiguous effects of $l$ and $B$ changes in Eq. 3. That said, the elasticity of property value in the model (capital investment plus land value) with respect to plot size is considerably larger (at 0.82) than the elasticity of reported property value with respect to plot size in the our survey data (0.45).[29]

### 6.3 Counterfactuals

In this section, we use the model to change the number of small, medium, and large plots planned within Near and Far 20k areas according to different objective functions, holding constant the total land area in Near and Far. First, we solve for the welfare-maximizing allocation of a Kaldor-Hicks social planner. Second, we solve for the allocation that maximizes the land values of either a monopolist developer who chooses plot size allocations in 20k to extract maximal profits or a local government aiming to maximize its revenues from land value capture.

The social planner's solution maximizes the sum of total land values in 20k areas and the compensating variation of residents. The latter is the present discounted value of the annual income supplement $\Delta$, which would leave 20k residents indifferent between being in their 20k plot and the city, i.e. $U_{m^*}(w, R_{m^*}; \Theta) + \tilde{\mu} = U_0(w + \Delta; \Theta)$. This money-metric approach, rather than utility summation, circumvents the need to impose assumptions on the allocation of land values across households. Since it weighs the consumer surpluses of the rich and poor equally, it does not drive our finding (below) that welfare is maximized by further accommodating the housing needs of poorer residents.[30]

The results in Table 9 show that compared to the baseline (column 1), welfare maximization (column 2) increases consumers' surplus by 24% and total surplus by 3.2%. The number of plots increases by 37%, average plot size declines by 27%, and the median owner income of 20k residents declines by 20%. Thus, most crucially, this allocation widens participation in the project to serve many more lower income people.[31] In contrast, a monopoly developer (column 3), chooses a very different allocation. Relative to the baseline, in Table 9, it *reduces* the supply of plots by 16%, *increases* average plot size by 20% and *increases* median owner income by 11%. This increases total land values by 2.1% and reduces consumers' surplus by 12%. Notably, on all dimensions, the

[29] To validate the model on moments not used in estimation, we use the assessments of residents in Near areas of their current property values, which correspond to $R + rk$. We focus on the elasticity of these values with respect to plot size, mitigating concerns about assessment levels. The model elasticity for the same plots is computed using property values of small, medium, and large plots for residents with $\tau \leq 15$ in the Near areas.

[30] Furthermore, the welfare gains of accommodating poorer residents are stronger when measured in utility rather than surplus, as these poorer residents have higher marginal utility.

[31] In this counterfactual, the share built upon by 2020 declines by 8.6%.

baseline lies between the solutions entailed by welfare maximization and land value maximization, although it is closer in percentage terms to the land value maximization.[32]

The plot type allocations, prices, and lowest income residents for each of the baseline, social planner, and monopoly solutions are shown in Table 10. In the baseline (panel A), as in the reduced-form results, price per square meter decreases in plot size for the same $B$. Under welfare maximization (panel B), prices per square meter *increase* in plot size, although they are almost equalized.[33] In Appendix Section B.4, we show under fairly general conditions that the planner wants to subdivide large plots into smaller plots past the point of land price per sqm equalization and up to the point where the consumer surplus gain from accommodating more households is offset by the loss in land values. That point occurs when land prices sqm rise with size, which provides a sufficient statistic that could rule out some planning misallocations.

For a land-value-maximizing monopolist (panel C), price per square meter declines in plot size, and more steeply than in the baseline. The intuition is that subdividing a large plot increases land values only if the 'partial equilibrium' gains from splitting that plot outweigh the overall general equilibrium loss in land values from overall land prices declining with the increased supply of plots. This only happens when price per square meter declines sharply in plot size.

Another takeaway is that welfare maximization increases population density. This, in turn, could generate (unmodeled) positive neighborhood agglomeration effects, such as greater provision of local services, which should induce the social planner to further increase the number of small plots. However, there may also be negative externalities from crowding or including lower-income residents who are less able to contribute public goods, which should lead the social planner to provide fewer small plots.

The model also allows us to consider an incremental counterfactual, where one medium-sized "near" $(l_M, B_N)$ plot is divided into two $(l_S, B_N)$ plots, increasing its total land value by $1200 \times (8.31 - 6.67) = 1968$ as a partial equilibrium effect. But adding one plot to the total supply induces general equilibrium effects, where land prices per sqm fall slightly (less than 0.001% on average), which on aggregate reduces the total land value by 2816. The overall total effect is therefore to reduce land values by 848. Conditional on unchanging allocations of households to plots, the general equilibrium decline in prices generates an equal loss in land values as a gain in consumer surplus, as shown in Appendix B.4. However, in this example, due to the reallocation of some households with mass and positive surplus from the city to 20k areas, consumer surplus increases beyond the price decline gains, raising overall welfare by 2189. This sheds light on the effects of rigidities that prevent the splitting of formal plots: a preponderance of large plots restricts entry and props up

[32] As discussed in the estimation section, we also examined the case where the minimum income in 20k is exogenously set at 1200, which approximately excludes the bottom third of Dar es Salaam's population. Even then, welfare maximization entails roughly doubling the number of small plots in both areas compared to the baseline.

[33] Note that in the far community, no large plots are provided, so they do not have a well-defined price.

the value of 20k plots, but lowers aggregate welfare.

## 7 Concluding Remarks

De-novo urban planning provides a key policy option for developing country cities faced with large and rapidly growing informal areas. Despite its importance for Africa's large and growing cities, such planning is not sufficiently informed by economic analysis. This paper provides, to the best of our knowledge, the first systematic economic analysis of the decisions that de-novo planning entails. We construct and estimate a novel model, complemented by reduced-form analysis, using new data that we collect. The setting we study is the 20k plot project in Dar es Salaam.

Earlier World Bank de-novo projects were halted in the 1980s due to criticism that costs were not recouped and the poor were excluded. We find that 20k plots were cheaper and their costs were recouped, although the poor were still largely excluded. We also uncover two key reasons for the 20k project's success in roughly doubling land values: the protection of owners' property rights and the preservation of access through local unpaved roads, which connect to main paved ones.

Nevertheless, de-novo neighborhoods also have limitations, which can be mitigated. Key among these is an oversupply of large plots, which command lower land values, a small share of built space, and lower population density. A likely cause for this over-provision is the persistence of colonial-era rules and norms. Our evidence suggests that offering relatively more smaller plots would make de-novo projects more valuable and more inclusive, raising social welfare and allowing more people to benefit from affordably priced formal plots. The current equilibrium seems to be more aligned with one where planners sought to maximize land values, rather than social welfare. Our findings also indicate that non-residential amenities are largely ignored, generating neither land value appreciation nor more built activity, and 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 plot settlement.

## References

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

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

Adukia, A., Asher, S., Jha, K., Novosad, P. and Tan, B. (2024), Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.2m Neighborhoods. working paper.

Almagro, M. and Dominguez-Iino, T. (2024), Location Sorting and Endogenous Amenities: Evidence from Amsterdam. Forthcoming at *Econometrica*.

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

Anagol, S., Ferreira, F. V. and Rexer, J. M. (2021), Estimating the economic value of zoning reform. NBER working paper 29440.

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

Balboni, C., Bryan, G., Morten, M. and Siddiqi, B. (2020), Transportation, gentrification, and urban mobility: The inequality effects of place-based policies.

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', *Journal of Political Economy* **115**(4), 588–638.

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

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

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.

Calonico, S., Cattaneo, M. D., Farrell, M. H., Palomba, F. and Titiunik, R. (2025), Treatment effect heterogeneity in regression discontinuity designs.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014), 'Robust data-driven inference in the regression-discontinuity design', *The Stata Journal* **14**(4), 909–946.

Castells-Quintana, D. (2017), 'Malthus Living in a Slum: Urban Concentration, Infrastructure and Economic Growth', *Journal of Urban Economics* **98**, 158–173.

Centre for Affordable Housing Finance (2024), Housing finance in africa yearbook: 15th edition – 2024, Technical report, Centre for Affordable Housing Finance.

Chiumenti, N., Kulka, A. and Sood, A. (2022), 'How to Increase Housing Affordability: Understanding Local Deterrents to Building Multifamily Housing', (4241563).

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.

Combes, P.-P., Duranton, G. and Gobillon, L. (2021), 'The Production Function for Housing: Evidence from France', *Journal of Political Economy* **129**(10), 2766–2816.

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

Davis, O. A. and Whinston, A. (1964), 'The economics of complex systems: The case of municipal zoning', *Kyklos* **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"', *Econometrica* **78**(6), 1863–1903.

Dixit, A. (1973), 'The optimum factory town', *The Bell Journal of Economics and Management Science* **4**(2), 637–651.

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

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

Gechter, M. and Tsivanidis, N. (2023), Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills.

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

Glaeser, E. L. and Ward, B. A. (2009), 'The causes and consequences of land use regulation: Evidence from Greater Boston', *Journal of Urban Economics* **65**(3), 265–278.

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.

Gyourko, J. and Krimmel, J. (2021), 'The impact of local residential land use restrictions on land values across and within single family housing markets', *Journal of Urban Economics* **126**, 103374.

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

Harari, M. (2024), Residential Patterns and Local Public Goods in Urban Brazil. Working paper.

Harari, M. and Wong, M. (2024), Long-term impacts of slum upgrading: Evidence from the Kampung Improvement Program in Indonesia. Forthcoming at *Review of Economic Studies*.

Henderson, J. V., Regan, T. and Venables, A. J. (2021), 'Building the city: From slums to a modern metropolis', *The Review of Economic Studies* **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', *Land Use Policy* **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', *Current Urban Studies* **3**(4), 348–367.

Larson, W. D. and Shui, J. (2022), 'Land valuation using public records and kriging: Implications for land versus property taxation in cities', *Journal of Housing Economics* **58**, 101871.

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

Manara, M. and Pani, E. (2023), 'Institutional hybrids through meso-level bricolage: The governance of formal property in urban tanzania', *Geoforum* **140**, 103722.

Manara, M. and Regan, T. (2024), 'Unbundling tenure security and demand for property rights: Evidence from urban Tanzania', *Urban Studies* **61**(11), 2080–2098.

Manara, M. and Regan, T. (2025), 'Ask a local: Improving the public pricing of land titles in urban Tanzania', *The Review of Economics and Statistics* **107**(1), 188–203.

Mayo, S. K. and Gross, D. J. (1987), 'Sites and services—and subsidies: The economics of low-cost housing in Developing Countries', *The World Bank Economic Review* **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', *Journal of Political Economy* **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.

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.

Nagpal, G. and Gandhi, S. (2024), Scaling Heights: Affordability Implications of Zoning Deregulation in India . working paper.

National Bureau of Statistics (2019), 'National Panel Survey 2014-2015 [Data set]'.

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', *Philosophy & Geography* **4**(1), 25–48.

Rios-Avila, F. (2020), 'Smooth varying-coefficient models in Stata', *The Stata Journal* **20**(3), 647–679.

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', *Journal of Urban Economics* **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', *Journal of Planning History* **6**(1), 3–47.

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

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', *Memoirs of the American Academy in Rome* **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', *Habitat International* **49**, 413–418.

Turner, M. A., Haughwout, A. and van der Klaauw, W. (2014), 'Land use regulation and welfare', *Econometrica : journal of the Econometric Society* **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 (2022), Administrative units population distribution report, Technical report, MFP, NBS and PO-OCGS, Tanzania.

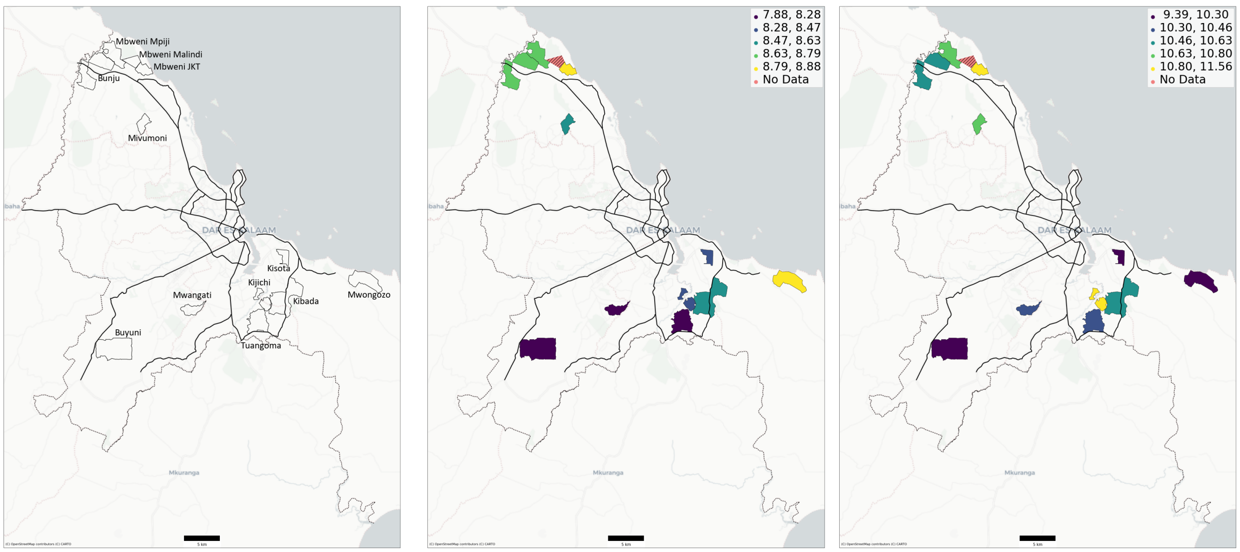
United States Geological Survey (2000), 'SRTM1 Arc-Second Global'. DOI: /10.5066/F7PR7TFT.

Yamasaki, J., Nakajima, K. and Teshima, K. (2023), 'From samurai to skyscrapers : How transactions costs shape Tokyo', *TDB-CAREE Discussion Paper Series* .

# Figures

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

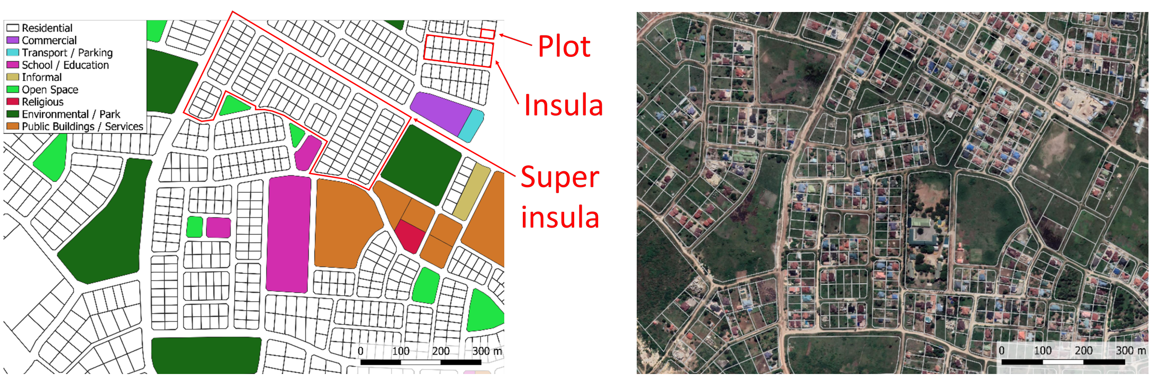
(A) 20k project area names (B) Initial ln price per sqm (C) Current ln price per sqm




*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) Satellite imagery



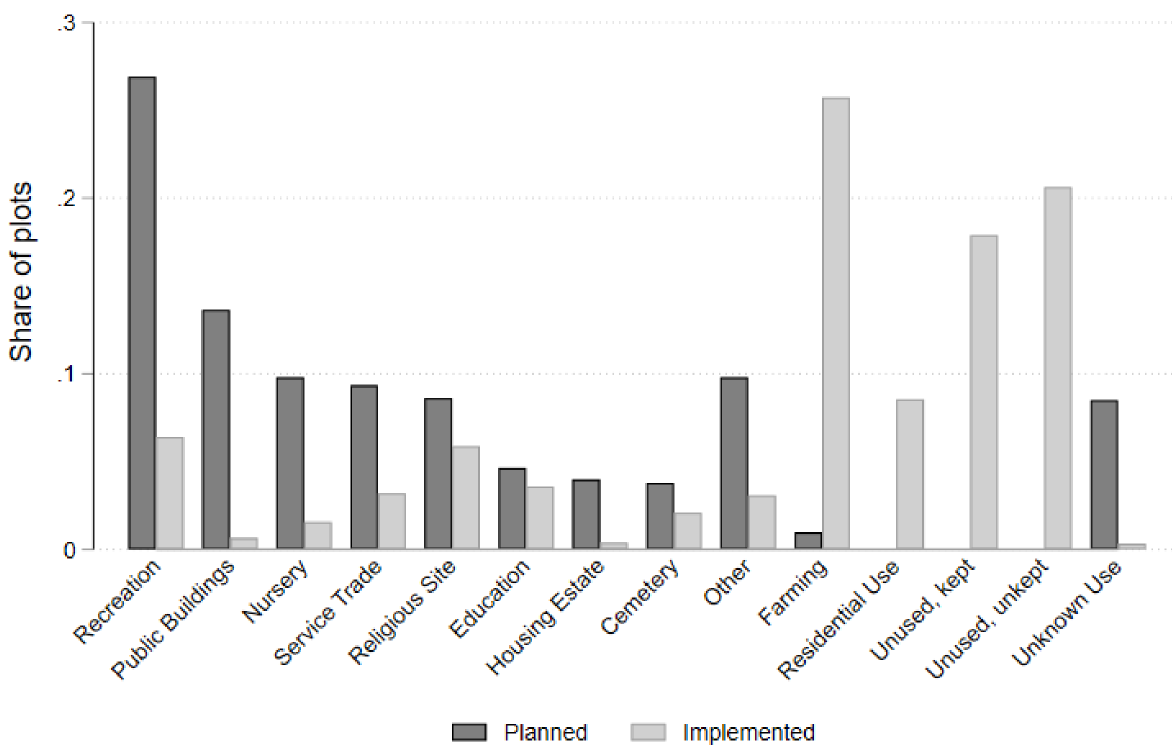

*Notes:* This figure shows an example of planned plot boundaries in Mbweni Mpiji. In Panel A, each plot is colored by its planned use. In Panel B, satellite imagery is displayed in the background.

Figure 3: Example of 20k boundary in Tuangoma

(A) Satellite imagery in 2001 (pre-implementation) (B) Satellite imagery in 2021 (post-implementation)

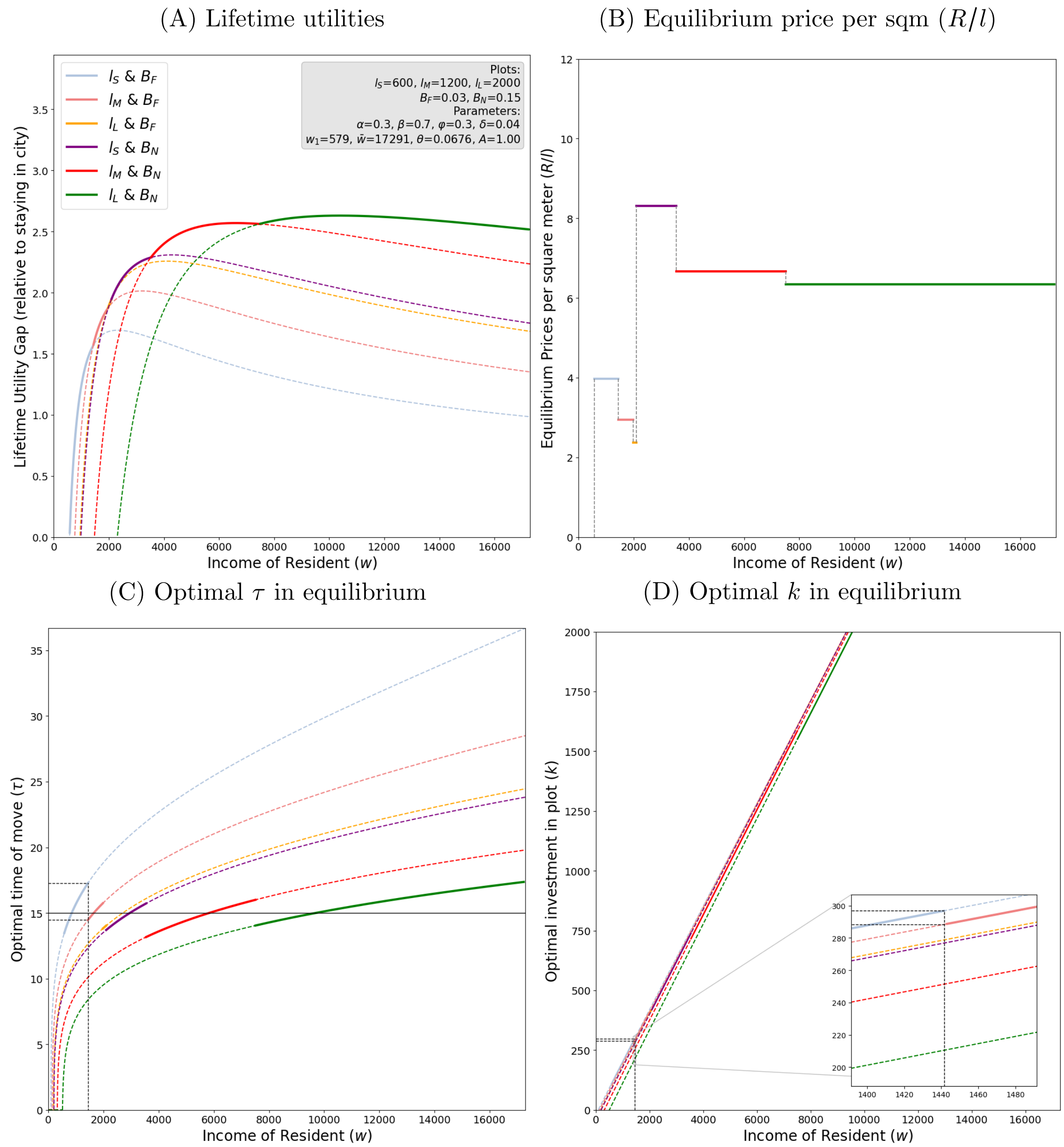

*Notes:* This figure shows an example of a 20k project boundary in Tuangoma. In Panel A, background satellite imagery is from 2001 (pre-implementation). In Panel B, background satellite imagery is from 2021 (post-implementation).

Figure 4: Non-residential plots by planned and implemented uses




*Notes:* This figure shows the share of planned non-residential plots by planned use (dark grey) and current/implemented use (light grey).

Figure 5: Equilibrium with varying amenities




*Notes:* This figure shows equilibrium model outcomes vs resident income for different plot sizes and amenity values. In panels (C) and (D), the vertical dashed black line denotes the cross-over level of $w$ from the lowest valued plot to the second lowest, and the horizontal lines depict the optimal choice for each type of plot at this income level. Panel (A) plots the lifetime utilities as outcomes. Note the solid parts form an outer-envelope of realized net utilities for $\mu = \tilde{\mu}(w_1) = 0.81$, 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 $\tau$ varies by income for each plot size-$B$ combination. The solid parts of the lines show the realized $\tau$'s in equilibrium, while the dashed lines show out-of-equilibrium choices of $\tau$. In panel (D), the outcome is the optimal capital investment.

# Tables

Table 1: Land price inside and nearby 20k areas

| | (1) | (2) | (3) |
|---|---|---|---|
| | Ln Price | Ln Price | Ln price per sqm |
| Ln plot size | 0.71 | 0.69 | |
| | (0.054) | (0.041) | |
| Non-20K Surveyed | -0.23 | -0.27 | -0.24 |
| | (0.16) | (0.12) | (0.11) |
| Non-20K Unsurveyed | -0.70 | -0.71 | -0.43 |
| | (0.099) | (0.079) | (0.17) |
| Mean Outcome | 17 | 17 | 10 |
| 20K or Nearest FE | | ✓ | ✓ |
| N | 2074 | 2074 | 2074 |

*Note:* This table presents regressions of log price on log plot size and planned/surveyed status. The outcome in cols 1-2 is the log price of a bare land transaction, and in col 3 it is the log price per square meter. Each observation is a transaction: 1246 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). Cols 2-3 include fixed effects for the the nearest 20k area. Standard errors in parentheses are clustered by 20K area.

Table 2: Prices and plot sizes in 20k areas (OLS)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Ln Price per sqm | Ln Price per sqm | Ln Price per sqm | Ln Price |
| Ln plot size | -0.45 | -0.55 | -0.52 | 0.48 |
| | (0.072) | (0.053) | (0.060) | (0.060) |
| Mean Outcome | 10 | 10 | 10 | 17 |
| Mtaa*20k FE | | ✓ | ✓ | ✓ |
| Amenities | | | ✓ | ✓ |
| N (gridcells) | 4074 | 4074 | 4074 | 4074 |
| N (plots) | 1446 | 1446 | 1446 | 1446 |

*Note:* This table presents regressions of plot price on plot size. Prices combine bare land transactions from the dalali and occupier surveys. The outcome in cols 1-3 is log plot price per square meter, and col 4 it is log plot price. 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: 34 Mtaa*20k Area FEs (cols 2-4), and amenities (cols 3-4). 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. NB: wetland within 100m dummy is perfectly collinear, and so dropped. Standard errors in parentheses are clustered by insula.

Table 3: 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: Mtaa*20k 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: Mtaa*20k FE + Amenity controls** | | | | |
| Ln plot size | -0.078 | 0.000075 | 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 up to 42 Mtaa*20k Area FEs and panel B adds amenities. Amenities are the same as described in Table 2. Standard errors in parentheses are clustered by insula.

Table 4: 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 | -0.18 | -0.017 | 0.0096 | -0.0089 | 0.031 |
| | (0.052) | (0.0024) | (0.0076) | (0.017) | (0.012) |
| Mean Outcome | 9.9 | 0.11 | 0.49 | 5.3 | 0.38 |
| N (gridcells) | 3581 | 93580 | 93580 | 45383 | 45383 |
| N (plots) | 1253 | 36035 | 36035 | 17583 | 17583 |
| **Panel B: gap≥400sqm** | | | | | |
| Own Larger | -0.50 | -0.038 | -0.0055 | 0.015 | 0.090 |
| | (0.14) | (0.0048) | (0.015) | (0.041) | (0.029) |
| Mean Outcome | 9.9 | 0.092 | 0.47 | 5.3 | 0.42 |
| N (gridcells) | 1021 | 23872 | 23872 | 10974 | 10974 |
| N (plots) | 341 | 9661 | 9661 | 4410 | 4410 |
| **Panel C: gap<100sqm** | | | | | |
| Own Larger | -0.11 | -0.0072 | 0.010 | -0.0076 | -0.0044 |
| | (0.063) | (0.0041) | (0.013) | (0.025) | (0.018) |
| Mean Outcome | 10 | 0.13 | 0.50 | 5.2 | 0.35 |
| N (gridcells) | 1048 | 32780 | 32780 | 16209 | 16209 |
| N (plots) | 485 | 16448 | 16448 | 8108 | 8108 |

*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 3 notes. Controls always include Mtaa*20k Area and insula-segment FEs, and amenities. Amenities are the same as described in Table 2. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table 5: Prices, built outcomes and plot sizes in 20k areas (RD with size gap interaction)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ln Price per sqm | Share gridcell built | Plot is built | Log area of buildings | Multiple buildings on plot |
| **Panel A: RD across insulae with interaction for difference in log mean size** | | | | | |
| Own Larger $\times\Delta$ ln mean size | -0.63 | -0.056 | -0.055 | 0.053 | 0.17 |
| | (0.25) | (0.0086) | (0.026) | (0.076) | (0.051) |
| Own Larger | -0.060 | -0.0037 | 0.021 | -0.020 | -0.0052 |
| | (0.054) | (0.0031) | (0.010) | (0.023) | (0.016) |
| Mean Outcome | 9.9 | 0.11 | 0.49 | 5.3 | 0.38 |
| N (gridcells) | 3581 | 93580 | 93580 | 45383 | 45383 |
| N (plots) | 1253 | 36035 | 36035 | 17583 | 17583 |
| **Panel B: OLS with RD sample from panel A** | | | | | |
| Ln plot size | -0.53 | -0.079 | -0.0015 | 0.14 | 0.19 |
| | (0.065) | (0.0026) | (0.0094) | (0.019) | (0.012) |
| Mean Outcome | 9.9 | 0.11 | 0.49 | 5.3 | 0.38 |
| N (gridcells) | 3581 | 93580 | 93580 | 45383 | 45383 |
| N (plots) | 1253 | 36035 | 36035 | 17583 | 17583 |

*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. To estimate heterogeneity in treatment effects we follow interact the difference in log mean plot size across the boundary of the insula-pair with all independent variables in the model (Calonico et al., 2025). 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 3 notes. Controls always include Mtaa*20K Area FEs, and amenities. Amenities are the same as described in Table 2. 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 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 | 0.48 | -0.078 | 0.000075 | 0.14 | 0.19 |
| | (0.060) | (0.0026) | (0.0094) | (0.018) | (0.012) |
| Dist (km) paved major road | -0.14 | -0.015 | -0.041 | -0.063 | -0.040 |
| | (0.032) | (0.0016) | (0.0071) | (0.012) | (0.0088) |
| Elevation (m) | 0.0024 | 0.00089 | 0.0028 | 0.0032 | 0.00035 |
| | (0.0024) | (0.000098) | (0.00043) | (0.00067) | (0.00049) |
| Ruggedness | -0.0097 | -0.0058 | -0.016 | -0.011 | -0.0096 |
| | (0.022) | (0.00098) | (0.0039) | (0.0090) | (0.0052) |
| River/stream 100m | -0.012 | -0.027 | -0.11 | -0.061 | -0.039 |
| | (0.17) | (0.0052) | (0.022) | (0.058) | (0.048) |
| Water/wetland 100m | | 0.0078 | -0.067 | -0.080 | 0.052 |
| | | (0.0089) | (0.032) | (0.16) | (0.22) |
| Z-index: 3 Ins. Characteristics | 0.025 | 0.0029 | 0.016 | 0.0052 | 0.0061 |
| | (0.028) | (0.0014) | (0.0058) | (0.010) | (0.0068) |
| 20k edge in 100m | 0.012 | -0.0041 | -0.0099 | -0.033 | 0.011 |
| | (0.043) | (0.0023) | (0.0096) | (0.016) | (0.011) |
| Pln. recreation in 100m | -0.016 | -0.00089 | -0.0085 | -0.011 | -0.0060 |
| | (0.040) | (0.0019) | (0.0071) | (0.012) | (0.0089) |
| Pln. nursery school in 100m | 0.079 | 0.0060 | 0.017 | 0.029 | 0.0051 |
| | (0.043) | (0.0026) | (0.0097) | (0.017) | (0.013) |
| Pln. religious site in 100m | 0.036 | 0.0021 | 0.015 | -0.0083 | -0.0077 |
| | (0.054) | (0.0030) | (0.012) | (0.020) | (0.015) |
| Pln. education in 100m | 0.15 | -0.0048 | -0.0095 | -0.026 | -0.0025 |
| | (0.074) | (0.0030) | (0.011) | (0.021) | (0.014) |
| Pln. service trade in 100m | 0.039 | -0.0023 | -0.011 | -0.0041 | -0.012 |
| | (0.086) | (0.0040) | (0.015) | (0.028) | (0.019) |
| Pln. housing estate in 100m | -0.015 | 0.00092 | 0.0013 | 0.011 | -0.036 |
| | (0.093) | (0.0073) | (0.030) | (0.046) | (0.032) |
| Pln. public building in 100m | 0.097 | -0.0060 | -0.018 | -0.041 | -0.029 |
| | (0.079) | (0.0041) | (0.015) | (0.027) | (0.019) |
| Pln. cemetery in 100m | 0.15 | 0.0034 | 0.029 | -0.045 | 0.0014 |
| | (0.12) | (0.0048) | (0.018) | (0.032) | (0.022) |
| Pln. any other non-res in 100m | 0.019 | -0.0017 | -0.011 | 0.0038 | 0.025 |
| | (0.059) | (0.0025) | (0.0100) | (0.019) | (0.013) |
| 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 3 notes. Controls always include Mtaa*20k FEs and amenities. Amenities are the same as described in Table 2. 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 perfectly collinear with other controls, and so dropped from the regression. 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 | -0.0020 | -0.0018 | -0.00049 | 0.024 |
| | (0.0025) | (0.010) | (0.020) | (0.015) |
| Own Smaller × Dist. (km) | 0.056 | 0.23 | -0.044 | 0.011 |
| | (0.017) | (0.067) | (0.12) | (0.084) |
| Own Larger × Dist. (km) | -0.031 | 0.026 | 0.10 | -0.039 |
| | (0.018) | (0.070) | (0.13) | (0.089) |
| Ln plot size | -0.065 | 0.027 | 0.18 | 0.20 |
| | (0.0032) | (0.013) | (0.027) | (0.018) |
| Mean Outcome | 0.11 | 0.49 | 5.3 | 0.38 |
| N (gridcells) | 93025 | 93025 | 45712 | 45712 |
| N (plots) | 35658 | 35658 | 17562 | 17562 |

*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 64m, median 40m, 75th percentile 97m, and 95th percentile 214m. In columns 1-4 the outcomes are the same built outcomes as described in Table 3 notes. Controls always include Mtaa*20k FEs, super-insula-segment FEs, and amenities. Amenities are the same as described in Table 2. Standard errors in parentheses are clustered by insula.

Table 8: Parameters

| Parameter | Description | Value | Source |
|---|---|---|---|
| *—Panel A. External sources—* | | | |
| $\alpha$ | Consumption housing elasticity | 0.3 | Combes et al. (2021) |
| $\beta$ | Consumption numeraire elasticity | 0.7 | Combes et al. (2021) |
| $\delta$ | Interest rate | 0.04 | Henderson et al. (2021) |
| $\rho$ | Discount rate | 0.04 | Henderson et al. (2021) |
| $\phi$ | Housing land elasticity | 0.3 | Combes et al. (2021) |
| $p$ | Rental price of city housing | 2.19 | Henderson et al. (2021) |
| *—Panel B. City Distribution estimation—* | | | |
| $w^{\text{shape}}$ | Shape of city distribution | 1.1703 | Balboni et al. (2020) + Estimation |
| $w^{\text{scale}}$ | Scale of city distribution | 2354.2 | Balboni et al. (2020) + Estimation |
| *—Panel C. Moment estimation—* | | | |
| $w_1$ | 20k marginal entrant wage | 578.6 | Joint Internal Estimation |
| $B_F$ | Far Amenity | 0.029 | Joint Internal Estimation |
| $B_N$ | Near Amenity | 0.146 | Joint Internal Estimation |
| $\theta$ | City deterioration rate | 0.0671 | Joint Internal Estimation |
| $\mu^{\text{shape}}$ | Shape of taste shock | 0.043 | Joint Internal Estimation |
| $\mu^{\text{scale}}$ | Scale of taste shock | 1.710 | Joint Internal Estimation |

*Notes:* We structurally estimate the 6 parameters using SMM. To do so, we minimize the distance between the 12 model-simulated moments, $M$(Model), and their empirical counterparts, $M$(Data), by searching over the parameter space, using the Differential Evolution (DE) algorithm, of the family of Evolutionary Algorithms. Simulated Annealing performed worse than DE. The loss function is the relative absolute deviations of 6 land prices and 6 occupancy rates by 2020 weighted by the share of plots of each type.

Table 9: Counterfactual

| Description | Baseline (1) | Welfare Max (2) | Land Values Max (3) |
|---|---|---|---|
| | *—Panel A. Info—* | | |
| Total supply of 20k plots | 36,215 | 49,714 | 30,249 |
| Average plot size | 1046 | 762 | 1252 |
| 20k mean wage ($ 2021) | 3858 | 3247 | 4211 |
| 20k median wage ($ 2021) | 3200 | 2549 | 3564 |
| 20k Occupancy Rate % (by 2020) | 56.9 | 52.0 | 55.6 |
| | *—Panel B. Values—* | | |
| Total Land Values (M) | 225.5 | 198.1 | 230.3 |
| Consumer Surplus (M) | 166.3 | 206.2 | 146.7 |
| Total Surplus (M) | 391.8 | 404.2 | 377.0 |

*Notes:* The number of potential 20k households, $N^D$, is 800,000. Mean and median household yearly "disposable" income in Dar Es Salaam is $2755 and $2120 , respectively.

Table 10: Counterfactual plot type information

| **Panel A: Baseline** | | | | | | |
|---|---|---|---|---|---|---|
| | **Far Areas** | | | **Near Areas** | | |
| | **S** | **M** | **L** | **S** | **M** | **L** |
| Number of Plots | 4975 | 4192 | 938 | 10068 | 12661 | 3381 |
| Land Price per Sqm | 3.98 | 2.94 | 2.37 | 8.31 | 6.67 | 6.34 |
| Lowest Income | 579 | 1442 | 1977 | 2097 | 3544 | 7507 |
| **Panel B: Welfare Maximization** | | | | | | |
| | **Far Areas** | | | **Near Areas** | | |
| | **S** | **M** | **L** | **S** | **M** | **L** |
| Number of Plots | 15096 | 696 | 0 | 23871 | 8036 | 2015 |
| Price per Sq Meter | 1.93 | 2.07 | - | 6.28 | 6.47 | 6.56 |
| Lowest Income | 579 | 1646 | - | 1706 | 4820 | 8733 |
| **Panel C: Land Values Maximization** | | | | | | |
| | **Far Areas** | | | **Near Areas** | | |
| | **S** | **M** | **L** | **S** | **M** | **L** |
| Number of Plots | 3839 | 3927 | 1437 | 5628 | 7246 | 7872 |
| Price per Sq Meter | 4.83 | 3.47 | 2.76 | 9.67 | 7.34 | 6.14 |
| Lowest Income | 579 | 1627 | 2273 | 2508 | 3565 | 5404 |

*Notes:* This table compares plot characteristics under three scenarios: Baseline, Land Value Maximization, and Land Values Maximization. Each panel includes six columns representing different plot sizes (S, M, and L) in Far and Near Areas.

# Supplementary Appendix
# Economics of Greenfield Urban Planning

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

## 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 the planners, which we have for all project areas, except Mwongozo. These TPDs are also called "neighborhood layouts", since they depict residential plots, non-residential plots with their planned use, and roads with road reserve widths. These drawings were created as hard copies and approved by the Town Planning Department of the Ministry of Lands, Housing and Human Settlements Development (MLHHSD) between 1997 and 2009. They were scanned and shared with us and we georeferenced them.

Second are survey maps (SMs), which were prepared by the MLHHSD after approval of the 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 involves placing beacons in the ground, typically at block corners, to determine exact coordinates (latitude and longitude) using theodolites and then adding more beacons that align with each plot's corners. Each beacon is then associated with its coordinates, which enables the plot boundaries to be precisely recorded using software. The SMs were also given to us as digital copies, and we transformed them from vectoral drawings (.dwg) into polygon shapefiles (.shp) and georeferenced them.

Finally, we obtained cadastral data from the MLHHSD for the municipalities of Kigamboni, Kinondoni, and Ubungo. These data cover all our project areas except Buyuni, Mwanagati, Tuoangoma, and Kijichi. The cadastral database contains registered SMs, which are approved and recorded in GIS software by the MLHHSD, ready for the issuance of title deeds and land rent (tax) bills. Therefore, while SMs are implemented town planning drawings, cadastral drawings constitute the legally registered version of SMs.[34].

Given the limitations of our three sources, we carefully designed a procedure to assemble a dataset as complete and accurate as possible. Our procedure involved discussions with project

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

secretary and town planners of the the 20,000 plots project, who approved our procedure. 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 the cadastral data. Where SMs are not available (i.e., Mwongozo) we instead use cadastral data. To ensure that the cadastral data are restricted to plots implemented as part of the 20k project, we restrict them to plots that were registered between 2000 and 2010, and fall within the boundaries of the Mwongozo project area.

We draw on the TPDs for two purposes. First, to update the planned use of non-residential plots in our plot boundary data where the SMs are missing this information.[35] Second, we digitize the planned road reserves and their widths by manually tracing the georeferenced TPDs.[36]

## 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.[37] The images cover all project areas with a 500m buffer outside them. We aimed to obtain the most recent clear image of each area, the precise dates of which 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).

**Buildings and fences.** We employed a company, Ramani Geosystems, which specializes in geospatial digitization, to trace buildings from the satellite imagery. Ramani digitized data on (i) building footprints, (ii) roof quality (painted metal or tiled; unpainted metal; or rusted metal), and (iii) whether each building was still under construction. Ramani also traced fences and hedges.

**Roads.** We also use the satellite imagery to trace and classify existing roads. This was done using trained research assistants (RAs) from our field staff team in Dar es Salaam. First, the RAs took the digitized road plans as the starting point. Second, they added road extensions (polylines) wherever a road appeared in the image, but not in the plan. Third, they segmented the roads wherever the roads intersected. Fourth, they classified each road segment's type (footpath, dirt road, or paved) and width (in meters). Road segments that were planned but do not appear in the image were classified as 'no road' and width of zero.

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

[36] 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) our enumeration of implemented uses and current maintenance of non-residential plots (see Section D.5).

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

## A.3 Additional data sources

**Elevation and ruggedness.** We measure elevation (relative to sea level) and ruggedness using a digital elevation model with a horizontal resolution of 1 arc second, or approximately 30 meters (United States Geological Survey, 2000), and a vertical resolution of 1 meter. Following Nunn and Puga (2012), we use the data to calculate the local ruggedness as the standard deviation of elevation over the eight neighbors of each 30 x 30m-cell in the SRTM data.

**Openstreetmap (local geographic features).** We measure the proximity to natural features using data from OpenStreetMap contributors (2017). Specifically, we use Openstreetmap to map (i) rivers and streams; (ii) water bodies and wetlands. We then measure the distance from each of our gridcells to the nearest feature in each of these two features.

**Inflation-adjusted prices.** Throughout the paper, we report prices in 2021 Tanzanian Shillings (TZS), unless otherwise noted. To do so, we use annual inflation rates from Statista (2022), which compiles data published by the International Monetary Fund (IMF). We inflate the prices in Tanzanian shillings for the year $y_0 < 2021$ by the product $\prod_{y=y_0+1}^{2021}(1+i_y)$, where $i_y$ is the inflation rate for the year $y$. According to 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).

**Project costs.** We use the total expected project costs reported in Mero (2008, 2009) to estimate the total project cost and cost per sqm of residential plot. The total expected costs were 29,344mn TZS: 19,968mn TZS (compensation for farms and buildings), 7,376mn TZS (roads), 2,000mn TZS (planning, survey, and overheads). We get a total project cost of $33.1mn USD 2021 by inflating from 2007 to 2021 by a factor of 2.61 and converting to 2021 USD using the exchange rate above. For a total area of 75 square kilometers that is $0.44 per sqm. For 36,000 residential plots that is $919 per plot, and $1.15 per square meter of plot (assuming a mean of 800sqm).

**Initial price of government-sold plots.** We obtained data on the initial price that the government charged when it sold the 20k plots from the 20k project secretary. These data agree with the partial data reported in Mero (2008, 2009), and are used in Mwiga (2011) and Kironde (2015). These initial prices per sqm, which were fixed within each 20K area, are reported in Table A.3.

**Price of plots sold in market transactions.** We collected data on the prices of plots sold in market transactions from questionnaires we administered 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 Questionnaires, interviews, and enumerations data

With the aid of our research assistants, who were 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 example, by having them work in pairs and report to the local mtaa office daily.

### A.4.1 Preliminary interviews with experts

From July 2021 - October 2022 we held ten interviews with eight experts, including government officials and academics, who were involved in key aspects of the 20k project. These interviews focused primarily 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; mtaa 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, who are 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 the mtaa offices in the areas covering all the '20,000 plots' project, which allowed us to collect relevant research permits, ensure stakeholder cooperation, and gather preliminary information, through a questionnaire to local leaders.

To identify the relevant mitaa, we overlapped a map showing the project area boundaries (Section A) with a government map of the mtaa boundaries in Dar es Salaam. We identified 38 mitaa containing the planned project areas. Two research assistants visited these mitaa to verify that the 20k project had been implemented locally and interview the mtaa leaders. We found that the program provided private residential plots in (parts of) 34 mitaa.[38]

**Interview details and protocol.** We interviewed the local leaders from September 2021 - October 2021 and recorded their responses using both paper questionnaires and an ODK app.[39] Two research assistants conducted the interviews in the local language (Swahili), and one of the authors participated remotely. Each interview lasted between 90 minutes and three hours. The target

[38] 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.

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

respondents were the mtaa chairpersons, whose responses we recorded, while executive officers and wajumbe were occasionally present. Given the objective of the questionnaire, the presence of multiple respondents was useful to triangulate and complete the picture. Together, our interviewees included the 34 mtaa chairpersons, 22 mtaa executive officers, and 18 wajumbe.

The interviews we conducted with the mtaa leaders provided information that was directly useful and that we also used to design the questionnaires with real estate agents and residents, which are described in the following sections. Furthermore, we asked the mtaa leaders to provide lists and contacts of real estate agents operating in their mitaa, which proved essential to sample them (see Section A.4.3). Finally, the mtaa leaders confirmed the boundary of their mitaa and the location of the 20k plots within them, which allowed us to amend the digitized boundary layers.

**Interview questions.** The interviews with the local leaders were divided into 11 parts. Part 1 gathered information on the respondents. Part 2 asked about residential plots in the mtaa, including land use statistics, built construction, and processes and opinions on opportunities and constraints to land development. Part 3 inquired about other formal plots with each mtaa, outside of the 20k areas. Part 4 focused on land markets for local 20K plots and non-20K informal plots, including questions about volumes of land sales and estimates of bare land current prices of plots of different sizes; this provided us with one of the sources used for the price data. We also asked about local leaders' involvement in land sales and collected contacts for our real estate agent questionnaire. Part 5 focused on residents' profiles, for example, asking questions about household income in 20K and non-20K areas within the mtaa. Part 6 asked about land titles and other documentation held by landowners. Other parts asked about infrastructure provision in 20K and non-20K areas, including roads and open spaces (Part 7), and electricity, water, and sanitation (Part 8). Part 9 asked about housing units provided by real estate firms and obtained the contact information of those firms. Part 10 asked about other services, including public safety, transportation, and schools, and Part 11 concluded by asking the local leaders to confirm their 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 from November to December 2021 and the second 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 (see above). We obtained from the real estate agents preliminary information including the mitaa in which they operate; whether they operate in 20,000 plot 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 (at least 20 transactions) with the sale of plots in

20k areas; (ii) had experience (at least five transactions) with the sale of non-20k plots in the same mitaa, where such plots exist.[40] 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. This was in part because some real estate agents have other primary occupations, so they could not afford to spend enough time answering our questions. However, through a process of snowballing, we recruited six (round one) and 21 (round two) additional real estate agents who met 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 the real estate agents' reliability, our research assistants (RAs) enumerated all land transactions that the interviewed real estate agents had facilitated. The RAs were supplied with A1 printed maps that displayed the mtaa boundaries and the 20k project area boundaries overlaid on satellite imagery. Using these maps, the real estate agents were asked to identify the plots whose sale they facilitated and physically accompany the RAs to the actual plots. The RAs recorded the sales using paper questionnaires and an ODK app. In some cases, the RAs also manually recorded the plot boundaries on their A1 map. For example, if a transaction involved subdivision (typically outside 20k areas), the RAs traced the original plot boundaries and the subdivided plot boundaries. Furthermore, the RAs traced the boundaries of informal transacted plots. Finally, the data on the sold plots were digitized and added to our digital project map.

**Phase 1 – phone questionnaire.** The phone questionnaire asked questions about the real estate agents' demographics and their experience in supervising sales in the mitaa. 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 also asked about volumes of sales and current prices of bare land for plots of different sizes in 20K versus non-20K areas. Finally, we asked questions about rental prices for unfurnished properties of different sizes in 20K versus non-20K areas. This background information was helpful in designing our questionnaires, but not used as data in the analysis.

**Phase 2 – field questionnaire.**

The field questionnaire recorded for each plot the transaction id and area type (e.g., 20K versus non-20K), the estimated plot size, the period and year of transaction, whether a written record of the time of transaction exists, the price in million TZS and whether a written record of price exists, the plot's development status at the time of the sale, the real estate agents' assessment of information reliability (e.g., of the quality of their recollection) and the enumerator assessment. Open-ended

[40] The threshold numbers of sales were selected since we anticipated that one day of fieldwork would enable us to visit at most 25 plots

questions asked real estate agents to talk about the processes and stages of land transactions in the mtaa, in both 20K and non-20K areas. We also asked questions about the involvement of the mtaa office or formal lawyers in the ratification of bills of sale in 20K and non-20K areas.

**Sample Selection.** In total, we collected information on 2,588 transactions from the field questionnaire , including: 1,666 sales of 20k plots , 311 sales of formal non-20K plots , and 611 sales of non-20K informal plots. We note that formal non-20K plots are surveyed, included in town planning drawings, and eligible for land titles (as 20K plots); however, they were not provided as part of the 20K project. Typically, they result from ex-post regularization of informal plots. Thus, they are formal plots predominantly located in informal neighborhoods. We also note that most real estate agents were able to read maps and were familiar with the mitaa in which they operate, which made the data collection process relatively smooth.

From the data we collected, we assemble a set of transactions of bare land inside and outside 20k areas. We keep only the transactions of plots that were unbuilt 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 involving 20K plots, 297 involving non-20K formal plots, and 600 involving non-20K informal plots.

Next, we match the 1,507 bareland transactions in 20k areas uniquely to our 20k plot data. First, we discard any transactions inside 20k areas that do not match planned 20k plots (possibly subdivisions or formal or informal plots added later on), leaving us with 1,370 bareland transactions. We then impose further restrictions on the data. For plots with multiple bare land records, we keep only the most recent transaction and only if there were no transactions, we keep the listing price. This leaves us with 1,319 plots.

Finally, we discard non-residential plots and plots for which we only know the listing price and where no transaction had yet occurred. This left 1,122 20k plots with transaction prices. We add to these 324 plots with prices recorded from our resident questionnaire to get a total sample size of 1,446 plots with market sales prices.

#### A.4.4 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 with a minimum size of 30 sqm. One of the 20k areas (Mwongozo) was excluded from the resident questionnaire, due to cost-effectiveness considerations, since it has a 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.

Given our budget, 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 meet this target, we assigned each enumerator team a weekly cluster of randomly selected plots, through a process that we hereby describe. Of the 17,333 plots mentioned above, we randomly selected 5,900, and grouped them into questionnaire clusters of approximately 35 plots each.[41]

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).[42] 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.[43]

**Interview details and protocol.**

In June 2022, the fourteen local town planning graduates whom we selected as our enumerators received four weeks of training on the questionnaire, including two weeks under the supervision of one of the authors. These enumerators 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 spared the enumerators the need for long commutes and allowed them to embed themselves in the 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 authors.

Each interview team worked from Wednesday to Sunday each week, to maximize the likelihood of finding the landowners 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 allowed the enumerators to identify plots that were ineligible for data collection or whose eligibility was undetermined.[44]

To all the plots that were eligible for interviews (4,683) and those deemed undetermined (23), the enumerators delivered leaflets written in the local language (Swahili) and signed by the local mtaa chairperson. This leaflet provided introductory information on the research project and the interview that was planned for the weekend. When possible, the enumerators spoke to people living on the plots, and otherwise they left the leaflet attached to the gate or under the door.[45] Prospective

[41] 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.

[42] 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.

[43] Given the complexity of the questionnaire protocol and questionnaires (see next sections), we did not collect statistics on the reasons why some eligible respondents declined to be interviewed. We note that none of our respondents dropped out during their interviews.

[44] 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 had undetermined eligibility, as the enumerators were unable to establish whether their building was inhabited or whether the residents were eligible to be interviewed.

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

respondents could contact the enumerators using the contact details provided in the leaflet to ask the enumerators for clarifications and schedule interview times. These weekly preparatory activities took place in parallel with the enumeration of non-residential plots (see next section).

Within each interview plot, the target respondent was designated as one of the following, in declining order. First, the landowner (named on any property document); second, 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's household, but allowed to live there for free); finally, 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 the preferred category (e.g., joint landowners, multiple usufructuary households, or multiple renting households), we interviewed only one. Guardians and servants (those not part of the landowner's household but paid to live and/or work on the plot) were not interviewed. Where possible, the enumerators tried to interview their target rather than another respondent.

Every four weeks, a catch-up week was organized to allow the enumerators to revisit plots assigned to them in previous weeks, where they did not find the target respondent at home. If the target respondent was still unavailable, the enumerators interviewed 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.

**Questionnaire content.** The questionnaire was structured in 13 parts. Part 1 asked questions about the residents and identified respondents, including the target and (where needed) the proxy. Part 2 collected information on current land uses, while Part 3 focused on road access and plot characteristics (e.g., counts of buildings with residential and non-residential use). Part 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. Part 6 inquired about rental income (where applicable), while Part 7 asked questions about the history of plot acquisition and development, such as the year and mode of acquisition, and the timing of construction. Parts 8, 9, and 10 asked about the respondents' education and employment, including current work or last work before retirement, while Part 11 asked about household wealth and how it is held. Part 12 contained questions about neighborhood amenities and disamenities, residents' contributions to public goods, and perceptions of the local mtaa office. Finally, Part 13 recorded the respondents' assessments of the current value of the property and the enumerators' assessments of the building materials and maintenance condition.

### A.4.5 Non-residential plots enumeration

**Sampling frame.** To enumerate the non-residential plots in the 20k project areas (again excluding Mwongozo), we first used as a reference the 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 planned non-residential land uses, of which we enumerated 1,530 (98%).

**Enumeration details and protocol.** The data on the 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, maintenance, and ownership status. The data the enumerators collected were based 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 activities, enumerators then sought to identify the specific use from a list of sixteen precoded activities. Finally, enumerators noted the maintenance condition of each plot (very well kept, reasonably well kept, abandoned, or encroached by squatters), its ownership (e.g., government or public institutions or private individuals or firms), and the source(s) of information they (the enumerators) had used (e.g., own observation, people who live or work in this plot, local leaders, or neighbors).

### A.4.6 Enumeration of public transport nodes

**Sampling frame.** We also enumerated all public transport access points (e.g., bus stops and others described below) available to residents of the 20k project areas. We note that Mwongozo was included in these surveys since they required less time in the field.

**Enumeration details and protocol.** This enumeration took place from December 2022 - February 2023. For each mtaa that covers part of the 20k areas, 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 the mtaa (i.e., the closest minibus collection point). Our enumerators then visited each access point, asked drivers and passengers questions (described below), 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 it has is a direct route to Kariakoo (the most central location accessible by informal transport, beyond which only formal transport can enter the city center). If not, 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 (a neighborhood in the CBD that is a common commuting destination) 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 model residents face a nested choice problem that involves first choosing whether to remain permanently in the city or move to 20k. If they decide to move, they also choose the type ($m$) of plot they buy, how much capital ($k$) to invest in housing on their plot, and when to move $\tau$. Residents also choose their city housing ($h$) and their consumption in the city ($z_1$) and in 20k ($z_2$). As explained in the text, this problem can be solved by backward induction.

In the final stage, each resident faces the optimization problem as described in the text Eq. 1. Here, the plot size $l$ and amenity $B$ are taken as given (having been determined in earlier stages of the optimization problem). A plot of type $(l, B)$ has a price $R_{l,B}$ at 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, $w$, an individual discount rate $\rho$ and $\delta$ is the interest rate. We equate $\rho$ and $\delta$, so that optimized consumption is constant over time. $A$ is the initial amenity level in the city at time 0 which declines at a rate $\theta$.

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) $\frac{w-z}{\delta} - \frac{ph}{\delta}(1 - e^{-\delta\tau}) = R_{l,B} + rke^{-\delta\tau}$, (5) $\varphi lnh + Ae^{-\theta\tau} - \varphi ln(l^{\alpha}k^{1-\alpha}) - B + \omega[\delta rk - ph] = 0$. 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 FOCs and $\omega$, we can derive expressions for $\tau$ and $k$:

$$
\begin{aligned}
ln(w - \delta R) = &- \frac{1}{\alpha\varphi}[Ae^{-\theta\tau} - B - \varphi\alpha ln(\beta + \varphi(1 - \alpha e^{-\delta\tau})) \\
&+ \varphi ln(\frac{\varphi}{p}) - \varphi\alpha lnl - \varphi(1-\alpha)ln(\frac{\varphi(1-\alpha)}{r\delta}) - \alpha\varphi]
\end{aligned} \tag{6}
$$

$$k = \frac{\varphi(1-\alpha)(w-\delta R)}{r\delta(\beta+\varphi(1-\alpha e^{-\delta\tau}))} \tag{7}$$

where, notably, the expression for $k$ is an implicit function of $\tau$, the model parameters, and non-choice variables. We therefore solve for $\tau$ computationally in the model estimation below. These expressions 6 and 7 correspond to the conditions used to derive the comparative statics in the text.

Given the optimal choices from Eq. 1, we calculate the indirect utility of a plot owner conditional on their entry to 20k and their choice of plot type $m = (l, B)$, i.e., $U_m(w, R_m; \Theta)$. Because $\tau$ has no closed-form solution, neither does $U_m$, so it too requires solving computationally. The indirect utility of staying in the city permanently ($\tau = \infty$), on the other hand, does have a closed form solution $U_0(w;\Theta) = (\varphi ln\frac{\varphi w}{p(\beta+\varphi)} + \beta ln\frac{\beta w}{\beta+\varphi})/\rho + \frac{A}{\rho+\theta}$ which comes from Eq. 1 after substituting $\tau = \infty$ (thus rendering choices of $k$ and $z_2$ irrelevant).

Continuing with the backward induction. In the penultimate stage, residents choose type of plot that gives them the highest indirect utility; $m^*$ where $U_{m^*}(w, R_{m^*};\Theta) \geq U_m(w, R_m;\Theta)\ \forall\ m \neq m^*$. In the first stage, residents draw a preference shock $\mu$ for living in 20k areas and choose whether to (ever) move there or stay in the city permanently. Residents move to their preferred plot type in 20k areas instead of permanently staying in the city if $U_{m^*}(w, R_{m^*};\Theta) + \tilde{\mu} \geq U_0(w;\Theta)$.

## B.2 Solving for an equilibrium

Here we show that given $w_1$ we can solve for $\tilde{\mu}(w)$, $\{w_m\}_{m=2}^6$ and $\{R_m\}_{m=1}^6$. Recall from the main text that we use the index $m \in \{1,6\}$ to denote the ordering of plot-type attractiveness $\tilde{m} \equiv \phi\alpha ln(l) + B$.[46] Further, we solve the locus piece-wise, so it is helpful to consider $\tilde{\mu}(w) = \bigcup_{m=1}^6 \tilde{\mu}_m(w)$. For now, we take the minimum income among 20k owners, $w_1$, as given, and estimate it as explained below.

We solve the model computationally with the following sequential algorithm. We start by assuming a value $\widehat{\tilde{\mu}_1(w_1)}$ which is the lowest taste shock of the lowest income individual who is allocated a 20k plot. This reflects the preferences of the individual who is just indifferent between living in the city permanently and moving to the least attractive type of plot. Each iteration of the algorithm solves for the highest income individual in 20k areas $\hat{\bar{w}}$. If $\hat{\bar{w}} \neq \bar{w}$ then we update $\widehat{\tilde{\mu}_1(w_1)}$ computationally using a standard root-finding procedure until $\hat{\bar{w}} - \bar{w} = 0$.

Each iteration starts by solving the price of land on the least attractive type of plot $R_1$, then the segment of the locus that defines all individuals who are indifferent between living in the city and the least attractive type of plot $\tilde{\mu}_1(w)$, and then the income level of the marginal individual

[46] In our equilibrium case, the order will be small, medium and large plots with low amenity $B_F$, followed by small, medium and large plots with high amenity $B_N$. This specific ordering is not necessary for an equilibrium solution, and we have solved examples where, for example, higher income people prefer low amenity, large plots $(l_L, B_F)$ to high amenity, small plots $(l_S, B_N)$.

between the least attractive type of plot and the second least $w_2$.

$$
\begin{aligned}
\widehat{R_1} &\rightarrow U_1(w_1, R_1; \Theta) + \widehat{\tilde{\mu}_1(w_1)} = U_0(w_1; \Theta) \\
\widehat{\tilde{\mu}_1(w)} &\rightarrow \widehat{\tilde{\mu}_1(w)} = U_0(w; \Theta) - U_1(w, \widehat{R_1}; \Theta) \\
\widehat{w_2} &\rightarrow N_1 = 800,000 \int_{w_1}^{\widehat{w_2}} \int_{\widehat{\tilde{\mu}_1(w)}}^{\infty} f_w(w) \cdot f_{\tilde{\mu}}(\tilde{\mu})\, d\tilde{\mu}\, dw
\end{aligned}
$$

For the next four types of plots, $m = (2, 3, 4, 5)$, we solve for the price of land $R_m$, the locus segment $\tilde{\mu}_m(w)$, and the income level on the margin with the next more desirable plot type $w_{m+1}$. The solutions can be expressed generically:

$$
\begin{aligned}
\widehat{R_m} &\rightarrow U_m(\widehat{w_m}, \widehat{R_m}; \Theta) = U_{m-1}(\widehat{w_m}, \widehat{R_{m-1}}; \Theta) \\
\widehat{\tilde{\mu}_m(w)} &\rightarrow \widehat{\tilde{\mu}_m(w)} = U_0(w; \Theta) - U_m(w, \widehat{R_m}; \Theta) \\
\widehat{w_{m+1}} &\rightarrow N_m = 800,000 \int_{\widehat{w_m}}^{\widehat{w_{m+1}}} \int_{\widehat{\tilde{\mu}_m(w)}}^{\infty} f_w(w) \cdot f_{\tilde{\mu}}(\tilde{\mu})\, d\tilde{\mu}\, dw
\end{aligned}
$$

For the most desirable plot type $m = 6$ we then solve for $R_6$, $\tilde{\mu}_6(w)$, and $\bar{w}'$:

$$
\begin{aligned}
\widehat{R_6} &\rightarrow U_6(\widehat{w_6}, \widehat{R_6}; \Theta) = U_5(\widehat{w_6}, \widehat{R_5}; \Theta) \\
\widehat{\tilde{\mu}_6(w)} &\rightarrow \widehat{\tilde{\mu}_6(w)} = U_0(w; \Theta) - U_6(w, \widehat{R_6}; \Theta) \\
\widehat{\bar{w}} &\rightarrow N_6 = 800,000 \int_{\widehat{w_6}}^{\widehat{\bar{w}}} \int_{\widehat{\tilde{\mu}_6(w)}}^{\infty} f_w(w) \cdot f_{\tilde{\mu}}(\tilde{\mu})\, d\tilde{\mu}\, dw
\end{aligned}
$$

Therefore, each iteration gives a solution for $\widehat{\bar{w}}$ and we continue to search values of $\widehat{\tilde{\mu}_1(w_1)}$ until $\widehat{\bar{w}} = \bar{w}$ at which point we take all solutions from that iteration for $\tilde{\mu}(w)$, $\{w_m\}_{m=2}^{6}$ and $\{R_m\}_{m=1}^{6}$. This solves the equilibrium prices for each plot type. The allocation of plots to individuals is also easily computed, e.g. $N_1$ plots of type $m = 1$ will be allocated to the individuals $(w, \mu) \in [w_1, w_2] \cap (\tilde{\mu} \geq \tilde{\mu}(w))$.[47]

## B.3 Parameters of the model

First, we discuss the parameters in panel A of Table 8. We assume that the housing consumption share, $\varphi$, and the land share in housing production, $\alpha$, both equal 0.3, which are typical numbers in the literature (e.g., Combes et al. (2021)); and we assume that $z$'s share parameter $\beta$ is 0.7.[48] We set the real interest rate $\delta$ to 0.04, consistent with Kenyan data Henderson et al. (2021), and

[47] If $\bar{w}$ were several orders of magnitude higher, the richest people would not obtain a positive surplus from 20k, since high income diminishes the importance of amenity deterioration in the city relative to its unrestricted choice of housing.

[48] Combes et al. (2021) report estimates around 0.3 for land's share for larger plots in more sparsely populated areas.

as noted above, we assume that $\rho = \delta$. We set the purchase price of capital $r$ to 25 so the rental rate on capital is 1. We set the rental price of a unit of housing in the city to $p = 2.19$ which from the cost function implies that the rental price of a unit of land in the city is 1.8. The suburbs are appropriately much cheaper in our data, ranging in unit price from 0.16 to 0.44, consistent with urban land rent gradients in Kenya (e.g., Henderson et al. (2021)).

The parameter estimates for the city income distribution are in Panel B of Table 8. The estimation of the household's disposable income distribution in the city uses Balboni et al. (2020)'s survey of random households of Dar Es Salaam, collected in 2015, namely the monthly net income response by the household head, converted to 2021 dollars. We approximated the data using log-normal, Pareto, and Gamma distributions, and found that the latter has the best fit. The Gamma distribution has two parameters, shape and scale, the estimates of which are reported in Table 8. Since in principle we are modeling permanent income, we tried a version in which the mean of the Gamma (\$2755) was kept fixed and the Gini coefficient was reduced by a third, in line with literature (Bönke et al. (2015)) . The results were broadly similar, but the loss function was slightly worse, so we kept the current distribution.

We structurally estimate the six parameters in Panel C of Table 8 $(w_1, B_F, B_N, \theta, \mu^{shape}, \mu^{scale})$ using the Simulated Method of Moments (SMM), by minimizing the relative absolute deviation of 12 moments: land prices $R_m$ and occupancy rates $S_m$ for each of the six plot types). The structural estimation chooses the set of parameters that minimizes the following weighted objective function:

$$\sum_{m=1}^{6} \frac{N_m}{N} \left( \frac{|R_m^{Model} - R_m^{Data}|}{R_m^{Model} + R_m^{Data}} + \frac{|S_m^{Model} - S_m^{Data}|}{S_m^{Model} + S_m^{Data}} \right)$$

The global optimization algorithm that we use is Differential Evolution, of the family of Evolutionary Algorithms, which proved very stable.[49]. The parameter search is over (0, 2775); (-0.5,1) ; (-0.5,1); (0,1) ; (0,1); (0,10) for $(w_1, B_F, B_N, \theta, \mu^{shape}, \mu^{scale})$ respectively. The shape and scale of the $\mu$ distribution, $\theta$, $w_1$ are all strictly positive, so we impose a lower bound of zero on all these parameters. The $B$s could theoretically be negative, but in practice even very tiny negative values (e.g. -0.05) cause the loss function to be very large where the model converges, so we picked -0.5 as a lower bound. $A$ is normalized to 1, and we assume that city amenity is higher than 20k amenity at time zero, hence the upper bound for the $B$s. The average wage in the city is 2775, so we use it as an upper bound for $w_1$. $\theta$ is a depreciation rate, so we assume that it is bounded from above by 1. Scaling the upper bounds of the parameters by a large amount (e.g. more than tenfold) led only to minor changes in the estimation results.

Moment matches are reported in Appendix Table A.12 and the matches are generally good. For prices, matches are close, with more deviation for small and medium plots in near areas. For shares

[49] Simulated Annealing, as in Akcigit et al. (2025), was slower and less stable

built, the biggest deviation is for large plots in the far area, but the number of plots involved is a fraction of other pair counts.

In panel C of Table 8, the income cutoff for those who can afford to live in 20k areas is \$579 per year. For 20k area amenities $B_F = 0.029, B_N = 0.146$, compared to the undiscounted $A$ value of 1 in Dar, while the rate of deterioration of $A$, $\theta$, is high at 0.067. At this rate, the city amenity depreciates to the Near 20k amenity in approximately 28 years, and to the Far 20k amenity in 50. The scale and shape parameters of the Gamma distribution for the $\mu$ draws are given.

## B.4 Welfare Effects of Splitting

Here we detail the analysis of the welfare effects of splitting plots. We evaluate splitting a single plot of type $m_1 = (B, l)$ into two plots of type $m_2 = (B, l/2)$. We consider the partial equilibrium effects (PE), general equilibrium effects (GE), and overall total effects (TE), i.e. the sum of PE and GE. What we call the consumer surplus change (CS) is the compensating variation, and the total surplus (TS) is the sum of land values (R) and CS. The change in total surplus from splitting one plot into two half-sized plots has four components:

$$\Delta TS = \Delta R^{PE} + \Delta R^{GE} + \Delta CS^{\Delta R} + \Delta CS^{\Delta Alloc}$$

where $\Delta R^{PE}$ is the PE effect on land values, $\Delta R^{GE}$ is the GE effect on land values, $\Delta CS^{\Delta R}$ is the CS change from the change in prices keeping allocations fixed, and $\Delta CS^{\Delta Alloc}$ is the CS change from the change in allocations at the new prices.

### B.4.1 Land value effects of splitting a plot

For each plot type $m$, denote the land price and supplies before the split as $R_m, N_m$ and after the split as $R'_m, N'_m$. The supply effect of adding one more plot changes the price of every plot type by a small amount, i.e. $R'_m = R_m - \epsilon^R_m$. Further the new supplies are $N'_{m_1} = N_{m_1} - 1$, $N'_{m_2} = N_{m_2} + 2$, and $N'_m = N_m \ \forall \ m \neq m_1, m \neq m_2$. The effect on aggregate land values can be written:

$$\Delta R = \sum_m R'_m N'_m - \sum_m R_m N_m = \underbrace{2R'_{m_2} - 1R'_{m_1}}_{\Delta R^{PE}} + \underbrace{\sum_m (R'_m - R_m) N_m}_{\Delta R^{GE}}$$

The PE effect on land values is approximately the same at either $R'_m$ or $R_m$ because $\epsilon^R_m$ is small:

$$\Delta R^{PE} = 2R'_{m_2} - 1R'_{m_1} = 2R_{m_2} - R_{m_1} + \underbrace{\epsilon^R_{m_1} - 2\epsilon^R_{m_2}}_{\approx 0} \approx 2R_{m_2} - R_{m_1}$$

which is positive if land prices per square meter decline with size, i.e. when $\frac{R_{m_2}}{l/2} > \frac{R_{m_1}}{l}$, and negative otherwise. The GE effect on land values equals:

$$\Delta R^{GE} = \sum_m (R'_m - R_m) N_m = -\sum_m \epsilon^R_m N_m$$

### B.4.2 Consumer surplus effects of splitting a plot

Here we consider the consumer surplus effects. First, to ease notation, we rewrite the indirect utility of plot type $m$ in terms of the lifetime income $W = w/\delta$ rather than period income $w$ and drop the $\Theta$ from the arguments, i.e. $U_m(W, R_m)$. Also, $U_m(W, R_m - a) = U_m(W + a, R_m)$ for any value $a$, which is clear from the household budget constraint. Further, by definition, $\Delta CS^{\Delta Alloc}_{m,m'} > 0$ is the transfer that makes makes the individual indifferent between their previous allocation and their new plot allocation at current prices, i.e. $U_m(W + \Delta CS^{\Delta Alloc}_{m,m'}, R'_m) = U_{m'}(W, R'_{m'})$. Finally, by definition, $\Delta CS^{\Delta R}_m = \epsilon^R_m$ is the transfer that makes makes the individual indifferent between new prices and old prices for their original plot allocation, i.e. $U_m(W + \Delta CS^{\Delta R}_m, R_m) = U_m(W, R'_m)$.

We note four types of potential households following a split that accommodates more people:

1) Households that stay in the city before and after the split. Have no CS change.

2) Households that are in the city before the split and in 20k plot $m'$ after the split. Have a CS *gain* from reallocation, $\Delta CS^{\Delta Alloc}_{0,m'}$, as they could have chosen to remain in the city.

3) Households that are in 20k areas and remain on the same plot $m' = m$. Have a CS change from the change in prices, $\Delta CS^{\Delta R}_m = -(R'_m - R_m) = \epsilon^R_m$, but no reallocation.

4) Households that are in 20k areas in plot $m$ and move to other plot $m' \neq m$. Have a CS change from the change in prices and from reallocation.

For households of type 4),we decompose the CS change into the change from price changes $CS^{\Delta R}_m$ and the change from reallocation $\Delta CS^{\Delta Alloc}_m$. Then, we can decompose the utility change for households 4) as follows:

$$\begin{aligned} U_{m'}(W, R'_{m'}) - U_m(W, R_m) &= U_{m'}(W, R'_{m'}) - U_m(W, R'_m) + U_m(W, R'_m) - U_m(W, R_m) \\ &= \underbrace{U_m(W + \Delta CS^{\Delta Alloc}_{m,m'}, R'_m) - U_m(W, R'_m)}_{>0} + \underbrace{U_m(W + \Delta CS^{\Delta R}_m, R_m) - U_m(W, R_m)}_{>0 \text{ if } \epsilon^R_m > 0 \text{ and } \leq 0 \text{ otherwise}} \end{aligned}$$

So the utility change is composed of two terms. The first is positive because households have chosen plot $m'$ over plot $m$ at the new prices, so these households gain $\Delta CS^{\Delta Alloc}_{m,m'} > 0$. The second terms depends on the GE effect to prices $\Delta CS^{\Delta R}_m = \epsilon^R_m$.

### B.4.3 Total surplus effects of splitting a plot

From the previous section, only households in 2) and 4) face a change in CS from a change in allocation. The total CS change is therefore:

$$\Delta CS^{\Delta Alloc} = \underbrace{\Delta CS^{\Delta Alloc}_{0,m'}}_{>0} + \underbrace{\sum_m \Delta CS^{\Delta Alloc}_{m,m'}}_{>0} > 0$$

From the previous section, only households 3) and 4) face a change in CS from a change in prices. The total CS gain from the change in prices is therefore: $\Delta CS^{\Delta R} = \sum_m \Delta CS^{\Delta R}_m = \sum_m \epsilon^R_m N_m$. Note that this implies $\Delta CS^{\Delta R} = -\Delta R^{GE}$, and therefore total surplus is simplified to:

$$\Delta TS = \Delta R^{PE} + \Delta CS^{\Delta Alloc}$$

To summarize, the total surplus will be the sum of $\Delta CS^{\Delta Alloc} > 0$ and $\Delta R^{PE}$ which is positive for initial prices that decline in plot size.

### B.4.4 Corollary: Relative land prices and surplus from splitting

The above analysis implies that the social planner will choose to split plots past the point of land price per square meter equalization. Define the land prices per square meter as $r_{m_1} \equiv \frac{R_{m1}}{l}$ and $r_{m_2} \equiv \frac{R_{m2}}{l/2}$. As described above $\Delta R^{PE} \approx 2R_{m_2} - R_{m_1} = l(r_{m_2} - r_{m_1})$. As long as prices per square meter are decreasing in plot size there are PE gains to land values. Further, since $\Delta CS^{\Delta Alloc} \geq 0$ the social planner will want to continue splitting plots until $\Delta R^{PE} = -\Delta CS^{\Delta Alloc} \leq 0$. Therefore the land price per square meter on small plots will be *lower* than that on large plots at the planner's optimum. The intuition is that, given a fixed amount of formal land, the social planner is willing to split a plot into two parts of lesser total value if this allows a new entrant to access the community.

There is a caveat. While we found no numerical examples, it might be possible that there is a fifth type of household: those on plot $m$ before the split but who stay permanently in the city after the split. They would have have a CS *loss* from reallocation, $\Delta CS^{\Delta Alloc}_{m,0}$, as their city utility is unchanged, but they now chose to remain in the city. Having such households would imply price increases on some plots after the total increase in plot supply. If a such case exists, such households are likely to have a smaller total decrease in consumer surplus than the total gain from households of type 2), i.e. $\Delta CS^{\Delta Alloc}_{m,0} + \Delta CS^{\Delta Alloc}_{0,m'} > 0$ and thus $\Delta CS^{\Delta Alloc} > 0$ in the above corollary. This occurs because, to equalize demand and supply following a split, the mass of households of type 2) must be exactly 1 + the mass of households of this fifth potential type. So, for small price changes the total gains from type 2) will outweigh the total losses from this type.

## C Data set construction

**Insulae construction.** Insulae (singular insula) are contiguous groups of plots, defined by the planners of the 20k project, which can be thought of as city blocks.[50] Insulae are typically separated by roads, or in some cases by natural spaces that cannot be built on (e.g., streams). Insulae typically contain only residential plots or only non-residential ones, although a few contain a mix of residential and non-residential plots. We often characterize residential insulae by their plot size, which is measured as the mean or median size of residential plots within the insula.

Our RD specifications measure distance to the nearest insula boundary segments. We create insula boundaries that lie at the midpoint between pairs of adjacent insulae (of any type) as follows. First, we define a grid (raster) of 1m x 1m pixels, each classified by the unique insula ID that it spatially overlaps with (pixels that do not overlap any insulae, e.g., roads, are not given an empty value). Second, we expand our set of pixels by iteratively replacing any empty value pixel with the unique ID of its adjacent pixel. This morphological operation, dilation, is common in image processing. We continue this process until no open space remains. In the end, we have well-defined edges where 1m x 1m pixels switch from one ID to another, and we use these edges as our boundaries. Further, this allows us to segment the boundary lines by each unique pair of insulae.

**Super-insulae construction.** "Super-insulae" is a term that we use to group insulae that are similar in size and close to each other. First, residential insula are classified into three size groups where 'small' insulae have a median plot size of less than 800sqm; 'medium' insulae between 800sqm and 1600sqm; and 'large' insulae above 1600sqm. These classifications follow the official planning definitions of high, medium, and low density, where higher density corresponds to smaller plot sizes. We then create super-insula by aggregating spatially "proximate" residential insulae with 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 following a similar method to the insula boundaries. First, we define a grid (raster) of 1m x 1m pixels, each classified as small, medium, large, non-residential, or open space based on the type of insula that they spatially intersect with. Second, we expand our set of small, medium, large, and non-residential pixels by iteratively replacing any open-space pixel by the class of its adjacent pixel. We continue this process until no open space remains. In the end, each set of contiguous pixels becomes a super-insula with a unique ID and particular classification (small, medium, large, or non-residential).

[50]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


Akcigit, U., Pearce, J. and Prato, M. (2025), 'Tapping into talent: Coupling education and innovation policies for economic growth', *Review of Economic Studies* **92**(2), 696–736.

Balboni, C., Bryan, G., Morten, M. and Siddiqi, B. (2020), Transportation, gentrification, and urban mobility: The inequality effects of place-based policies.

Bönke, T., Corneo, G. and Lüthen, H. (2015), 'Lifetime earnings inequality in Germany', *Journal of Labor Economics* **33**(1), 171–208.

Combes, P.-P., Duranton, G. and Gobillon, L. (2021), 'The Production Function for Housing: Evidence from France', *Journal of Political Economy* **129**(10), 2766–2816.

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

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', *Current Urban Studies* **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`.

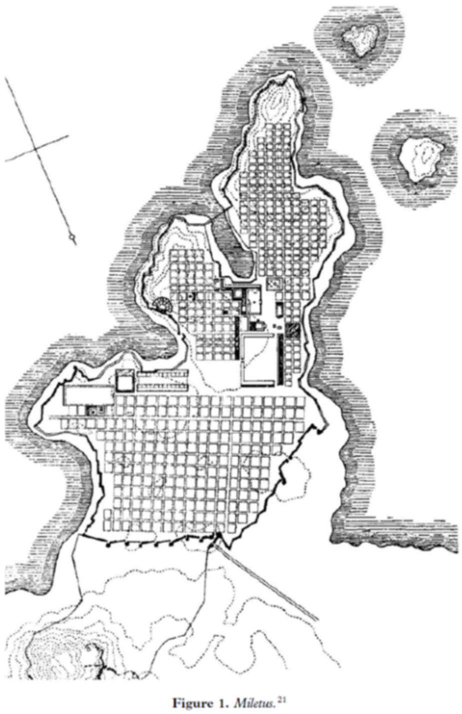
Statista (2022), 'Tanzania: Inflation rate from 1988 to 2027 retrieved on Oct. 9, 2022', `https://www.statista.com`.

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

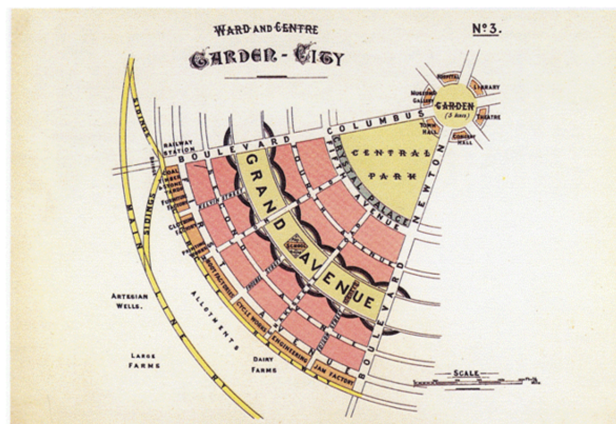
# D Appendix Figures

Figure A.1: Historical examples of urban planning

(A) The urban plan of Miletus, Ancient Greece

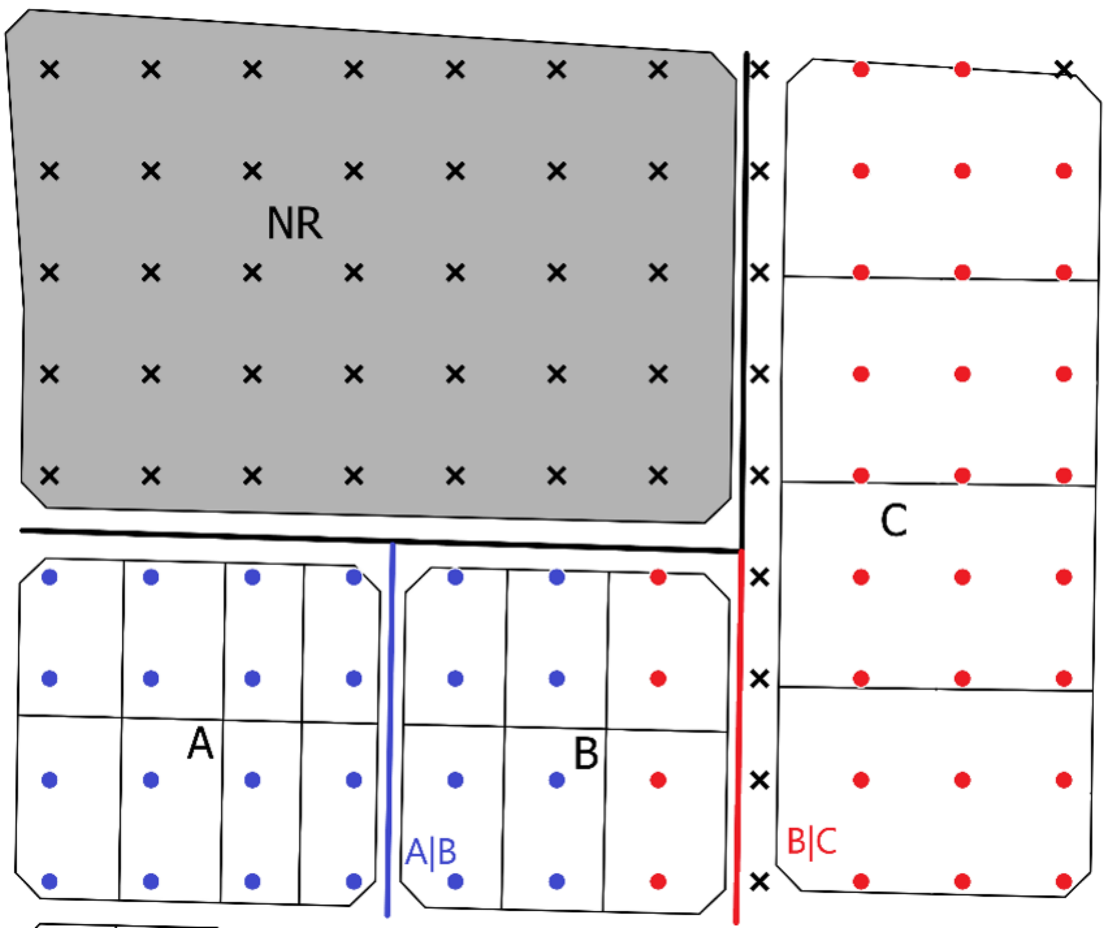



(B) Diagram from Howard (1902) "Garden Cities of To-Morrow"




*Notes:* This figure plots historical examples of urban planning.

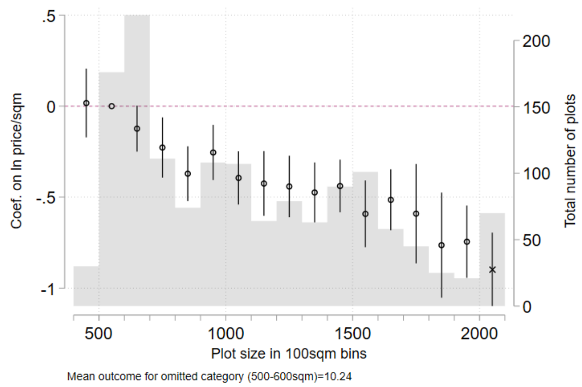
Figure A.2: Diagram of insula, plots, gridcells, and boundaries



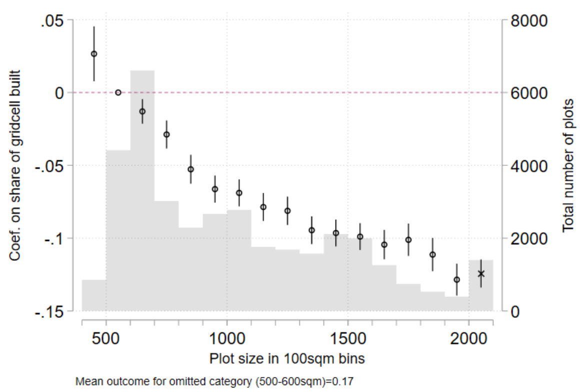

*Notes:* This figure provides a diagram demonstration of our data construction of insula, plots, gridcells, and boundaries. Plots are denoted by black outlines, and are colored white (residential), and grey (non-residential "NR"). Insulae are made up of the contiguous plots, each with a unique ID (A, 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)

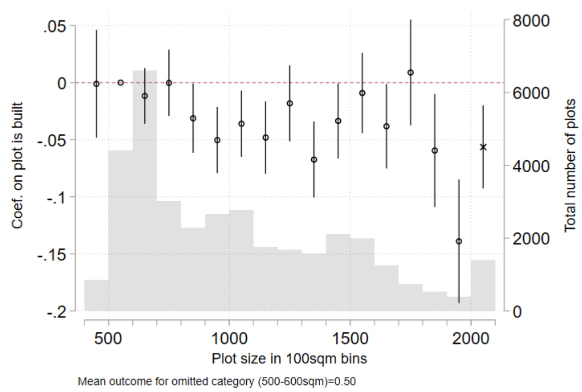
(A) Ln price per sqm



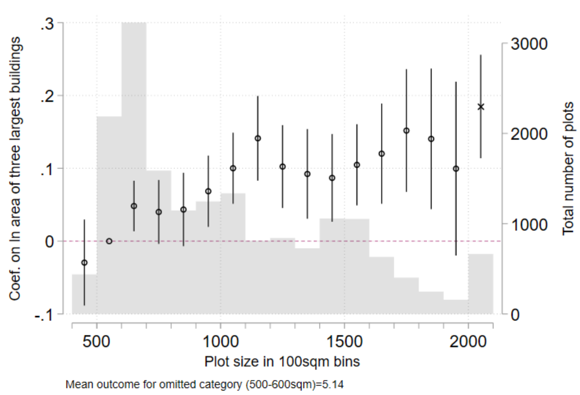

(B) Share of gridcell built



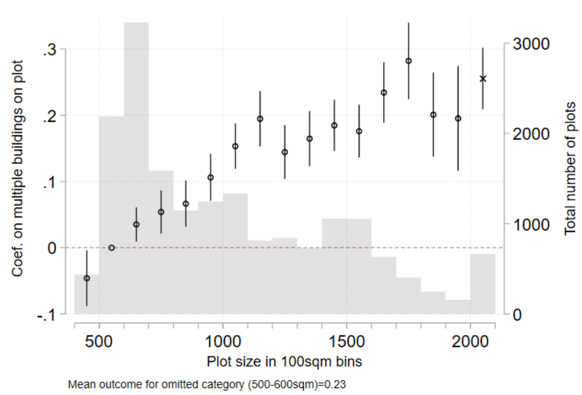

(C) Plot has a building




(D) Ln area of 3 largest buildings



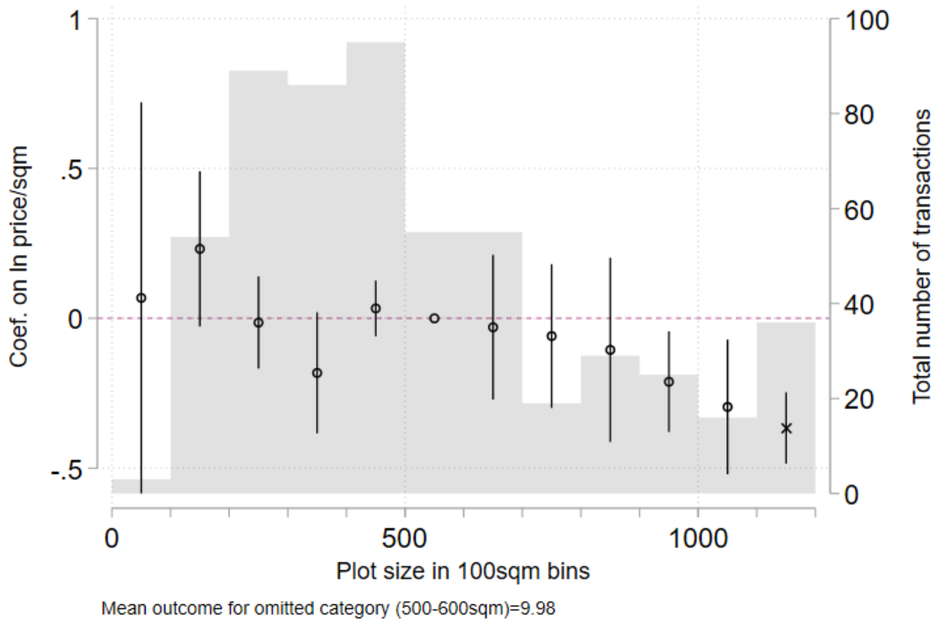

(E) Plot has multiple buildings



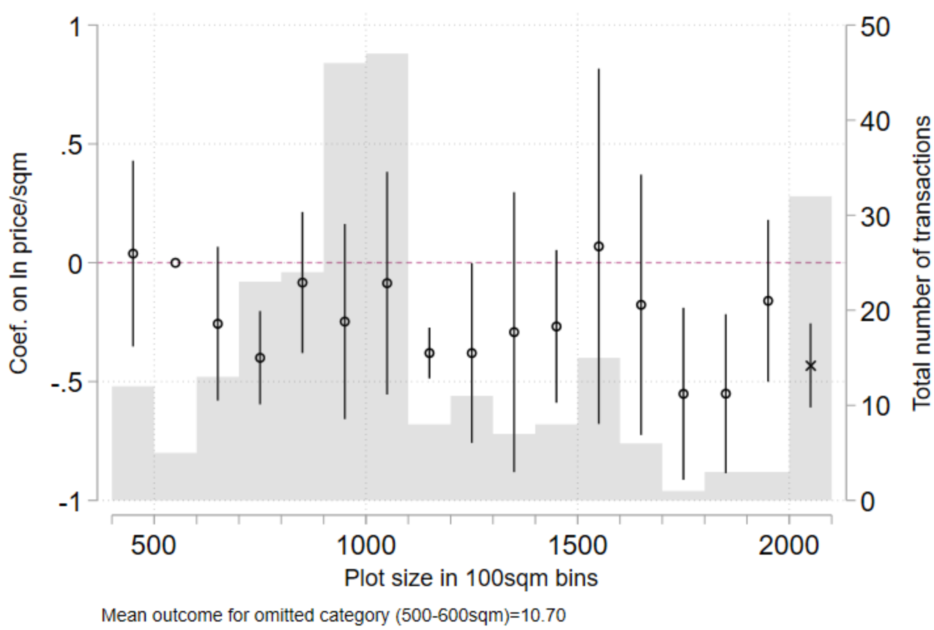

*Notes:* This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter and building quantity measures on 100sqm plot size bins (plots with size above 2000sqm are pooled into one bin, marked by an 'x'). The omitted bin is 500-600sqm. Outcomes vary by panel. In panel (a) the outcome is the log price per square meter. Outcomes in panels (b-e) are the same as those from Table 3. 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)

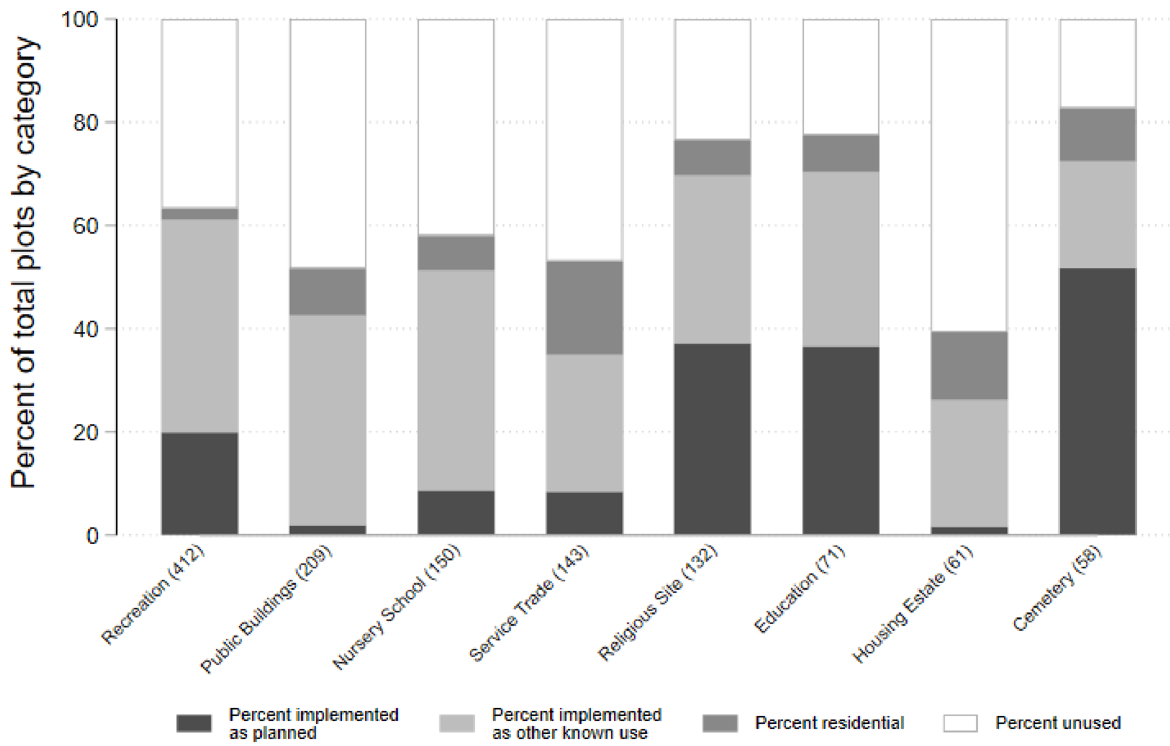



(B) Ln price per sqm (non-20k surveyed)



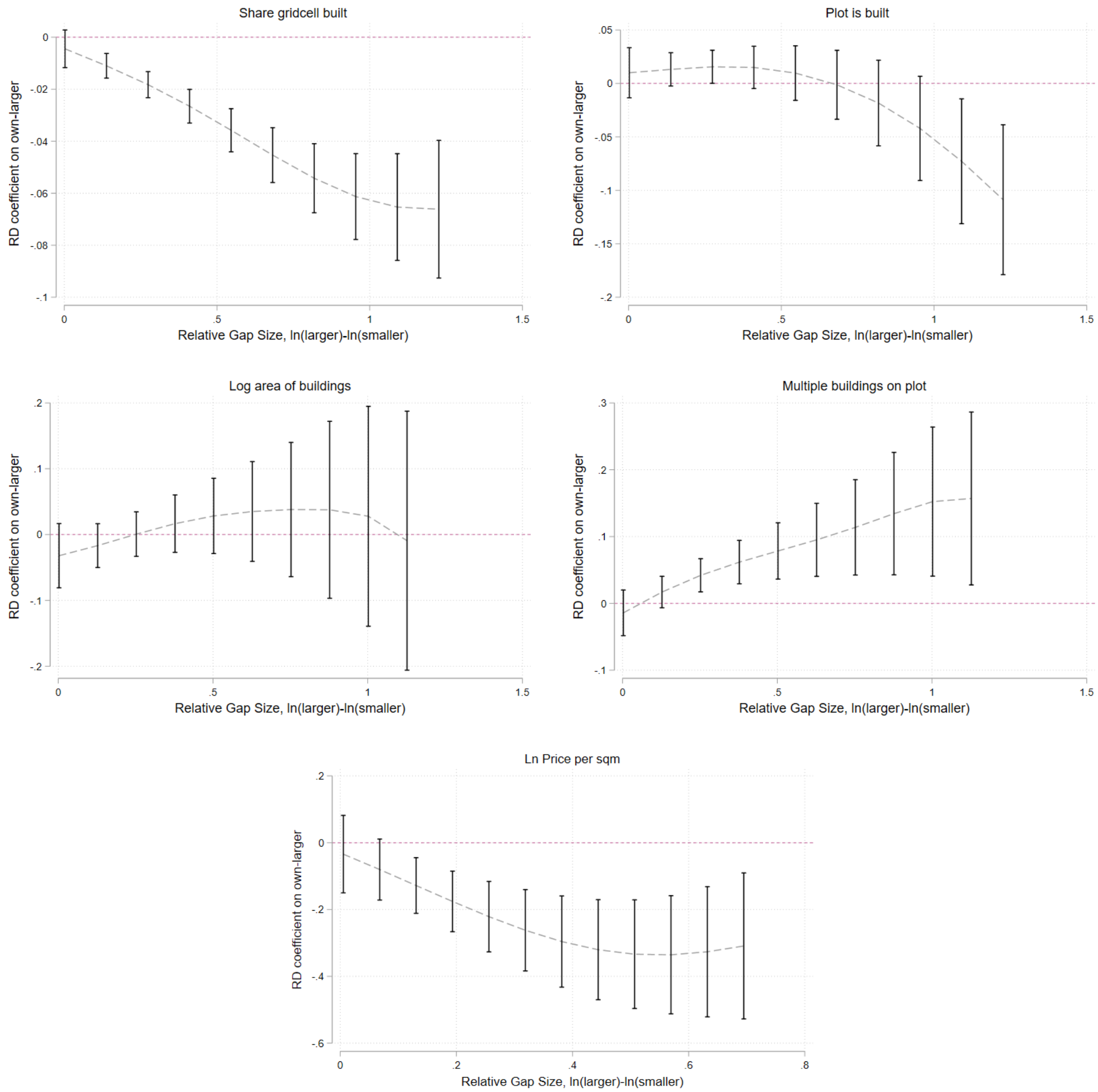

*Notes:* This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter in non-20k areas. The omitted bin is 500-600sqm. Samples vary by panel: unsurveyed plots in non-20k areas (a), and surveyed plots in non-20k areas (b). Controls include fixed effects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

Figure A.5: Planned non-residential broken into implemented uses




*Notes:* This figure shows the breakdown of each planned non-residential use into its current/implemented uses.

Figure A.6: RD coefficients by gap size




*Notes:* This figure plots estimated coefficients on the own plot larger dummy following the specification outlined in Eq. 5. We use the `vc_pack` stata package (Rios-Avila, 2020) to implement a semi-parametric approach called smooth varying-coefficient model to estimate heterogeneity in the coefficient of interest along a particular dimension of interest. Here we look at heterogeneity by the relative plot size gap for a particular boundary segment, i.e. ln(mean size of plots in larger insula) - ln(mean size of plots in smaller insula). For computational purposes we first estimate residuals of each outcome and the own plot larger dummy (using all controls from equation 5 except the own plot larger dummy), and then run the `vc_pack` semi-parametric approach.

# 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 stream. |
| Water/wetland 100m | Gridcell is within 100m of water feature or a 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 planned recreation plot. |
| Pln. nursery school in 100m | Gridcell is within 100m of planned nursery school plot. |
| Pln. religious site in 100m | Gridcell is within 100m of planned religious (e.g. church, mosque, etc.) plot. |
| Pln. education in 100m | Gridcell is within 100m of other education (e.g. primary, secondary, etc.) plot. |
| Pln. service trade in 100m | Gridcell is within 100m of planned service or trade plot. |
| Pln. housing estate in 100m | Gridcell is within 100m of planned public housing plot. |
| Pln. public building in 100m | Gridcell is within 100m of planned public building plot. |
| Pln. cemetery in 100m | Gridcell is within 100m of planned cemetary plot. |
| Pln. any other non-res in 100m | Gridcell is within 100m of planned any other type of non-residential use plot. |
| 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 non-residential 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 | 6.34 |
| Elevation (m) | 48.96 | 28.07 | 2.50 | 111.00 | 56.69 | 27.21 | 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 | -3.47 | 1.12 | -0.00 | 0.82 | -2.64 | 1.11 |
| Z1: Rectangularity | -0.00 | 1.00 | -3.58 | 1.08 | -0.01 | 1.02 | -3.41 | 1.07 |
| Z2: Insula alignment | 0.00 | 1.00 | -2.78 | 0.89 | 0.03 | 0.98 | -2.78 | 0.89 |
| Z3: Regularity | 0.00 | 1.00 | -7.82 | 1.42 | -0.02 | 1.06 | -6.65 | 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 | 0.23 | 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 | 0.31 | 0.46 | 0.00 | 1.00 | 0.32 | 0.47 | 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. nursery school in 100m | 0.02 | 0.13 | 0.00 | 1.00 | 0.01 | 0.12 | 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. service trade in 100m | 0.03 | 0.18 | 0.00 | 1.00 | 0.04 | 0.19 | 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. public building in 100m | 0.01 | 0.09 | 0.00 | 1.00 | 0.01 | 0.08 | 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. other non-res in 100m | 0.02 | 0.14 | 0.00 | 1.00 | 0.02 | 0.15 | 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. as residential in 100m | 0.07 | 0.25 | 0.00 | 1.00 | 0.06 | 0.24 | 0.00 | 1.00 |
| 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. unknown non-res in 100m | 0.00 | 0.03 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table A.3: 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 | 1.91 |
| Kisota | 1120 | 8.34 | 1.88 |
| Mbweni JKT | 1920 | 8.88 | 2.60 |
| Mbweni Mpiji | 1632 | 8.72 | 2.07 |
| Mivumoni | 1344 | 8.52 | 2.14 |
| 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 A.4: Years of schooling, plot size and date of occupancy

| Year Occupied | Years of Schooling Small | Medium | Large |
|---|---|---|---|
| <2011 | 11.3 | 13.1 | 12.4 |
| 2011-2015 | 12.6 | 14.8 | 14.7 |
| >2015 | 13.2 | 14.9 | 14.7 |
| N | 403 | 764 | 249 |

*Note:* This table presents the mean number of years of schooling for the head of landowner households. Means are broken down by the date in which the plot was first occupied (rows), and the size of the plot (columns). The plot size categories are based on the official definitions (below 800sqm, between 800sqm and 1600sqm, and above 1600sqm). Only 20k areas that are near to the city center are included, so Buyuni and Mwongozo are not included.

Table A.5: 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.6: Balance on natural amenities (RD)

| | (1) Elevat. | (2) Rugged. | (3) River or stream in 100m | (4) Wetland in 100m | (5) Z-index cols. 1-4 |
|---|---|---|---|---|---|
| **Panel A: all insula pairs** | | | | | |
| Own Larger | 0.037 | -0.030 | -0.0021 | 0.0015 | 0.0032 |
| | (0.054) | (0.016) | (0.0021) | (0.0012) | (0.0086) |
| Mean Outcome | 49 | 0.37 | 0.024 | 0.0018 | 0.0032 |
| N (gridcells) | 93580 | 93580 | 93580 | 93580 | 93580 |
| N (plots) | 36035 | 36035 | 36035 | 36035 | 36035 |
| **Panel B: gap≥400sqm** | | | | | |
| Own Larger | -0.11 | -0.053 | 0.0069 | 0.00014 | 0.00070 |
| | (0.12) | (0.039) | (0.0058) | (0.00063) | (0.015) |
| Mean Outcome | 50 | 0.47 | 0.032 | 0.0021 | -0.0327 |
| N (gridcells) | 23872 | 23872 | 23872 | 23872 | 23872 |
| N (plots) | 9661 | 9661 | 9661 | 9661 | 9661 |
| **Panel C: gap<100sqm** | | | | | |
| Own Larger | 0.024 | -0.0026 | -0.0057 | 0.0012 | 0.0036 |
| | (0.074) | (0.019) | (0.0018) | (0.00093) | (0.0081) |
| Mean Outcome | 50 | 0.31 | 0.017 | 0.0015 | 0.035 |
| N (gridcells) | 32780 | 32780 | 32780 | 32780 | 32780 |
| N (plots) | 16448 | 16448 | 16448 | 16448 | 16448 |

*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: eleveation in metres, ruggedness, an indicator for a river or stream within 100m, and an indicator for 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 ther three (disamenities) are negative. Controls always include 20K*Mtaa Area and insula-segment FEs. Standard errors in parentheses are clustered by insula.

Table A.7: Optimal bandwidth selection (RD)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ln Price per sqm | Share gridcell built | Plot is built | Log area of buildings | Multiple buildings on plot |
| Own Larger | -0.22 | -0.014 | 0.0033 | 0.013 | 0.039 |
| | (0.055) | (0.0059) | (0.0078) | (0.016) | (0.013) |
| Bandwidth (m) | 29 | 21 | 48 | 40 | 33 |
| Mean Outcome | 10 | 0.091 | 0.49 | 5.3 | 0.37 |
| N (gridcells) | 1651 | 32532 | 79441 | 33588 | 27683 |
| N (plots) | 856 | 23303 | 33683 | 15697 | 14810 |

*Note:* This table presents optimal bandwidth selection results for RD regressions across neighbouring insula boundaries. First, an optimal bandwidth for each column is selected by run the `rdbwselect` in Stata for the outcome against the running variable (distance to the insula-pair boundary) following the approach in (Calonico et al., 2014). The selected bandwith in meters is given at the bottom of the table. Second, we run the same specifications as Table 4 Panel A, but restrict to observations within the selected bandwidth rather than within 100m. Standard errors in parentheses are clustered by insula.

Table A.8: 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.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 | 0.47 | -0.083 | -0.015 | 0.12 | 0.19 |
| | (0.059) | (0.0026) | (0.0096) | (0.018) | (0.012) |
| Insula rectangularity (standardized) | 0.014 | 0.0058 | 0.024 | 0.014 | 0.00057 |
| | (0.029) | (0.0014) | (0.0058) | (0.011) | (0.0068) |
| Insula alignment (standardized) | -0.011 | 0.0018 | 0.0052 | 0.0072 | 0.0088 |
| | (0.023) | (0.0011) | (0.0044) | (0.0080) | (0.0054) |
| Insula homogeneity (standardized) | 0.026 | -0.0027 | -0.0068 | -0.011 | -0.0018 |
| | (0.027) | (0.0012) | (0.0049) | (0.0096) | (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 3 notes. Controls always include 20K*Mtaa FEs and amenities. Amenities are the same as described in Table 2, 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 perfectly 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 | (2) Perfect | (3) Random Implementation | (4) | (5) | (6) N Plots | (7) |
|---|---|---|---|---|---|---|---|
| | Ratio | Ratio | Median | 95-pct | 99-pct | Plan | 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 | 1.5 | 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 | 1.7 | . | . |
| 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: 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 | 0.0014 | 0.14 | 0.19 |
| | (0.0025) | (0.0093) | (0.018) | (0.012) |
| Dist (km) paved major road | -0.015 | -0.041 | -0.062 | -0.039 |
| | (0.0016) | (0.0071) | (0.012) | (0.0087) |
| Elevation (m) | 0.00089 | 0.0028 | 0.0032 | 0.00033 |
| | (0.000099) | (0.00043) | (0.00067) | (0.00048) |
| Ruggedness | -0.0055 | -0.015 | -0.0097 | -0.0089 |
| | (0.00098) | (0.0039) | (0.0090) | (0.0053) |
| River/stream 100m | -0.028 | -0.12 | -0.067 | -0.046 |
| | (0.0052) | (0.023) | (0.058) | (0.048) |
| Water/wetland 100m | 0.0054 | -0.075 | -0.081 | 0.053 |
| | (0.0084) | (0.031) | (0.16) | (0.23) |
| Z-index: 3 Ins. Characteristics | 0.0029 | 0.016 | 0.0051 | 0.0056 |
| | (0.0013) | (0.0057) | (0.010) | (0.0068) |
| 20k edge in 100m | -0.0033 | -0.0078 | -0.032 | 0.0096 |
| | (0.0023) | (0.0095) | (0.016) | (0.011) |
| Impl. recreation in 100m | 0.0064 | 0.0026 | 0.034 | 0.0043 |
| | (0.0031) | (0.011) | (0.019) | (0.013) |
| Impl. nursery school in 100m | 0.0074 | 0.037 | 0.000049 | 0.018 |
| | (0.0072) | (0.027) | (0.039) | (0.026) |
| Impl. religious site in 100m | 0.011 | 0.028 | 0.012 | 0.013 |
| | (0.0043) | (0.016) | (0.027) | (0.018) |
| Impl. education in 100m | 0.00012 | 0.0018 | -0.027 | -0.0015 |
| | (0.0044) | (0.015) | (0.027) | (0.017) |
| Impl. service trade in 100m | 0.016 | 0.050 | 0.053 | 0.058 |
| | (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 | -0.0061 | -0.034 | -0.013 | 0.083 |
| | (0.010) | (0.040) | (0.060) | (0.045) |
| Impl. cemetery in 100m | 0.0025 | 0.021 | -0.021 | 0.031 |
| | (0.0054) | (0.018) | (0.034) | (0.024) |
| Impl. other non-res in 100m | -0.0050 | -0.032 | 0.0057 | 0.0089 |
| | (0.0076) | (0.025) | (0.042) | (0.025) |
| Impl. farming in 100m | -0.00012 | 0.0069 | -0.011 | -0.0035 |
| | (0.0023) | (0.0087) | (0.015) | (0.010) |
| Impl. as residential in 100m | -0.0033 | -0.0016 | -0.013 | -0.0051 |
| | (0.0034) | (0.014) | (0.023) | (0.017) |
| Unused, kept in 100m | 0.0030 | 0.00053 | -0.021 | -0.0021 |
| | (0.0026) | (0.010) | (0.016) | (0.012) |
| Unused, unkept in 100m | -0.0085 | -0.035 | -0.042 | -0.023 |
| | (0.0027) | (0.010) | (0.017) | (0.012) |
| Impl. unknown non-res in 100m | -0.0016 | 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 3 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, alignment, and homogeneity), 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 A.12: Moments

| Label | # Plots | Land Price | | | Share Built | | |
|---|---|---|---|---|---|---|---|
| | | Moment | Data | Model | Moment | Data | Model |
| Far, Small | 4975 | $R_1$ | 2388 | 2388 | $\%(\tau_1 < 15)$ | 30.45 | 30.45 |
| Far, Medium | 4192 | $R_2$ | 3272 | 3532 | $\%(\tau_2 < 15)$ | 30.75 | 30.75 |
| Far, Large | 938 | $R_3$ | 4571 | 4741 | $\%(\tau_3 < 15)$ | 27.50 | 100 |
| Near, Small | 10068 | $R_4$ | 6776 | 4983 | $\%(\tau_4 < 15)$ | 59.42 | 59.46 |
| Near, Medium | 12661 | $R_5$ | 10906 | 8001 | $\%(\tau_5 < 15)$ | 54.74 | 71.28 |
| Near, Large | 3381 | $R_6$ | 12478 | 12685 | $\%(\tau_6 < 15)$ | 54.63 | 54.63 |

*Notes:* This table presents the moments for both the data and the model. The first three columns under "Land Price" provide the parameter, data, and model values for land prices, while the last three columns under "Share Built" show the percentage built in 2020. The data for Land Prices comes from a regression of log price on dummies for each of the six size-location categories and fixed effects for transaction date and data source. The estimated equation with the omitted category as sold in 2021 & dalali transaction is $\ln \text{price}_{t,n} = 16.81(0.05) + 0.39(0.05)\text{small} \times \text{near} + 0.87(0.04)\text{medium} \times \text{near} + 1.00(0.04)\text{large} \times \text{near} + -0.65(0.05)\text{small} \times \text{far} + -0.33(0.04)\text{medium} \times \text{far} + 0\text{large} \times \text{far}$. For each of the six categories we plug in the dummies to the equaiton above, take the anti-log, convert to USD and deflate from 2021 back to 2005.