Greenfield Urban Planning*

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

August 10, 2026

Abstract

Urban planning shapes cities, demarcating property rights and mitigating coordination failures, but its rigidities constrain market-driven development. In developing country cities, characterized by rapid growth and urban informality, greenfield urban planning, if done well, creates value. We develop a framework for analyzing how planning choices affect land and individual-level outcomes and, thus, the magnitude and distribution of this value. We study Tanzania's flagship "20,000 Plots" project in Dar es Salaam, which partitioned peri-urban land into over 36,000 formal residential plots. Using administrative records, satellite imagery, and primary surveys, we estimate causal effects of planning decisions using highly local within-neighborhood variation and spatial regression discontinuities, and we develop and estimate a dynamic structural model. We find that the price per square meter falls with plot size (elasticity of -0.5) within project neighborhoods. The model yields a sufficient statistic for planners: inefficient allocations have decreasing price per square meter with plot size. Model counterfactuals indicate that welfare maximization would increase the relative supply of small plots and spread gains more evenly, though still attracting mostly higher-income households. Land value maximization would, in contrast, have many fewer plots overall, focused even more on higher-income households. Extending the current project would generate large net gains, reflecting model shadow values well above cost. Land values are roughly twice those in nearby informal areas conditional on plot size, mostly due to secure property rights and road access.

KEYWORDS: Urban Planning, Economic Development, Africa.

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

* An earlier version of this paper circulated as "Economics of Greenfield Urban Planning" (Henderson et al., 2025). We thank our field managers, Prosper Kaigarula and Erick Makori, 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, Daniel 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

The roles of planners and markets in determining land allocation have been debated throughout history. Economists have long been critical of both market intervention broadly, e.g. Smith (1759) on the "man of system" organizing lives, and strict urban planning specifically, e.g. Jacobs (1962) on the top-down urban renewal projects of Le Corbusier and Robert Moses. Modern research remains generally critical of land-use regulation (Gyourko and Molloy, 2015). Yet regulation is pervasive: in high-income countries, planning governs the roughly 50% of urban land in public use and, through zoning, the private remainder.[1] Duranton (2017) frames the challenge as balancing market-based development against planning's role in setting the "rules of the game" (e.g., property rights) and providing public goods, yet mutual learning between the two fields has been limited (Bertaud, 2018). The stakes are high as cities house most of the world's population, and new urbanization is concentrated in developing countries (Glaeser, 2012; Moretti, 2012), where—particularly in Africa's large and expanding cities—planning is often ineffective and informal settlements predominate. In that context, prior work has established that de-novo planning raises land values relative to unplanned areas (Michaels et al., 2021). Greenfield planning thus creates value (a rent), but its size and distribution depend on how planning is done, a topic that has received little scrutiny from economists.

This leaves open a crucial, policy-relevant question: *which planning choices matter, and how could planning be improved?* Planners face many choices, including where to site projects, how much land to develop, and how to subdivide it into plots of different sizes, uses, and configurations. Yet economic evaluation of these decisions is limited. We focus on how to diagnose and correct within-project misallocation, and then extend the logic to examine two other margins: scale and location.

We combine novel data, reduced-form evidence, theory, and counterfactual simulations to study a large greenfield ('de-novo') urban planning project. First, we assemble a rich and highly spatially disaggregated dataset, including: detailed project maps of residential and non-residential plots, roads, and planned amenities; primary surveys with local leaders, informal real estate agents, and residents; and high-resolution satellite imagery. Second, we estimate causal effects of planning treatments on land outcomes using OLS and spatial regression discontinuity (RD) designs, focusing on variation within very local neighborhoods. Third, guided by these results, we develop a model of planners' plot-size and amenity decisions that captures the sorting of owners across plots by income, the timing and intensity of development, land values, and welfare. We estimate a structural version of our model by matching data moments. We then use the estimated model to simulate the optimal plot size distribution under two alternative objectives: welfare maximization (the social planner's

[1] In developed countries, roads alone take up 20-30% of urban land (American Planning Association, 1950), but in developing countries this figure is typically lower (e.g., Bertaud (2018), Figure 5.11).

criterion) and land value maximization (the land monopolist's criterion).

We study the "20,000 Plots" project (hereafter, 20k), a flagship Tanzanian initiative implemented in the early 2000s on the fringes of Dar es Salaam, one of Africa's fastest-growing large cities. The project surveyed, titled, and partitioned undeveloped agricultural land into serviced plots that people could buy and build on, a strategy pursued in many developing countries under the World Bank's 'Sites and Services' agenda and later by several African governments. It delivered more than 36,000 residential plots, almost double what was originally intended.

This greenfield context is well-suited for identifying causal effects of planning decisions because it is free from confounding pre-existing development. The units in our empirical analysis are small land parcels (the natural units with which planners work). To address concerns about the potential endogeneity of planning decisions, we employ two complementary strategies. First, we estimate OLS regressions with small area fixed effects, which greatly restrict preexisting locational differences. Second, we estimate spatial regression discontinuity (RD) designs, which compare land parcels allocated to different planning treatments on opposite sides of roads, holding natural amenities and local conditions fixed. Reassuringly, both approaches yield quantitatively similar estimates, supporting the credibility of our methodology.

A key interpretive point concerns what our parcel-level estimates capture. Consider a planner reallocating a tiny piece of land from a small plot to a neighboring large plot, which has trivial general equilibrium effects on the prices of plots and does not affect equilibrium sorting. However, it affects parcel-level outcomes through two channels: (i) a direct effect of plot size, holding owner income constant, and (ii) an owner income effect, since under the existing sorting pattern, the large plot may be occupied by a higher-income owner who invests differently. This parallels boundary-based designs where population composition changes at the threshold (Abdulkadiroğlu et al., 2014). An estimated effect of marginal changes in plot size on outcomes reflects the total effect (sum of both channels). To move beyond this counterfactual, our model allows us to assess the effects of infra-marginal planning decisions on parcel-level outcomes, sorting, and aggregate welfare. The model also decomposes the total effect into its direct and owner-income components.

We find planning creates a substantial rent: conditional on plot size, bare land in the project is worth roughly double that in nearby unplanned areas. This premium appears to arise from two factors. The first is stronger property rights, evidenced by the premium that formalized informal plots command over unformalized ones. The second is access: a planned layout gives all residents side streets connected to the main paved roads to the city center, proximity to which is itself highly valued. Surveys of local leaders corroborate both aspects of access.

Our main findings are based on variation in planning choices within planned areas. We find the elasticity of price per square meter with respect to plot size is approximately -0.5. Strikingly, we find that in nearby unplanned areas, where plot sizes emerge from an unregulated process, this elasticity is much lower and not significantly different from zero. This contrast suggests that the rigidity of formal plot sizes, which cannot easily be adjusted ex post, may generate a substantial price wedge across plot sizes. We also find that larger plots have much less intensive development, though are no less likely to be built upon and tend to contain more buildings. Second, we find that better access to main paved roads connecting to the city center is highly valued.

Planning choices that might have been expected to matter are less consequential for building outcomes or land values. A more gridded layout, higher elevation, and distance from flood-prone areas have modest effects on construction, while planned non-residential amenities are largely ignored due to low implementation rates. We also find modest evidence of plot-size externalities: smaller plots bunched together with other small plots are only slightly more likely to be built on.

Building on our empirical evidence, we develop a model in which households who differ in income and taste for planned areas choose whether to remain in the city, where housing choices are largely unrestricted, but amenities deteriorate, or move to planned areas, where the plot size distribution is fixed. Households who choose planned areas purchase a plot with a size and an amenity and decide when to move there and build housing. The model focuses on plot sizes and connectivity to the city center, the two planning choices that our empirical results identify as most consequential.

We match moments of the model to the data and conduct a counterfactual analysis to assess how planning could have been improved. Motivated by the negative elasticity of price per square meter with respect to plot size, which in our model reflects misallocation, we assess the effects of changing the initial distribution of formal plot sizes on sorting, prices, housing investment, development rates, and welfare. Welfare maximization entails providing more small plots to reduce the price-size elasticity, accommodate more households, and extend formal ownership to lower-income people, raising consumer surplus (equivalent variation) by 11.1% and total surplus (equivalent variation and land value) by 2.3%. Land value maximization moves in the opposite direction, providing even fewer and larger plots, which reduces welfare.

Reallocation of land has consequential distributional effects: it changes who gains with only a modest effect on total surplus. Welfare maximization distributes the gains far more evenly than the baseline or land-value maximization, though the poorest remain largely excluded. The larger consumer-surplus gain under welfare maximization comes with a fall in aggregate land value; the model treats this as a transfer from landowners to residents. We focus on the incidence across income groups, rather than on the split between land value and consumer surplus.

These counterfactuals hold the project's total formal land fixed; relaxing that reveals a further lever. At the welfare optimum, the shadow value of formal residential land is about \$6.83 per square meter on sites nearer the city center and \$2.27 farther out, against roughly \$1 to acquire and service land. This value remains above cost even if the project's land were doubled. So, too little de-novo land was provided, especially in less peripheral areas. Well-connected greenfield land can be costlier to acquire than remote land, but the large gap in gains suggests that planners should prioritize it where possible. Because such choices, like the plot-size distribution, cannot be undone ex post without violating property rights, they must be made at the planning stage.

The model also offers a sufficient statistic for evaluating greenfield urban planning efficiency: land price per square meter should be weakly increasing in plot size. When prices fall, as in our setting, splitting larger plots into smaller ones increases welfare. Research in developed-country contexts consistently finds that prices fall with plot size (Brownstone and De Vany, 1991; Colwell and Sirmans, 1993; Combes et al., 2021; Larson and Shui, 2022), consistent with the diagnostic where re-division is constrained by a high "regulatory tax" (Gyourko and Krimmel, 2021). The sufficient statistic allows planners to spot inefficient plot size distributions that may arise from the application of rules of thumb of planning, underscoring how planning and markets can be combined to improve outcomes.

Our paper relates to debates on the roles of planners and markets in determining the allocation of land, as noted earlier [2] We extend the literature on land use regulation and zoning in cities in high- and low-income countries (Glaeser and Ward, 2009; Turner et al., 2014; Chiumenti et al., 2022; Anagol et al., 2021; Nagpal and Gandhi, 2024) by studying explicit planning decisions like plot and road configurations, in a greenfield context. The de-novo setting is important, as previous research has demonstrated that zoning itself is influenced by prior development (Shertzer et al., 2018).[3] For the literature on suburban areas in high-income countries cited above, which finds 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; Harari and Wong, 2025), which we add to by studying the effects of residential planning regulations carried over from colonial rule, which emphasized large plot sizes. Related are descriptive literatures on the prevalence of large minimum plot sizes in Africa (e.g., Gulyani and Connors, 2002; Collier and Venables, 2014; Tipple, 2015) and on de-novo planning in Tanzania (Tiba et al., 2005; Mwiga, 2011; Kironde, 2015), to which we contribute by studying planning in a quantitative economic framework. The literature emphasizes 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 planning decisions per se from those due to property rights.

We also contribute methodologically to a literature that uses land market prices as data. In

[2] Urban economics has long 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).

[3] A related literature studies the costs of overly segmented plots near large city centers (Harari and Wong, 2026; Yamasaki et al., 2023), but we focus on a suburban setting, where greenfield land is abundant.

particular, since land market transactions are frequently off-the-record in developing countries, most studies rely on government assessments (Harari and Wong, 2026; Gechter and Tsivanidis, 2023), self-reports (Gonzalez-Navarro and Quintana-Domeque, 2009), or listings data with partial coverage (Henderson et al., 2021). We survey informal real estate agents who facilitate transactions and rentals within a locale as an alternative method of collecting price data. These agents each report on dozens of transactions and match these at the property level, providing scale in data collection and direct measures of transaction prices.

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; 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 the paper proceeds as follows. Section 2 discusses institutional background. Section 3 describes the data. Section 4 presents the empirical framework and reduced-form results. Section 5 presents and estimates the model and the counterfactuals. Section 6 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 (panel A of Figure A.1) and Piraeus (Athens's port) with gridded layouts (Paden, 2001). 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 Urban planning, segregation, and colonial legacy

Urban planning has also long served to facilitate segregation. In the US, exclusionary land-use regulation has operated both explicitly (Sood and Ehrman-Solberg, 2026) and implicitly through restrictions on residential density (Ihlanfeldt, 2004; Song, 2025). Colonial administrations from

Mexico to Indonesia explicitly separated European from indigenous quarters, each under distinct land institutions (Baldomero-Quintana et al., 2025; Harari and Wong, 2025), with standards that often persist today, restricting formal development and raising its cost (Visagie and Turok, 2020).

Kironde (1994) and MLHHSD (2018) provide an overview of urban planning history 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). During this time, de-jure planning standards were retained, sometimes with new justifications, and a series of masterplans were developed, in which colonial-era plot size categories were relabeled as density classes but retained similar standards (Kironde, 2006). The causes of institutional inertia may have included the insistence of officials on retaining high standards, perhaps to preserve bureaucratic authority (Kironde, 2006), possibly reinforced by political economy pressures. In contrast to these high (and rigid) planning standards, however, almost the whole city comprises informal settlements.

2.3 De-novo planning in Tanzania and elsewhere

From the 1970s, some de-novo planned neighborhoods were developed in Tanzania, notably through "Sites and Services" in collaboration between the Tanzanian government and the World Bank. Similar projects were financed by the World Bank in 53 countries around the same time (Kamunyori et al., 2022). 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 at least fourteen African governments and renewed interest from the World Bank (Centre for Affordable Housing Finance, 2024; Kamunyori et al., 2022).

2.4 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 5.5 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.[4] In recent years, projects in Tanzania have been initiated to create formal plots in about 40 additional cities.

The 20k project was implemented around 2000-2005 and delivered over 36,000 residential plots in twelve project neighborhoods spanning a total area of 75 sqkm. The residential plots, which take up roughly half the area (~38 sqkm), were formally surveyed and titled. Approximately 1,550 additional plots (~12 sqkm) 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.

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 price per square meter of the 20k residential plots sold by the government was lower than in the earlier Sites and Services project (Michaels et al., 2021). Nevertheless, the poor were largely underserved by 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

[4] 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.

[5] We include in our analysis one area, Mbweni 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 A.3.

[7] Prospective buyers had to collect application forms from municipalities or the Ministry of Lands, fill them in, and submit them to the municipal land office. Priority was given to applicants who: (i) had owned land in this specific area; (ii) could pay for the 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.

with richer buyers. Finally, the plots' large sizes 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 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). 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.

[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).

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 an 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-sized 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. Details on insulae and super-insulae construction are given in Appendix B.

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 and outcomes of interest

The main units for our empirical analysis are small square land parcels ("gridcells"), which planners can "treat", e.g., by assigning them to residential plots of different sizes or varying their proximity to planned amenities. 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. Appendix Figure A.2 illustrates gridcell assignment to plots and to boundaries of insula. We focus on the $\sim 95,000$ 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 4 are intuitive and also relate to the model in Section 5,

[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 several adjacent insulae.

which has price and quantity aspects. We begin with 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]

Most other outcomes cover the universe of 36,215 plots with a grid square overlay and are mostly derived from high-resolution satellite imagery. For these, we measure 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. Other outcomes measure the intensity of investment using two intensive margin measures: (i) the logarithm of the total footprint of (up to) three largest buildings on the gridcell's plot; and (ii) an indicator for multiple buildings in the gridcell's plot. Another key outcome combines the extensive margin of whether a plot is built upon by 2020 and the intensive margin of how much is invested. We look at construction intensity overall, as the share of each gridcell that is built in 2020.

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, is 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 An empirical analysis of the consequences of planning decisions

4.1 Methodology

The units of analysis in most of this section are land parcels, which planners treat in the 20k plan by assigning land parcels to plots and planned amenities, as discussed in Section 3.3. An assignment results in different prices for different plots and in the sorting of owners across plots, as discussed in Section 5.1.[12]

The reduced form estimates in this section consider the effects of local (hypothetical) planning changes. For example, consider a local planning decision to treat a tiny land parcel differently by reassigning it between two geographically proximate plots of different sizes.[13] Since this parcel is

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

[12] In the model (Section 5), the units are individual households, allowing us to consider not only prices and quantities, but also welfare.

[13] The RD discussed below corresponds directly to this thought experiment: a gridcell just inside the boundary of a larger-plot insula would have been assigned to a neighboring smaller-plot insula had the planner drawn the road a

tiny, its reassignment does not induce owners to re-sort (beyond their initial sorting, based on the plan), and its general equilibrium effects are negligible. However, this tiny parcel itself is affected due to a "total" combination of a plot size direct effect and an owner-income effect, if owner income differs across the plots. The decomposition of the total effect into these two components is made explicit in Section 5.2.

Such total effects are relevant for assessing the impacts of incremental changes that planners or individuals may have considered. For example, as noted in the Introduction, in the absence of plot size externalities (of which we find little evidence) in a welfare-maximizing allocation in our context, plot prices per square meter should increase weakly with plot size. If they do not, that signals misallocation. Total effects 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 and rarely happens, as discussed in Section 2.4. For example, if land prices per square meter vary by plot size and if owners could subdivide their plots, then there would be a tendency to equalize unit prices of land across plot sizes.[14]

In practice, we consider effects on 20 x 20 m 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 are 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:

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

where y_i is an outcome in the gridcell or its plot. 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 preexisting major paved road; elevation; ruggedness; and indicators for being within 100m of a river

few meters over, so the discontinuity identifies the effect of reassigning that marginal parcel between plot sizes.

[14]This is true even with fixed costs in construction, diminishing marginal utility of floor-space, or low demand for large plots (see e.g. Appendix 5 in Gyourko and Krimmel, 2021).

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. ϵ_{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 3,225 in our full sample.[15]

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

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

where Own_larger_i is indicator for gridcell i's insula having a larger mean plot size than the insula with which it shares a boundary segment (a common edge); $\mathbf{Dist}_i$ is the distance in meters to the boundary segment and its interaction with Own_larger_i indicator; $\mathbf{Boundary}_i$ is vector of boundary segment FEs; and ϵ_{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 of 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. 2 for the absolute difference in log mean plot size across the insula-pair boundary and its interaction with 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), 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.

4.2 Main empirical findings

4.2.1 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 each area, the logarithm of real land prices increased in all project areas, with a mean

[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.

increase of about two log points, corresponding to final land values worth about 7.5 times the initial government price in real terms.[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 4.2.3 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 are formally recorded in the cadaster, 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]

4.2.2 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 (1). In columns 1-3 of Table 2, we show the elasticity varies modestly around -0.5 as we change

[16] Coefficients on time period effects (not shown) indicate that the market quickly anticipated the value of 20k, since the roughly sevenfold 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.

controls. Relative to only time and data source controls, adding neighborhood fixed effects so the specification relies on within-neighborhood variation does increase the absolute value by about 20%, but then adding on all the local characteristics of the plot has minimal effect. A main theme in the empirics is that the estimate of ~ -0.5 for the elasticity is robust to many different sets of controls and specifications, which deal with various types of externalities as discussed later. 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.4 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.[18]

Next, we examine the implications of plot size for housing outcomes. Table 3, which reports estimates from Eq. 1, 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 direct effect and the owner income effect discussed in Section 4.1 roughly offsetting. For built plots, the elasticity of built area with respect to plot size is around 0.11-0.14, so that accounting for richer people buying bigger plots, bigger plots have more investment as is intuitive, but with a relatively low size elasticity. The final column shows that large plots also have a higher likelihood of housing multiple buildings.[19] Panels B-E of Figure A.3 show the same relationships non-parametrically, and at least for well-populated bins the relationships are generally 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. 2. Panel A shows estimates for all insula pairs, irrespective of the mean plot size difference between them. Here, as in the OLS estimates above, plot size reduces the price and the share built, and increases the probability of multiple buildings; the only qualitative difference from OLS is that here there is no significant effect on log building size. 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 to 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.[20]

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

[19] 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).

[20] First, we verified that they are balanced on first-nature fundamentals in Table A.5. Second, we select optimal bandwidths (Calonico et al., 2014) for each outcome; in Table A.6 all are more narrow than our preferred 100m

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 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. Note that the point estimate of the price elasticity with respect to plot size under the RD is almost the same (slightly larger in absolute value) as the OLS. In Figure A.5, 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.5 suggests that the effect on plot built is somewhat non-linear. Given the earlier discussion of direct versus owner income effects, this suggests that as the plot size gap starts to grow, the owner income effect offsets and then dominates the plot size effect, so as to lower the likelihood of the plot being 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 two things. First is the equilibrium effects of changing supply, which we turn to in Section 5, and second is to examine the role of externalities, which we turn to in the next two subsections.

4.2.3 Connectivity

All 20k plots benefit from local road access, but connectivity, especially to the city center, varies substantially and is a key amenity in our setting, as we show using evidence across and within areas. To highlight the cross-area comparison, we 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

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).

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 (1) 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.

4.2.4 Other planning choices

Beyond plot sizes and connectivity, planners make other choices: where to locate plots relative to natural features, how to configure insulae, what non-residential uses to designate and what share of land to allocate to them relative to residential plots, and how to group plots of different sizes. We examine each of these and find that, while some matter, their effects are modest relative to plot size and connectivity. This motivates the model in Section 5, which focuses on the latter two while abstracting from land use mixing and local externalities.

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 to planned communities: insula features and the effects of being on the edge of the planned area. For the first, we construct a Z-index of three insula characteristics: rectangularity, alignment with neighboring insula, and homogeneity of plot sizes within the insula.[21] We find that a more regular layout increases the share built and the probability that a plot is built. In Table A.7 we unpack the estimated effects of the Z-index into its three components, and find that insula rectangularity seems to drive the result, with alignment being consistent with the result and homogeneity not valued. For the second

[21] 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 plot sizes within the insula. Thus, the best values for the raw measures are 1 for perfect rectangularity, perfect 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.

variable in the middle panel of Table 6, 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.9 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 residents, but many are unkept (have wild growth, garbage, etc.). Table A.9 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 uses 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.6 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.[22]

Next, we investigate local neighborhood plot size composition effects to look for spillovers. In particular, we examine whether people value having neighboring plots that are bigger vs. smaller.

[22] Even though implementation rates for non-residential amenities are low, we find that conditional on implementation, the planners' guidelines on land use locations were followed, compared to a benchmark where implemented use locations were randomly selected among the non-residential plots. To show this, Table A.8 reports the 'compliance 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.

While this allows for clearly interpretable reduced-form effects, apart from amenity effects of proximity to different plot sizes per se, such effects could reflect expectations of future sorting by socioeconomic status.

We repeat the specification in Eq. 2, but this time for boundaries between super-insulae rather than insulae, controlling for the log size of the own plot.[23] 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 do have a more limited sample for the price data.

We have two sets of results. The first concerns whether controlling for neighborhood plot size composition affects the measures of plot size elasticities. Here we note that the coefficient estimates for own plot size in Tables 6 and 7 are almost identical, with the exception of plot built or not, which now has a significant positive elasticity. Notably, the price elasticity remains ~ -0.5, suggesting that it is not related to changing neighborhood composition at superinsula boundaries. One also worries at borders if there are distinct differences in the extent, or timing of building and whether unbuilt plots are unkept. We explore this in Appendix Table A.10 where we add a control for fraction of residential plots built within 100 meters. The price and other elasticities are unaffected.

The second set of results relates to how people value the composition of nearby plot sizes. Here we find limited effects. 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.[24] Although the effects are small, the results may suggest that owners of small – rather than large – plot owners value neighborhood uniformity. However, the small magnitude of effects motivates our abstraction in the model below from plot size externalities.

4.2.5 Sorting

Having established the effects of planning choices on land outcomes, we now document patterns of owner sorting, which the calibrated model in Section 5 accounts for. Our survey shows that the mean years of schooling among 20k 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. These patterns confirm that higher-income households sort into 20k areas.

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

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

Sorting also operates within 20k areas. To study this, Table A.11 uses survey data for "Near" areas, as discussed in Section 4.2.3.[25] The table shows mean years of schooling of owners of built plots. Schooling increases by occupancy year within plot size group. It also increases from small to medium plot sizes in the same time period—that is, owners of larger plots are more educated, and late-moving owners are more educated than earlier ones. These sorting patterns underpin the owner income effect discussed throughout this section: because higher-income owners select into larger and better-amenity plots, the total effect of planning choices on outcomes reflects not only the direct effect of the planning treatment, but also the composition of owners.

5 Model: specification, estimation, and counterfactuals

5.1 Modeling the development of 20k plot areas

The model focuses on the two most consequential planning choices identified in Section 4: plot size and connectivity to the city center. It abstracts from non-residential land use mixing, layout geometry, and local externalities, which are empirically less important. In essence, we keep the current road structure where every plot already is serviced by a local road, and we set aside the land for non-residential uses.[26] The setting is one where a planner fixes a menu of formal plots which city residents then choose from. This imposes a rigidity that secures property rights but constrains allocations. We analyze the impact of alternative menus of plot size compositions across communities for the fixed amount of residential land.

We formulate the household's choice problem in Dar es Salaam, setting time 0 to be the year 2005. Each household faces a nested choice: whether to remain in the city, where housing is typically informal, or buy a formal plot in a 20k planned area and eventually move there. We focus on households who live in the city before moving to 20k, rather than poorer immigrants from rural areas. Those who choose 20k select a plot type m (a size-amenity bundle (l_m, B_m)), decide when to move (τ), and how much to invest in housing (k). Let $D_{20k} \in \{0, 1\}$ indicate whether a household

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

[26]With low implementation rates, we do not yet have enough evidence to evaluate non-residential land use reallocation.

eventually moves to a 20k area. The optimization problem is:

$$
\max_{D_{i,20k}} \{U_{i,0},\ \max_{m}\ U_{i,m} + \nu_{i,m}\} =
$$

$$
\begin{cases}
U_{i,0} = \max_{h_i(t),\, z_{1i}(t)} \int_0^\infty \left[\varphi \ln h_i(t) + \beta \ln z_{1i}(t) + Ae^{-\theta t}\right] e^{-\rho t}\, dt \\
\quad + \omega_i \left(\int_0^\infty w_i e^{-\delta t}\, dt - \int_0^\infty (p\, h_i(t) + z_{1i}(t)) e^{-\delta t}\, dt \right), & D_{i,20k} = 0 \\
U_{i,m} = \max_{h_i(t),\, z_{1i}(t),\, z_{2i}(t),\, k_i,\, \tau_i} \int_0^{\tau_i} \left[\varphi \ln h_i(t) + \beta \ln z_{1i}(t) + Ae^{-\theta t}\right] e^{-\rho t}\, dt \\
\quad + \int_{\tau_i}^\infty \left[\varphi \ln(l_m^\alpha k_i^{1-\alpha}) + \beta \ln z_{2i}(t) + B_m\right] e^{-\rho t}\, dt \\
\quad + \omega_i \left(\int_0^\infty w_i e^{-\delta t}\, dt - \int_0^{\tau_i} (p\, h_i(t) + z_{1i}(t)) e^{-\delta t}\, dt \right. \\
\quad \left. - \int_{\tau_i}^\infty z_{2i}(t) e^{-\delta t}\, dt - r k_i\, e^{-\delta \tau_i} - R_m \right), & D_{i,20k} = 1
\end{cases}
\tag{3}
$$

To lighten notation, we drop the household index i and t throughout the exposition that follows. The plot-specific preference shock decomposes as $\nu_m \equiv \tilde{\mu} + \sigma\varepsilon_m$ where $\tilde{\mu} \equiv \ln(\mu)$, with $\mu \sim \Gamma(\mu^{shape}, \mu^{scale})$, captures the household's 20k-versus-city taste (common across plots). $\sigma\varepsilon_m$, with i.i.d. Gumbel ε_m captures idiosyncratic taste for a particular plot type; the scale parameter $\sigma > 0$ governs the within-plot taste dispersion; and the location parameter is normalized to zero. The decomposition allows the dispersion in 20k-versus-city taste (μ) to differ from the within-20k plot dispersion (ε_m): for example, sharp disagreement on whether to move to 20k combined with little disagreement on which plot type, or vice versa.[27]

When $D_{20k} = 0$, the household stays in the city permanently, renting housing h at unit price p (or owning at a fixed opportunity cost, p) in the thick housing market, consuming non-housing goods z_1 (the numeraire), and enjoying a city amenity A that deteriorates at rate θ relative to the fixed 20k amenity B_m. We think of this deterioration as capturing the relative worsening of traffic conditions, air quality, and safety in the growing city, and it rationalizes the gradual movement of households to 20k areas. ω is the Lagrange multiplier on the budget constraint.

When $D_{20k} = 1$, the household purchases a formal plot of type $m = (l_m, B_m)$ at market price R_m at time 0.[28] Before moving at time τ, the household lives and consumes in the city as in the $D_{20k} = 0$ branch. Upon moving, it invests k in housing capital at unit cost r and consumes housing services $l_m^\alpha k^{1-\alpha}$ and non-housing goods z_2, enjoying a fixed location amenity B_m that varies

[27] The log-gamma distribution of $\tilde{\mu}$ exhibits the exponential tail behavior characteristic of extreme-value distributions like the Gumbel. We also tried a log-normal distribution; its heavier upper tail spread non-negligible mass over very large values of μ, making the welfare integrals hard to evaluate numerically.

[28] 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 they are willing to pay the market price, since this is the opportunity cost of keeping the plot.

across 20k areas.[29] The plot size l_m is fixed by the planner and the household cannot subdivide or expand it. Period income w measures permanent income, since, with a perfect capital market, all that matters is $W \equiv w/\delta$, the present discounted value of lifetime earnings (including any endowment).[30]

We set $\rho = \delta$, so optimized non-housing consumption is constant over the lifetime ($z_1 = z_2 \equiv z$). We also assume that owners cannot rent out their 20k plots, consistent with the small share of renters we observe. The first-order conditions, derived in Appendix C.1, yield expressions for the optimal timing of move τ and housing investment k, both of which lack closed-form solutions and are solved computationally. Denote the resulting indirect utilities by $V_0(w;\Theta)$ for the city, which has a closed form, and $V_m(w, R_m;\Theta)$ for 20k on plot type m, which is solved computationally. A household moves to its preferred plot type in 20k rather than staying permanently in the city if $\max_m(V_m(w, R_m;\Theta) + \nu_m) \geq V_0(w;\Theta)$.

There is a fixed supply of 20k plots of each type N_m, with $\sum_m N_m = 36{,}215$ in total. We assume the 800,000 households in Dar es Salaam in 2005 each potentially demand one plot. Their annual incomes follow $w \sim \Gamma(w^{shape}, w^{scale})$, which is fit to observed household incomes in a 2015 survey.[31] In equilibrium, prices and allocations are such that demand equals supply for each plot type simultaneously and no household prefers to switch. Higher-income households tend to sort into larger and higher-amenity plots. Sorting is probabilistic under $\sigma > 0$: every income assigns positive probability to every plot. Details on solving the equilibrium are in Appendix C.2.

5.2 How planning choices affect outcomes

The model generates predictions about how plot size and amenity affect three outcomes: land prices, development timing, and housing investment. Each outcome is shaped by a *direct effect* (holding owner income constant) and a *sorting effect* (larger or better-amenity plots attract richer owners). While the reduced-form estimates in Section 4 capture the total effect (the sum of both channels), the model provides a framework to decompose it into its direct and sorting components. The derivations are in Appendix C.3. Additionally, Figure 5 panels C and D, discussed in Section 5.4, illustrate the magnitude of the direct effect: at a given income level, the vertical gaps between plot types show how outcomes change with size or amenity holding w fixed.

Land prices (R_m). In any equilibrium, larger and higher-amenity plots command higher total prices (otherwise a consumer on the smaller plot would prefer to switch to the larger plot) but the

[29] We map B_m to proximity to the city center (Section 5.3). Because our counterfactuals hold each area's location fixed, B_m is a revealed-preference sufficient statistic for a location's net value, including commuting; a location-only commute cost would be absorbed into B_m, leaving the plot-size results unchanged.

[30] As we document, 20k owners are wealthier and more educated than the Dar es Salaam average, and thus less likely to face binding borrowing constraints.

[31] These survey data are from Balboni et al. (2020). 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 alternatives.

price *per square meter* need not rise with plot size, and indeed the estimated model (Section 5.3) reproduces the empirical elasticity of -0.5.

Development timing (τ). Two forces work in opposite directions. The direct effect of a larger or higher-amenity plot is to hasten development: the household has more to gain from the improved plot relative to the deteriorating city. This is reinforced by the indirect price channel: more desirable plots are more expensive, shifting expenditure from the city to 20k and inducing faster moves. But the sorting effect goes the other way: larger and better-amenity plots attract higher-income owners, who build later because the relative importance of city amenity deterioration is smaller for richer households, so they prefer to stay longer in the city's unrestricted housing market. The total effect is therefore ambiguous, consistent with the empirical finding in Section 4 that plot size has no clear effect on whether a plot is built (except when size gaps are large, in which case sorting dominates). For amenity, the direct effect appears to dominate: plots in Near areas are built sooner than those in Far areas.

Housing investment (k). The direct effect of a larger or higher-amenity plot on investment is ambiguous: the complementarity between plot characteristics and capital pushes investment up, but the higher plot price squeezes the budget and pushes it down. The sorting effect, however, is unambiguously positive with richer owners investing more. Empirically, sorting dominates: larger and higher-amenity plots are associated with greater investment, as documented in Section 4 and consistent with the model solution illustrated later in panel D of Figure 5.

5.3 Estimation

We estimate the parameters of the model focusing on the equilibrium observed in the data. Two simplifications streamline the estimation. First, as discussed in Section 4.2.3, we partition the 20k project areas into "Near" (amenity B_N) and "Far" (amenity B_F) groups; further differentiation within the Near group yielded very similar amenity values. This gives three locations: Near, Far, and the city. Second, following the planners' three-tier classification, we discretize plot sizes into small, medium, and large categories, each represented in the model by its mean size (600, 1200, and 2000 sqm, respectively). Together, these yield six plot types (two amenities × three sizes), indexed by m.

Some parameters are taken from the literature: the housing consumption share $\varphi = 0.3$, land share in housing $\alpha = 0.3$, non-housing consumption share $\beta = 0.7$, and real interest rate $\delta = 0.04$, as reported in Panel A of Table 8.[32] This leaves six parameters (B_F, B_N, θ, μ^{shape}, μ^{scale}, σ) to estimate. We do so by matching twelve moments (the average land price R_m and the fraction of plots built by 2020 S_m for each of the six plot types) using a loss function based on the relative absolute deviation of prices and shares built, weighted by the relative share of plots of each type.

[32] We conducted robustness checks around these values; results are not sensitive to plausible variations.

Details on the loss function and the optimization algorithm are in Appendix C.4.

The estimated parameter values are in Panel C of Table 8. One observation is worth highlighting: the deterioration rate $\theta = 0.008$ implies that the city amenity equals the Near 20k amenity ($B_N =$ 0.665) after approximately 50 years.

Table A.12 compares the model and data moments. The match is close in most cases. The main exceptions are small- and medium-size Near plots, where prices are off by 28% for small and 26% for medium. The model is also consistent with patterns not targeted in estimation: owner education rises from small to medium plots and over time within plot size groups, as documented in Section 4.2.5. The model also does well matching the income distribution of movers from our 20K survey circa 2020. Notably, the model parameters are calibrated on different income data for Dar es Salaam city circa 2015 (Balboni et al., 2020). As mentioned earlier, the 10th percentile of movers earns around $910 in the data and $1,068 in the model (rising to $1,134 among those who build by 2020). The median income of movers is $3,891 in the data and $3,447 in the model (falling to $2,805 among those who build by 2020).[33]

5.4 Characteristics of the equilibrium

Figure 5 illustrates the equilibrium with solid lines for Far and dashed for Near, with green, red, and purple respectively covering S, M, and L lots. Panel A plots the conditional logit choice probability $\Pr(m \mid \text{move}, w)$ for each plot type across the income distribution; at every income the six probabilities sum to one. In Near and Far, higher-income households are increasingly likely to choose larger and higher-amenity plots, as in the data. Panel B weights these probabilities by the move probability and the income density, so that the six curves now sum to the density of 20k movers, shown in black.

Panel C illustrates the opposing forces affecting development timing discussed in Section 5.2. The vertical gap between any two lines (plot types) shows the direct effect of moving to a larger plot, which reduces τ at each income level (holding prices fixed). But within each plot type, τ increases with income (again holding prices fixed), i.e., richer owners on the same plot build later, which is the force that drives the sorting effect. For amenity, plots in Near areas (dashed lines) are built sooner, consistent with the reduced-form evidence. Panel D shows that housing investment increases sharply with income and is little related to either l or B directly, confirming that sorting dominates the ambiguous direct effects, as discussed above.

[33]These figures correspond to movers, so are distinct from the average income in the city reported in the appendix to be $2,755 in the data.

5.5 Counterfactuals

Motivated by the strong negative elasticity of price per square meter with respect to plot size, which suggests misallocation in the current distribution, we use the model to assess how the allocation of residential land across plot sizes could be improved. We reallocate private residential land by changing the partition of small, medium, and large plots within Near and Far areas, holding constant the total residential land area in each. Plot size changes effectively carve up existing insulae into fewer or more plots.

We consider two objectives. First, a Kaldor–Hicks social planner maximizes the total surplus of the project: the sum of aggregate land values in 20k formal areas and a money-metric measure of consumer surplus. Equivalent variation, the one we adopt, is the present discounted value of the annual income supplement Δ that would leave each 20k resident indifferent between their plot and the city, i.e., $\max_m(V_m(w, R_m; \Theta) + \nu_m) = V_0(w + \Delta; \Theta)$. This money-metric approach circumvents the need to impose assumptions on how land values are distributed across households. Since it weighs the consumer surpluses of the rich and poor equally, it does not mechanically drive the finding below that welfare is maximized by accommodating more lower-income residents.[34] Second, a monopolist developer (or a local government pursuing land value capture) maximizes aggregate land values, a common approach in the urban literature.

The results in Table 9 show that compared to the baseline (column 1), welfare maximization (column 2) increases consumer surplus by 11.1% and total surplus by 2.3%. The number of plots rises by 39%, average plot size falls by 28%, and the median owner income declines by 11%. Most crucially, this allocation widens participation to serve many more lower-income households: the 10th-percentile income of 20k movers drops from \$1,068 to \$785. In contrast, the monopoly developer (column 3) moves in the opposite direction: the supply of plots falls by 35%, average plot size increases by 54%, and median owner income rises by 12%. Total land values increase by 3.8% while total surplus declines by 8.1%. On all dimensions, the baseline lies between the welfare-maximizing and land-value-maximizing solutions.

More detail on incidence across the household income distribution in Dar es Salaam is given in Appendix Table A.13. In the baseline equilibrium, 4.0% of the households in 20k are from the bottom quintile of the Dar income distribution, while 39.4% are from the top. Under welfare maximization, the bottom quintile share more than doubles to 8.9% and the number of households triples due to the increase in total plots. The share of the top falls to 30.8%, but the absolute number of households from the top quintile increases modestly by about 8%. Under land value maximization, the share of households in the bottom quintile falls to 0.0%, while the top quintile's share rises to 43.2% (but their absolute number falls by 29% given the large decline in plots offered).

[34] The welfare gains from accommodating poorer residents are even stronger when measured in utility rather than surplus, as these residents have higher marginal utility.

The table also shows that in all the scenarios, the higher deciles capture an even larger share of total consumer surplus than their population shares.

The occupancy rate circa 2020 is lower under welfare maximization (31.5%) than at baseline (48.7%), reflecting later building, not less, and already priced into the discounted surplus. The welfare-maximizing shift toward smaller plots delays optimal τ (Panel C of Figure 5). This equilibrium contains relatively more lower-income households who build sooner, but this income effect is partially offset by lower equilibrium prices in such an allocation, as a lower price has the same effect as higher income on net income in the budget constraint.

Table 10 reports the plot type allocations and prices for each regime. In the baseline, price per square meter decreases with plot size for a given amenity, as in the reduced-form results. Under welfare maximization, prices per square meter *increase* modestly with plot size. Under land value maximization, they decline more steeply than in the baseline.

The pattern of prices across regimes reflects a general principle arising in a second-best world where land supply for formal development is fixed.[35] As shown in Appendix C.5, the total surplus from any reallocation of the supply of plots decomposes as:

$$\Delta TS = \underbrace{\Delta LV^{PE}}_{\text{PE land value change}} + \underbrace{\Delta CV^{\text{reopt}}}_{\text{reallocation surplus}} \tag{4}$$

The intuition for this result comes from realizing that price changes on existing land are pure transfers between landowners and residents; they do not create surplus. What changes total surplus is the reallocation of land and residents themselves. The first term, ΔLV^{PE}, is the partial equilibrium land value change from a reallocation such as splitting plots; it is positive whenever land prices per square meter decline with size. The second term, $\Delta CV^{\text{reopt}} \geq 0$, is the households' surplus from re-optimizing their housing and consumption choices after a reallocation. Since the households' reallocation surplus is non-negative, the social planner may split plots past the point of price-per-square-meter equalization. At the welfare optimum, price per square meter therefore is weakly increasing in plot size. This provides a sufficient statistic for detecting misallocation: when price per square meter falls with plot size—as in our data—splitting large plots into smaller ones increases welfare. Planners could use this diagnostic to identify inefficient plot size distributions that may arise from colonial-era rules-of-thumb.

For the monopolist, the logic reverses. Subdividing a plot only increases land values if the partial equilibrium gains outweigh the general equilibrium losses from increased supply pushing all prices down. This requires price per square meter to decline with size.

To illustrate how these forces operate in practice, consider in the baseline the effects of splitting

[35] The general principle and resulting sufficient statistic can be shown exactly using the compensating variation, CV, as the measure of consumer surplus, which is only approximately equal to the equivalent variation, EV, used in the counterfactual computations.

one medium-size Near plot (l_M, B_N) into two small plots (l_S, B_N). The partial equilibrium land value gain is $1{,}200 \times (\$8.16 - \$6.71) \approx \$1{,}737$. But adding one plot to total supply induces a general equilibrium decline in prices (about 0.0014% on average) to attract one extra mass of households to locate in 20k areas, which reduces aggregate land values by $3,099. The net effect on land values is therefore −$1,362. However, the GE land value decline is a transfer to residents on existing plots, canceling in aggregate. The reoptimization itself generates additional consumer surplus. Together they raise consumer surplus by $3,413 and total surplus by $2,052. This example illustrates how rigidities that prevent the splitting of formal plots restrict entry and prop up land values, but lower aggregate welfare.

As an aside, welfare maximization increases population density, which could generate externalities not captured by the model.[36] Positive externalities, such as greater provision of local services with more residents, would strengthen the case for additional small plots. Negative externalities from crowding would weaken it. Our reduced-form estimates suggest that neighborhood composition effects are modest, though they point to small positive effects from grouping together small plots.

Next, we consider the shadow value of adding a square meter of formal residential land—that is, the change in total surplus when relaxing the planner's fixed-land constraint. At the welfare-maximizing allocation, this multiplier is approximately equalized across plot sizes within each location: about $\$6.83/\text{m}^2$ in Near areas and $\$2.27/\text{m}^2$ in Far areas, both above price per sqm values at the optimum for all plot sizes, given the consumer surplus gains beyond the price per sqm PE gain in land values. This is a direct consequence of the interior optimum—if an additional square meter of small plots generated more surplus than one of larger plots, the planner could raise welfare by reallocating land across sizes, contradicting optimality. All these values are well above the approximately $1 per square meter it costs the government to service these plots, as discussed in Section 4.

As a final exploration, we consider extensive margin effects by resolving the welfare problem after doubling the residential land in each area. The shadow values decrease to about $\$6.19/\text{m}^2$ in Near areas and $\$1.76/\text{m}^2$ in Far areas, still considerably above the servicing cost. This suggests the project could be increased well past its current scale. However, this figure is best read as an optimistic illustration. It holds the city house price fixed—a reasonable approximation given the city's thick, highly elastic informal housing supply, even as 20k's household share rises from 4.5% to above 10% under doubling. Any price response would be downward, improving the city outside option and dampening take-up, so fixing the price favors the expansion case. Servicing costs also likely rise with scale. We therefore treat this exercise as more speculative than the baseline

[36]Table A.14 shows that small plots have population densities that are almost double those of medium plots and roughly three times those of large plots.

shadow-value gap exercise above.

6 Conclusion

De-novo urban planning is a key policy option for developing-country cities facing large and rapidly growing informal areas. Despite its importance for Africa's 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 de-novo planning entails, combining a novel structural model with reduced-form evidence and new data we collect, in the setting of the "20,000 Plots" project in Dar es Salaam.

We identify two sources of the value de-novo planning creates—here, a rough doubling of land values relative to nearby informal areas: secure property rights, and access, both within projects (local roads connecting to the main paved network) and across space (where a project is sited). Relative to the earlier World Bank sites-and-services projects halted in the 1980s amid concerns that costs were not recouped and the poor were excluded, the 20k plots were cheaper, and their costs were recouped—though, as we discuss below, the poor remained largely excluded.

De-novo neighborhoods also have limitations that can be mitigated. Chief among them is an oversupply of large plots, which command lower land values, less built space, and lower density—likely a legacy of colonial-era rules and norms. Offering relatively more small plots, through a less rushed sales process, would make projects both more valuable and more inclusive, raising social welfare; the observed allocation looks closer to one that maximizes land values than social welfare. Inclusion has limits, however: because formal plots are worth most to higher-income households, even the welfare-maximizing allocation concentrates gains near the top of the income distribution, and the poorest remain largely excluded—least so under welfare maximization and most under land-value maximization. Non-residential amenities, by contrast, are largely ignored, generating neither land-value appreciation nor built activity, most likely because of low implementation rates.

Two further margins matter beyond how a project's land is subdivided. First, its scale and location: the marginal value of formal land far exceeds the cost of providing it, most of all in less peripheral areas, so the project was, if anything, too small and too peripheral. Where greenfield land is not prohibitively costly to assemble, expanding and better-siting such projects would create additional value and extend formal access to more households. Second, development is slow: only half the residential plots are built upon, reflecting both plot characteristics (smaller plots with worse amenities) and higher-income owners delaying their move, dynamics we pursue in follow-up work.

Finally, our analysis yields a diagnostic that travels beyond our setting: the price per square meter should be weakly increasing in plot size. Where it falls, as in 20k and as evidence from other contexts suggests is common, the plot-size distribution is skewed inefficiently toward large plots, where plots cannot be freely subdivided. More broadly, our results show how combining planning with market signals can make de-novo projects both more valuable and more inclusive.

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.

Abdulkadiroğlu, A., Angrist, J. D. and Pathak, P. A. (2014), 'The elite illusion: Achievement effects at Boston and New York exam schools', *Econometrica* **82**(1), 137–196.

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.

Baldomero-Quintana, L., Woo-Mora, L. G. and De la Rosa-Ramos, E. (2025), 'Infrastructures of race? colonial indigenous segregation and contemporary land values', *Regional Science and Urban Economics* **110**, 104065.

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.

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

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

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

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.

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.

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

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.

Gonzalez-Navarro, M. and Quintana-Domeque, C. (2009), 'The reliability of self-reported home values in a developing country context', *Journal of Housing Economics* **18**(4), 311–324.

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.

Gyourko, J. and Molloy, R. (2015), Chapter 19 - regulation and housing supply, *in* G. Duranton, J. V. Henderson and W. C. Strange, eds, 'Handbook of Regional and Urban Economics', Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, pp. 1289–1337.
URL: *https://www.sciencedirect.com/science/article/pii/B9780444595317000193*

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. (2025), 'Colonial legacy and land market formality', *Journal of Urban Economics* **149**, 103789.

Harari, M. and Wong, M. (2026), 'Slum upgrading and long-run urban development: Evidence from indonesia', *The Review of Economic Studies* **93**(4), 2646–2679.
URL: *https://doi.org/10.1093/restud/rdaf090*

Henderson, J. V., Libano-Monteiro, F., Manara, M., Michaels, G. and Regan, T. (2025), Economics of greenfield urban planning, CEP Discussion Paper 2101, LSE.

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.

Ihlanfeldt, K. R. (2004), 'Exclusionary land-use regulations within suburban communities: A review of the evidence and policy prescriptions', *Urban Studies* **41**(2), 261–283.

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

Jaupart, P., Chen, Y. and Picarelli, N. (2017), Cholera in Times of Floods Weather Shocks and Health in Dar es Salaam.

Kamunyori, S. W., Arif, R. S., Mwanga, K. M. and Tasker-Brown, J. K. (2022), Reconsidering sites and services: A global review, Technical Report 172260, World Bank Group, Washington, D.C.

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 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.

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.

Song, J. (2025), 'The effects of residential zoning in u.s. housing markets', *Journal of Urban Economics* **149**, 103784.

Sood, A. and Ehrman-Solberg, K. (2026), 'The long shadow of housing discrimination: Evidence from racial covenants', *The Review of Economic Studies* p. rdag079.

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.

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.

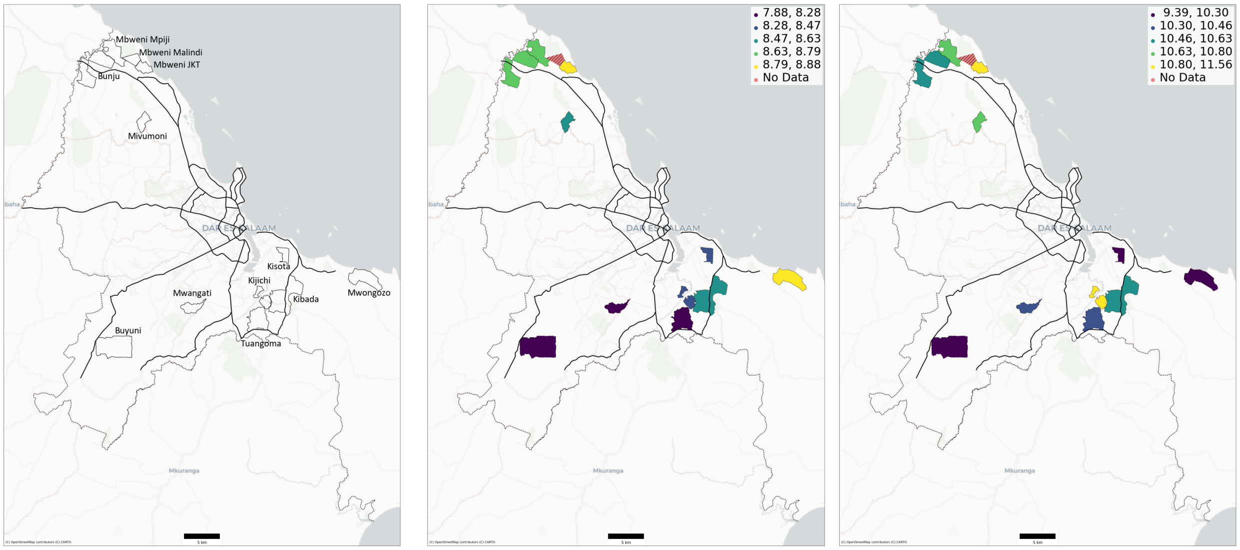
Visagie, J. and Turok, I. (2020), 'Getting urban density to work in informal settlements in africa', *Environment & Urbanization* **32**(2), 351–370.

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

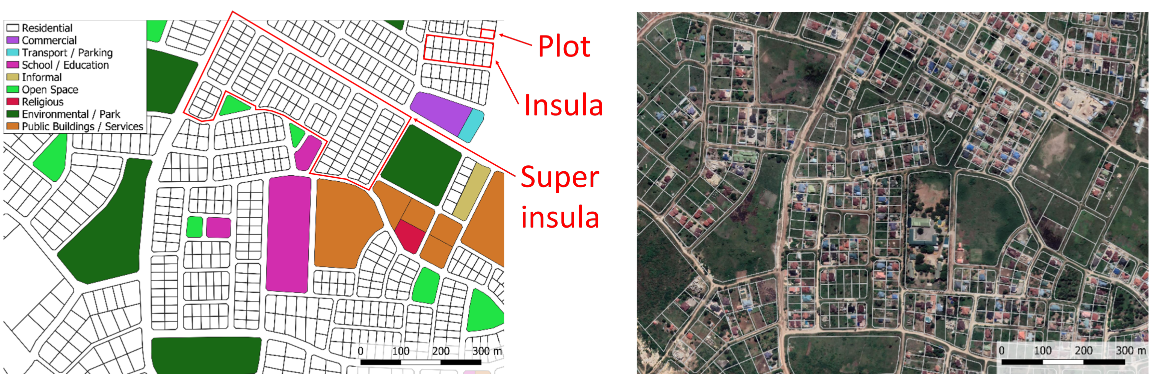
(A) 20k project area names (B) Initial ln price per sqm (C) Current ln price per sqm



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



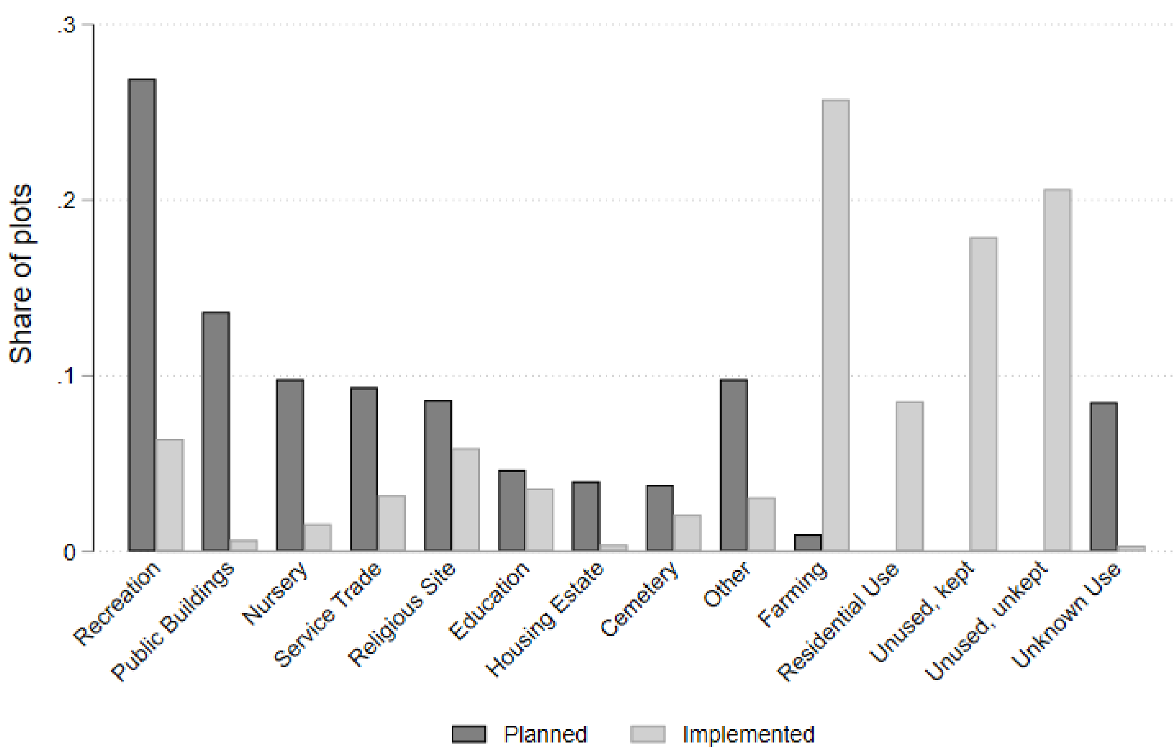
Notes: This figure shows an example of planned plot boundaries in Mbweni Mpiji. In Panel A, each plot is colored by its planned use. In Panel B, satellite imagery is displayed in the background.

Figure 3: Example of 20k boundary in Tuangoma

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

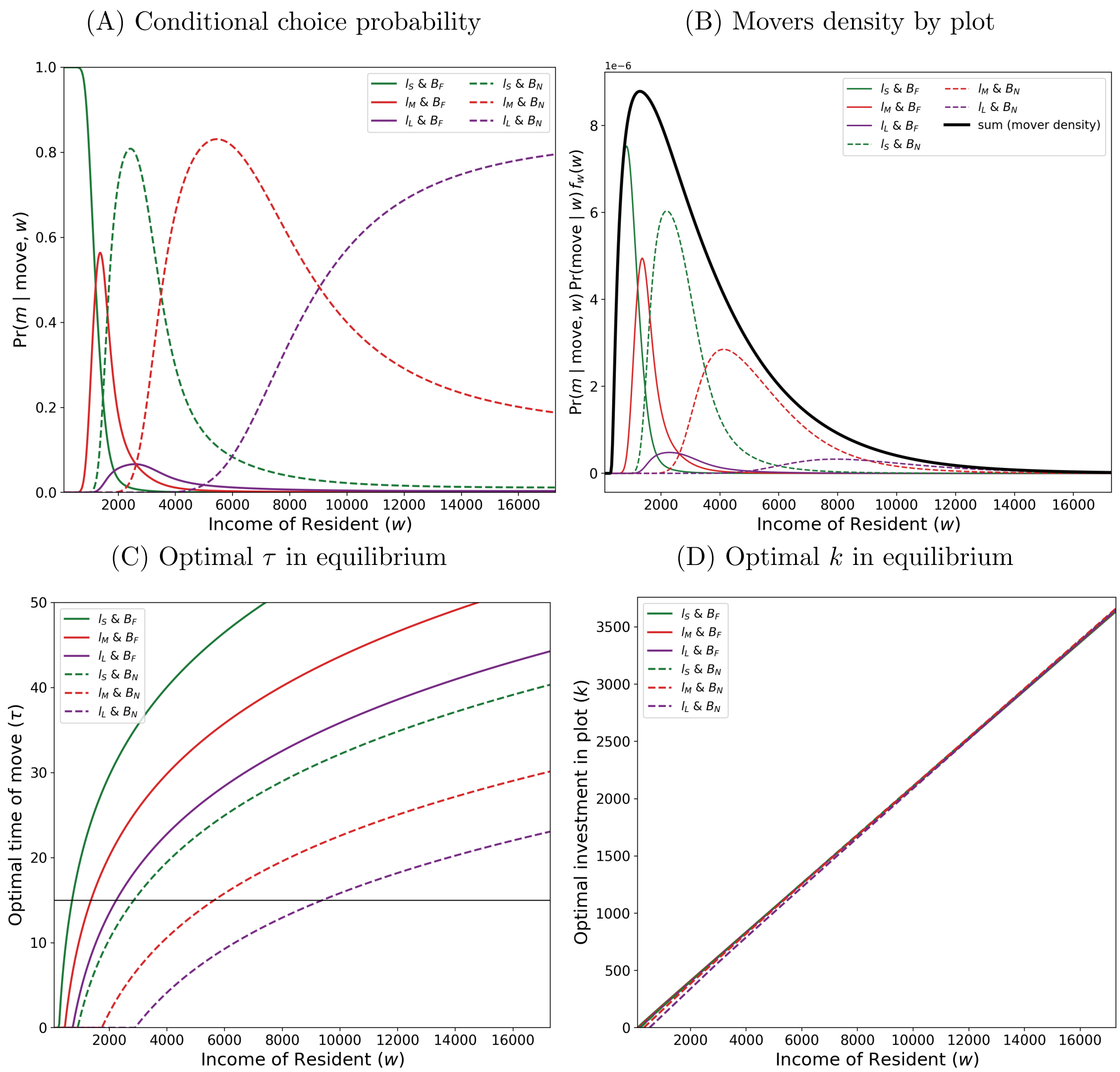
Notes: This figure shows an example of a 20k project boundary in Tuangoma. In Panel A, background satellite imagery is from 2001 (pre-implementation). In Panel B, background satellite imagery is from 2021 (post-implementation).

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. Panel (A) plots the conditional logit choice probability $\Pr(m \mid \text{move}, w)$ for each of the six plot types; at every wage the six probabilities sum to one. Panel (B) weights these by the probability of moving and the wage density, $\Pr(m \mid \text{move}, w)\ \Pr(\text{move} \mid w)\, f_w(w)$; the six curves now sum to the density of 20k movers shown in black. In panel (C), the outcome is the optimal time of move (τ). The horizontal line at $\tau = 15$ marks the data observation window. In panel (D), the outcome is the optimal capital investment (k).

Tables

Table 1: Land price inside and nearby 20k areas

	(1) Ln Price	(2) Ln Price	(3) 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 nearest 20k area. Standard errors in parentheses are clustered by 20K area.

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

	(1) Ln Price per sqm	(2) Ln Price per sqm	(3) Ln Price per sqm	(4) 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 four 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 less 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 ×Δ 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, 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: Prices and built outcomes on plot sizes in 20k areas (super-insula 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	0.028	-0.0020	-0.0018	-0.00049	0.024
	(0.060)	(0.0025)	(0.010)	(0.020)	(0.015)
Own Smaller × Dist. (km)	0.25	0.056	0.23	-0.044	0.011
	(0.43)	(0.017)	(0.067)	(0.12)	(0.084)
Own Larger × Dist. (km)	0.56	-0.031	0.026	0.10	-0.039
	(0.35)	(0.018)	(0.070)	(0.13)	(0.089)
Ln plot size	-0.52	-0.065	0.027	0.18	0.20
	(0.087)	(0.0032)	(0.013)	(0.027)	(0.018)
Mean Outcome	10	0.11	0.49	5.3	0.38
N (gridcells)	3962	93025	93025	45712	45712
N (plots)	1405	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 1-5 the outcomes are the same 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. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table 8: Parameters

Parameter	Description	Value	Source
—Panel A. External sources—			
ϕ	Consumption housing elasticity	0.3	Combes et al. (2021)
β	Consumption numeraire elasticity	0.7	Combes et al. (2021)
δ	Interest rate	0.04	Henderson et al. (2021)
ρ	Discount rate	0.04	Henderson et al. (2021)
α	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^{shape}	Shape of city distribution	1.1703	Balboni et al. (2020) + Estimation
w^{scale}	Scale of city distribution	2354.2	Balboni et al. (2020) + Estimation
—Panel C. Moment estimation—			
B_F	Far Amenity	0.534	Joint Internal Estimation
B_N	Near Amenity	0.665	Joint Internal Estimation
θ	City deterioration rate	0.008	Joint Internal Estimation
μ^{shape}	Shape of 20k vs city taste shock	0.011	Joint Internal Estimation
μ^{scale}	Scale of 20k vs city taste shock	77.76	Joint Internal Estimation
σ	Scale of plot-type taste shock	0.073	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 Value Max (3)
	—Panel A. Info—		
Total supply of 20k plots	36,215	50,168	23,489
Average plot size	1046	755	1613
20k mean wage ($ 2021)	4205	3648	4739
20k median wage ($ 2021)	3447	3078	3853
20k 10th-percentile wage ($ 2021)	1068	785	1625
20k Occupancy Rate % (by 2020)	48.7	31.5	77.5
	—Panel B. Values—		
Total Land Values ($M, 2021)	225.7	208.8	234.3
Consumer Surplus ($M,2021)	249.3	277.1	202.1
Total Surplus ($M, 2021)	474.9	486.0	436.3

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

	Far Areas			Near Areas		
Panel A: Baseline						
	S	**M**	**L**	**S**	**M**	**L**
Number of Plots	4975	4192	938	10068	12661	3381
Land Price per Sqm	3.98	2.73	2.29	8.16	6.71	6.57
Panel B: Welfare Maximization						
	Far Areas			**Near Areas**		
	S	**M**	**L**	**S**	**M**	**L**
Number of Plots	14433	683	206	25619	7287	1940
Land Price per Sqm	2.08	2.20	2.24	6.68	6.76	6.79
Panel C: Land Value Maximization						
	Far Areas			**Near Areas**		
	S	**M**	**L**	**S**	**M**	**L**
Number of Plots	69	1483	4035	3086	4360	10456
Land Price per Sqm	7.98	4.52	3.16	11.70	8.15	6.51

Notes: This table compares plot allocations and land prices under three scenarios: Baseline, Welfare Maximization, and Land Value Maximization. Each panel reports the number of plots and the land price per square meter for the six plot types, defined by size (S, M, L) and area (Far, Near).

Supplementary Appendix
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, Tuangoma, 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.[37].

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 sec-

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

retary and town planners of 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.[38] Second, we digitize the planned road reserves and their widths by manually tracing the georeferenced TPDs.[39]

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.[40] 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.

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

39 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).

40 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 ≈ $0.88 per square meter of plot (assuming a mean of about 1,046 sqm).

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.[41]

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.[42] 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

[41] 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.

[42] 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.[43] 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

[43] 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 impose restrictions on the price data for within-20k analysis. 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.

Separately, we impose a different set restrictions on the transaction data for analyses that also look outside 20k areas, e.g. Table 1. We start with the full 2,404 transactions from the dalali data, we drop any that are listings and do not add the prices from the resident questionnaire in order to keep the samples in and outside of 20k areas comparable. This leaves us with 2,177 transactions, including: 1,302 on 20K plots, 288 on non-20K formal plots, and 587 on non-20K informal plots. Finally, we use the plot size as reported by the dalali, since we do not have matched plots for transactions outside of 20k areas. After dropping transactions missing a plot size, we have 2,074 transactions, including: 1,246 on 20K plots, 266 on non-20K formal plots, and 562 on non-20K informal plots.

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.[44]

Of the 5,900 randomly selected plots, 4,683 plots were eligible for interview (for reasons explained below). Two additional questionnaire clusters with 35 plots each were dropped during the questionnaire's implementation, leaving 4,613 plots.[45] 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.[46]

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.[47]

[44] 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.

[45] One due to a local land conflict and another due to personal circumstances of enumerators.

[46] 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.

[47] plots were ineligible because: (i) they were undeveloped – possibly due to changes in land use since the imagery

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.[48] Prospective 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. Part 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

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) two clusters used for piloting (70 cases); and (vi) other reasons (168 cases). 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. Therefore, we had 1,217 ineligible plots in total.

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

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 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.[49] 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

[49] 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.

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).

C Model details

C.1 The optimization problem and first-order conditions

From the household's optimization problem in Eq. 3, consider the $D_{20k} = 1$ branch. Conditional on choosing to move to 20k and selecting plot type $m = (l_m, B_m)$, the household optimizes over city housing h, non-housing consumption z_1 and z_2, housing capital k, and timing of move τ. The plot size l_m, amenity B_m, and price R_m are taken as given. Recall that p is the rental price of housing in the city, r is the purchase cost of capital, z is the numeraire, A is the initial city amenity that declines at rate θ, and we set $\rho = \delta$ so that optimized consumption is constant over time. The plot-specific taste shock $\nu_m = \tilde{\mu} + \sigma\varepsilon_m$ enters utility additively and thus does not appear in any of the conditions below.

The first-order conditions are:

1. $\beta = \omega z_1 = \omega z_2$, so $z_1 = z_2 \equiv z$;
2. $h = \frac{\varphi}{\omega p} = \frac{\varphi z}{\beta p}$;
3. $k = \frac{\varphi(1-\alpha)}{\omega r \delta}$;

4. $\frac{w-z}{\delta} - \frac{ph}{\delta}(1 - e^{-\delta\tau}) = R_m + rke^{-\delta\tau}$ (budget constraint);

5. $\varphi \ln h + Ae^{-\theta\tau} - \varphi \ln(l_m^{\alpha} k^{1-\alpha}) - B_m + \omega[\delta rk - ph] = 0$ (optimal timing).

Substituting into the budget constraint (item 4) yields the Lagrange multiplier:

$$\omega(\tau, R_m) = \frac{\beta + \varphi(1 - \alpha e^{-\delta\tau})}{w - \delta R_m}.$$

Through substitution using the FOCs and ω, we derive expressions for τ and k:[50]

$$\begin{aligned}\ln(w - \delta R_m) = &- \frac{1}{\alpha\varphi}\Big[Ae^{-\theta\tau} - B_m - \varphi\alpha \ln\big(\beta + \varphi(1 - \alpha e^{-\delta\tau})\big) \\ &+ \varphi \ln\Big(\frac{\varphi}{p}\Big) - \varphi\alpha \ln l_m - \varphi(1-\alpha) \ln\Big(\frac{\varphi(1-\alpha)}{r\delta}\Big) - \alpha\varphi\Big]\end{aligned} \tag{5}$$

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

The expression for k is an implicit function of τ, the model parameters, and non-choice variables. We therefore solve for τ computationally in the model estimation. These expressions correspond to the conditions used to derive the comparative statics below.

Given the optimal choices, the indirect utility of a household that enters 20k on plot type m is $V_m(w, R_m; \Theta)$. Because τ has no closed-form solution, neither does V_m, so it too is solved computationally. The indirect utility of staying in the city permanently ($D_{20k} = 0$, i.e., $\tau = \infty$) does have a closed form:

$$V_0(w; \Theta) = \frac{1}{\rho}\left(\varphi \ln \frac{\varphi w}{p(\beta + \varphi)} + \beta \ln \frac{\beta w}{\beta + \varphi}\right) + \frac{A}{\rho + \theta}$$

which follows from Eq. 3.

A household chooses 20k over the city if $\max_m(V_m(w, R_m; \Theta) + \nu_m) \geq V_0(w; \Theta)$, where the inner maximum is taken over the realized plot-specific shocks $\nu_m = \tilde{\mu} + \sigma\varepsilon_m$. Conditional on entering 20k, the household picks the plot m that achieves the inner maximum. The 20k-or-stay margin is

$$\Pr(\text{move} \mid w) = \int_{-\infty}^{\infty}\left[1 - F_{\mu}\left(e^{V_0(w) - \sigma \log \sum_m e^{V_m/\sigma} - \sigma\xi}\right)\right] f_{\text{Gumbel}}(\xi)\, d\xi. \tag{7}$$

This follows by rearranging the move condition into $\tilde{\mu} \geq V_0(w) - \max_m(V_m + \sigma\varepsilon_m)$, or equivalently $\mu \geq \exp\big(V_0(w) - \max_m(V_m + \sigma\varepsilon_m)\big)$, using the Gumbel closure property $\max_m(V_m + \sigma\varepsilon_m) \overset{d}{=} \sigma \log \sum_m e^{V_m/\sigma} + \sigma\xi$ with $\xi \sim \text{Gumbel}(0, 1)$ to reduce the plot shocks to a single scalar Gumbel, and integrating μ and ξ out against their densities. With $\sigma > 0$ and Gumbel ε_m, the plot-choice

[50] Note that $w - \delta R_m > 0$, since a household would not spend its entire lifetime income $W = w/\delta$ on purchasing the 20k plot, which costs R_m.

probabilities follow the usual multinomial logit form,

$$\Pr(m \mid \text{move}, w) = \frac{\exp(V_m(w, R_m; \Theta)/\sigma)}{\sum_{j=1}^{6} \exp(V_j(w, R_j; \Theta)/\sigma)}. \tag{8}$$

C.2 Solving for the equilibrium

Equilibrium is the price vector $\mathbf{R}^* = (R_1, \ldots, R_6)$ at which demand equals supply on every plot type simultaneously,

$$D_m(\mathbf{R}^*) = N_m, \qquad m = 1, \ldots, 6,$$

where

$$D_m(\mathbf{R}) = H \int_0^\infty \Pr(\text{move} \mid w; \mathbf{R}) \cdot \Pr(m \mid \text{move}, w; \mathbf{R}) \cdot f_w(w)\, dw.$$

We solve it by computationally finding the roots on $\mathbf{x} \equiv \log \mathbf{R}$ of $\mathbf{F}(\mathbf{x}) = (D_m(e^{\mathbf{x}}) - N_m)_m = \mathbf{0}$. The allocation of plots to households follows from the choice probabilities: the mass of households on plot m from the income band $(w, w + dw)$ is $N \cdot \Pr(\text{move} \mid w; \mathbf{R}^*) \cdot \Pr(m \mid \text{move}, w; \mathbf{R}^*) \cdot f_w(w)\, dw$, where $H = 800,000$ is the number of households.

C.3 Comparative statics

As discussed in the main text, the outcomes affected by plot size and amenity are: land prices (R_m); the probability that a plot is developed by 2020, i.e., $\Pr(\tau < 15)$, which is inversely related to τ; and housing capital investment (k).

Land prices. Higher prices must be paid for higher l_m and B_m in any Nash equilibrium. For example, two plots with the same amenities but 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_m or l_m according to the equilibrium elasticities:

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

Development timing. Differentiating Eq. 5:

$$G\, d\tau = \underbrace{\frac{dw}{w - \delta R}}_{\text{owner income effect}} - \underbrace{\left(1 + \frac{\delta R}{w - \delta R}\eta_{R,l}\right) dl/l}_{\text{plot size effect}} - \underbrace{\left(\frac{B}{\alpha\varphi} + \frac{\delta R}{w - \delta R}\eta_{R,B}\right) dB/B}_{\text{amenity effect}} \tag{9}$$

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

Holding owner income constant, bigger plots and higher-amenity ones are developed sooner (τ is lower). The direct effect of a bigger plot makes the owner move sooner; 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 expenditure from the city to 20k and inducing faster development. But in equilibrium, bigger plots are purchased by higher-income households, and the impact of higher income in Eq. 9 is to raise τ and delay development. The total effect of plot size on τ is therefore ambiguous without controlling for owners' incomes. A similar argument applies to the effect of amenities B_m.

Housing investment. Differentiating Eq. 6 and substituting for $d\tau$ from Eq. 9:

$$\frac{dk}{k} = \frac{X}{X+Z}\left(\underbrace{\frac{dw}{w-\delta R}}_{\text{owner income effect}} + \underbrace{\left[\frac{Z}{X} - \frac{\delta R}{w-\delta R}\eta_{R,l}\right] dl/l}_{\text{plot size effect (ambiguous)}} + \underbrace{\left[\frac{B}{\alpha\varphi}\frac{Z}{X} - \frac{\delta R}{w-\delta R}\eta_{R,B}\right] dB/B}_{\text{amenity effect (ambiguous)}} \right) \tag{10}$$

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

Holding income constant, the direct complementary effect of an increase in l_m or B_m is to increase k. However, countervailing this is the indirect effect of the increase in price R_m, which squeezes the budget and reduces k. Thus, even holding income constant, the effect of an increase in l_m or B_m on k is ambiguous. With positive owner income effects, the reduced-form empirics and the structural estimation in Section 5.4 show that higher B_m or l_m is generally associated with higher investment.

C.4 Parameters of the model

Calibrated parameters (Panel A of Table 8). We set the housing consumption share $\varphi = 0.3$ and the land share in housing production $\alpha = 0.3$, both typical in the literature (e.g., Combes et al., 2021).[51] The non-housing consumption share is $\beta = 0.7$. We set the real interest rate $\delta = 0.04$, consistent with Kenyan data (Henderson et al., 2021), and $\rho = \delta$. The purchase price of capital is $r = 25$, so the rental rate on capital is 1. The rental price of a unit of housing in the city is $p = 2.19$, which from the cost function implies a rental price of a unit of land in the city of 1.8. The 20k areas are appropriately much cheaper in our data, with unit prices ranging from 0.16 to 0.44, consistent with urban land rent gradients in Kenya (Henderson et al., 2021).

Income distribution (Panel B of Table 8). We estimate the household income distribution in Dar es Salaam using the survey of Balboni et al. (2020). We use the monthly net income reported by household heads, converted to 2021 dollars. We approximated the data using log-normal, Pareto,

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

and Gamma distributions, and found that the Gamma has the best fit. The shape and scale parameters are reported in Table 8. Since we model permanent income, we also tried a version in which the Gamma mean ($2,755) was held fixed and the Gini coefficient was reduced by a third in line with the literature (Bönke et al., 2015); the results were broadly similar, but the loss was slightly worse, so we retained the current distribution.

Structurally estimated parameters (Panel C of Table 8). The six parameters (B_F, B_N, θ, μ^{shape}, μ^{scale}, σ) are estimated by Simulated Method of Moments (SMM), minimizing the following weighted loss function:

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

which matches twelve moments: the average land price R_m and the fraction built by 2020 S_m for each of the six plot types, weighted by the relative share of plots of that type.

The global optimization algorithm is Differential Evolution, from the family of Evolutionary Algorithms, which proved very stable.[52] The parameter search bounds are: $(0,1)$ for B_F and B_N; $(0,1)$ for θ; $(0,1)$ for μ^{shape}; $(0,100)$ for μ^{scale}; and $(0,10)$ for σ.

The shape and scale of the μ distribution, θ, and σ are all strictly positive, so we impose a lower bound of zero. While in principle the B's can be negative, even very small negative values cause the loss function to be very large, so we set 0 as a lower bound. A is normalized to 1, and we assume that the city amenity exceeds the 20k amenity at $t = 0$, giving the upper bound of 1 for the B's. θ is a depreciation rate, so we assume it is bounded from above by 1. The wide bounds on μ^{shape} and μ^{scale} let the 20k-vs-city taste distribution be highly right-skewed—most households with little idiosyncratic preference for 20k over the city and a long tail of households with strong 20k preferences.

Parameter estimates. The 20k amenities are $B_F = 0.534$ and $B_N = 0.665$, compared to the undiscounted city amenity $A = 1$. The rate of city amenity deterioration $\theta = 0.008$ implies that the city amenity reaches the Near 20k amenity level in approximately 50 years and the Far level in approximately 77 years. The estimated logit scale is $\sigma = 0.073$; the standard deviation of the plot-specific component, $\sigma\pi/\sqrt{6} \approx 0.09$ utility units, is modest relative to typical magnitudes of V_m, so the plot-specific shock provides moderate noise around the systematic income-driven plot ranking. Moment matches are reported in Table A.12 and are generally close, with the main deviations being prices on small- and medium-size Near plots and the fraction built on small-Far plots.

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

C.5 Welfare decomposition of a change in the plot allocation

We decompose the change in total surplus from any change in plot allocation, such as a split of a larger plot or adding an extra formal 20k plot to the existing supply. Total surplus is the sum of aggregate land value and aggregate compensating variation,

$$\text{TS} = \text{LV} + \text{CV}, \qquad \text{LV}(\mathcal{A}, R) = \sum_m R_m N_m,$$

where $R = (R_1, \ldots, R_M)$ is the vector of plot prices and an allocation $\mathcal{A}$ is a supply vector $(N_1, \ldots, N_M)$ together with an assignment of each household to the city or a plot. We note that while we use equivalent variation as the measure of consumers' surplus in the rest of the paper, for this proof (but not examples) we use compensating variation. While the result holds in all model simulations, the proof with equivalent variation involves an approximation (discussed below in footnote 54), while with compensating variation it is exact.

A household i with time-zero lifetime income W_i[53] and taste vector ν_i chooses the option that maximizes its realized utility. Households choose a "location" j, which is either to remain in the city indefinitely ($j = 0$) or a plot m, a (B, l) combination. Let $x(j, R)$ denote the decision choices (housing and non-housing consumption, capital, timing) that are optimal at location j under price vector R; the within-city choices $x(0)$ comprise only non-housing consumption z and housing h and do not depend on plot prices. Households thus choose

$$j_i^* = \underset{j \in \{0,1,\ldots,M\}}{\arg\max}\, \mathcal{W}_j\big(W_i;\, R,\, j,\, x(j,R)\big), \qquad \mathcal{W}_m(W;\, R,\, m,\, x) = V_m(W, R_m, x) + \nu_m.$$

A household's compensating variation $CV_i^{(R,j,x)}$ is the time-zero income *removable* from the household until it is indifferent to the city,

$$\mathcal{W}_0(W_i) = \mathcal{W}_{j_i^*}\big(W_i - CV_i^{(R,\, j_i^*,\, x(j_i^*,R))};\, R,\, j_i^*,\, x(j_i^*, R)\big), \qquad CV_i \geq 0 \ (= 0 \text{ if } j_i^* = 0).$$

Aggregating over the $N = 800{,}000$ households,

$$\text{CV}(\mathcal{A}, R) = H \iint CV_i\, f_w(W)\, f_\nu(\nu)\, dW\, d\nu.$$

States. Consider a change from an initial equilibrium $0 = (\mathcal{A}^0, R^0)$ to a new equilibrium $1 = (\mathcal{A}^1, R^1)$. Consider also the *fixed-allocation, fixed-choice scenario* $(\mathcal{A}^0, R^1)$, in which every household stays at its equilibrium-0 location j^0 with its equilibrium-0 choices $x(j^0, R^0)$ but faces equilibrium-1 prices R^1. This is not an equilibrium, since households neither re-sort to a different j nor re-

[53] $W_i \equiv \int_0^\infty w_i\, e^{-\rho t}\, dt = w_i/\rho$ is the present discounted value at rate ρ of the household's constant wage stream w_i. The compensating variation defined below is measured in these same time-zero lifetime-income units.

optimize their choices x given the new prices. We mark its objects with a tilde,

$$\widetilde{\mathrm{LV}} \equiv \mathrm{LV}(\mathcal{A}^0, R^1), \qquad \widetilde{CV}_i \equiv CV_i^{(R^1, j^0, x(j^0, R^0))}, \qquad \widetilde{\mathrm{CV}} \equiv \mathrm{CV}(\mathcal{A}^0, R^1),$$

where $\widetilde{CV}_i = 0$ for city dwellers ($j^0 = 0$), and we write the price change as $R_m^1 = R_m^0 - \zeta_m^R$.

Decomposition. The change in total surplus decomposes into a land-value change and a compensating-variation change, and we insert the fixed-allocation, fixed-choice scenario into each:

$$\Delta\mathrm{TS} = \Delta\mathrm{LV} + \Delta\mathrm{CV},$$

$$\Delta\mathrm{LV} = \underbrace{\left(\mathrm{LV}^1 - \widetilde{\mathrm{LV}}\right)}_{\Delta\mathrm{LV}^{PE}} + \underbrace{\left(\widetilde{\mathrm{LV}} - \mathrm{LV}^0\right)}_{\Delta\mathrm{LV}^{GE}}, \qquad \Delta\mathrm{CV} = \underbrace{\left(\mathrm{CV}^1 - \widetilde{\mathrm{CV}}\right)}_{\Delta\mathrm{CV}^{\mathrm{reopt}}} + \underbrace{\left(\widetilde{\mathrm{CV}} - \mathrm{CV}^0\right)}_{\Delta\mathrm{CV}^{R}}.$$

This yields four components

$$\Delta\mathrm{TS} = \Delta\mathrm{LV}^{PE} + \Delta\mathrm{LV}^{GE} + \Delta\mathrm{CV}^{\mathrm{reopt}} + \Delta\mathrm{CV}^{R}. \tag{11}$$

The land-value terms value the supply change at the new prices and the price change on the initial stock,

$$\Delta\mathrm{LV}^{PE} = \sum_m \left(N_m^1 - N_m^0\right) R_m^1, \qquad \Delta\mathrm{LV}^{GE} = \sum_m \left(R_m^1 - R_m^0\right) N_m^0 = -\sum_m \zeta_m^R N_m^0.$$

The price terms cancel exactly. Holding allocations and choices fixed, a change in prices is a pure transfer between landowners and residents, not a creation of surplus, so the two price terms $\Delta\mathrm{LV}^{GE}$ and $\Delta\mathrm{CV}^R$ must offset. Only households on a plot in equilibrium 0 contribute to $\Delta\mathrm{CV}^R$: for a city dweller ($j^0 = 0$) plot prices do not enter the budget, so freezing its allocation and choices leaves its compensating variation unchanged, $\widetilde{CV}_i = CV_i^0 = 0$. We therefore consider a resident of a plot m^0. The budget-constraint identity

$$\mathcal{W}_m(W + a;\, R,\, m,\, x) = \mathcal{W}_m(W;\, R - a,\, m,\, x)$$

says that receiving a more income is welfare-equivalent to paying a less for the plot. Setting $a = \zeta_{m^0}^R$ and $R = R^0$,

$$\begin{aligned} \mathcal{W}_0(W_i) &= \mathcal{W}_{m^0}\left(W_i - \widetilde{CV}_i;\, R^1,\, m^0,\, x(m^0, R^0)\right) \\ &= \mathcal{W}_{m^0}\left(W_i - \widetilde{CV}_i + \zeta_{m^0}^R;\, R^0,\, m^0,\, x(m^0, R^0)\right) = \mathcal{W}_{m^0}\left(W_i - CV_i^0;\, R^0,\, m^0,\, x(m^0, R^0)\right), \end{aligned}$$

where the first line uses the definition of $\widetilde{CV}_i$, the budget identity delivers the second, and the definition of CV_i^0 delivers the third. Monotonicity of $\mathcal{W}_{m^0}$ in income gives

$$\widetilde{CV}_i - CV_i^0 = \zeta^R_{m^0} :$$

the price fall raises this resident's compensating variation by *exactly* the land-value change on its plot. Because the plot price enters the budget constraint linearly, no first-order approximation is involved.[54] Aggregating over the residents of plot m—and noting that $N\iint_{\mathcal{H}^0_m} f_w(W)\, f_\nu(\nu)\, dW\, d\nu$ is the demand $D_m(R^0)$, which equals the supply N^0_m in equilibrium 0—

$$\Delta \text{CV}^R_m = \zeta^R_m\, H \iint_{\mathcal{H}^0_m} f_w(W)\, f_\nu(\nu)\, dW\, d\nu = \zeta^R_m\, D_m(R^0) = \zeta^R_m\, N^0_m,$$

where $\mathcal{H}^0_m = \{(W,\nu) : j^*_i = m \text{ at } (\mathcal{A}^0, R^0)\}$. Summing over m,

$$\Delta \text{CV}^R = \sum_m \zeta^R_m N^0_m = -\,\Delta \text{LV}^{GE}.$$

The reallocation term is non-negative. We show $\Delta CV_i^{\text{reopt}} = CV_i^1 - \widetilde{CV}_i \geq 0$ for every household, treating city dwellers and plot residents in turn.

For a city dweller in equilibrium 0 ($j^0 = 0$), $\widetilde{CV}_i = 0$, so $\Delta CV_i^{\text{reopt}} = CV_i^1 \geq 0$. This is pure reallocation, and it is positive precisely for those city dwellers whom the change draws into a 20k plot: by revealed preference they could have remained in the city, so their new option is at least weakly better.

For a resident of a plot m^0 in equilibrium 0 ($j^0 = m^0 \geq 1$), we split $\Delta CV_i^{\text{reopt}}$ into two non-negative pieces by inserting a second intermediate scenario, in which the household stays on m^0 but re-optimizes its within-plot choices at the new prices R^1: prices R^1, plot m^0, choices $x(m^0, R^1)$. Then

$$\Delta CV_i^{\text{reopt}} = \underbrace{\left[\, CV_i^1 - CV_i^{(R^1, m^0, x(m^0, R^1))} \right]}_{\text{re-optimization of allocation } \geq 0} + \underbrace{\left[\, CV_i^{(R^1, m^0, x(m^0, R^1))} - \widetilde{CV}_i \right]}_{\text{re-optimization of within-plot choices } \geq 0} .$$

The first bracket is non-negative by revealed preference: at prices R^1, household i weakly prefers its equilibrium-1 option to remaining on m^0. The second is non-negative by optimality of the within-plot first-order conditions: $x(m^0, R^1)$ weakly dominates the frozen $x(m^0, R^0)$ on plot m^0 at prices R^1. Aggregating over all households, $\Delta \text{CV}^{\text{reopt}} \geq 0$.

Two features of the reallocation are worth noting. First, both operations we consider—splitting a plot and adding a plot to the supply—expand the total mass of 20k plots and thus draw an

[54] This is the value of using CV over EV. The issue with equivalent variation is that changes in revenue enter $\max_m(V_m(w, R_m; \Theta) + \nu_m) = V_0(w + \Delta; \Theta)$ non-linearly on the LHS, while EV enters linearly on the RHS.

additional mass of households from the city into 20k; these entrants are exactly the city dwellers with $\Delta CV_i^{\text{reopt}} = CV_i^1 \geq 0$. Second, a resident whose plot price rises may prefer to return to the city in equilibrium 1; its price change is a negative ΔCV_i^R—a pure transfer to the landowner, which cancels against ΔLV^{GE}—while its reallocation component stays non-negative, since at the new prices it weakly prefers its equilibrium-1 option (here, the city).

Result. The price terms cancel in (11), leaving

$$\Delta\text{TS} = \Delta\text{LV}^{PE} + \Delta\text{CV}^{\text{reopt}}, \qquad \Delta\text{CV}^{\text{reopt}} \geq 0. \tag{12}$$

The price–size gradient is a sufficient statistic for welfare effects In particular, holding total land area fixed, the optimal allocation will have a weakly positive price-per-square-meter gradient in plot size. We will show this by contradiction. Assume that the optimal allocation has at least one pair of plot types (m, m') with $m = (B, l)$ and $m' = (B, al)$ and $a \in (0,1)$ where $r_m < r_{m'}$, i.e., the gradient is strictly negative across these plot types. A planner could take one plot of type m and split it into $1/a$ plots of type m'. The resulting $\Delta\text{LV}^{PE} = l(r_{m'} - r_m) > 0$, where $r = R/l$, which, in addition to the $\Delta\text{CV}^{\text{reopt}} \geq 0$ result above, contradicts the allocation being at an optimum.

There are two caveats to consider before applying this sufficient statistic in practice. First, a reallocation at the optimum can induce price changes, and a planner assessing an *existing* plot allocation can only observe the existing prices r^0, not the hypothetical post-split prices r^1 that determine ΔLV^{PE}. Nevertheless, a marginal split (e.g., one plot out of tens of thousands) moves equilibrium prices negligibly, and so r^0 and r^1 are approximately equal, and the sufficient statistic holds.[55] Second is that plots are discrete in the real world and $1/a$ may not always take an integer value; however, in such a case the planner could take multiple plots of type m so that they split cleanly into an integer number of type m' plots.[56]

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.

[55] For example, in the case where we split one medium plot of 1200 sqm (m) into two small plots of 600 sqm (m') at our baseline, the general-equilibrium price changes from such a split are about 0.0014% of price levels, so the difference term is of the same negligible order.

[56] That is, for $a = a_1/a_2$ with $a_2 > a_1$ both integers, the planner takes a_1 plots of type m and creates a_2 plots of type m'.

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.

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

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`.

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

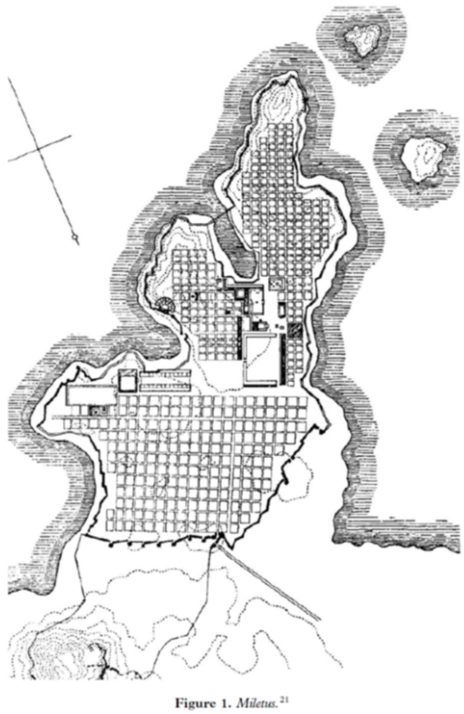
Statista (2022), 'Tanzania: Inflation rate from 1988 to 2027 retrieved on Oct. 9, 2022', `https://www.statista.com`.

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

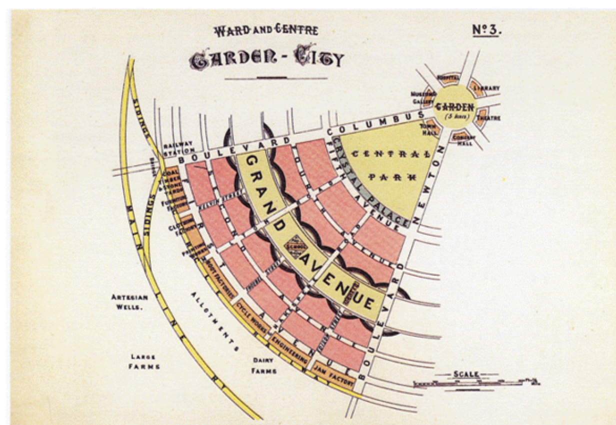
D Appendix Figures

Figure A.1: Historical examples of urban planning

(A) The urban plan of Miletus, Ancient Greece

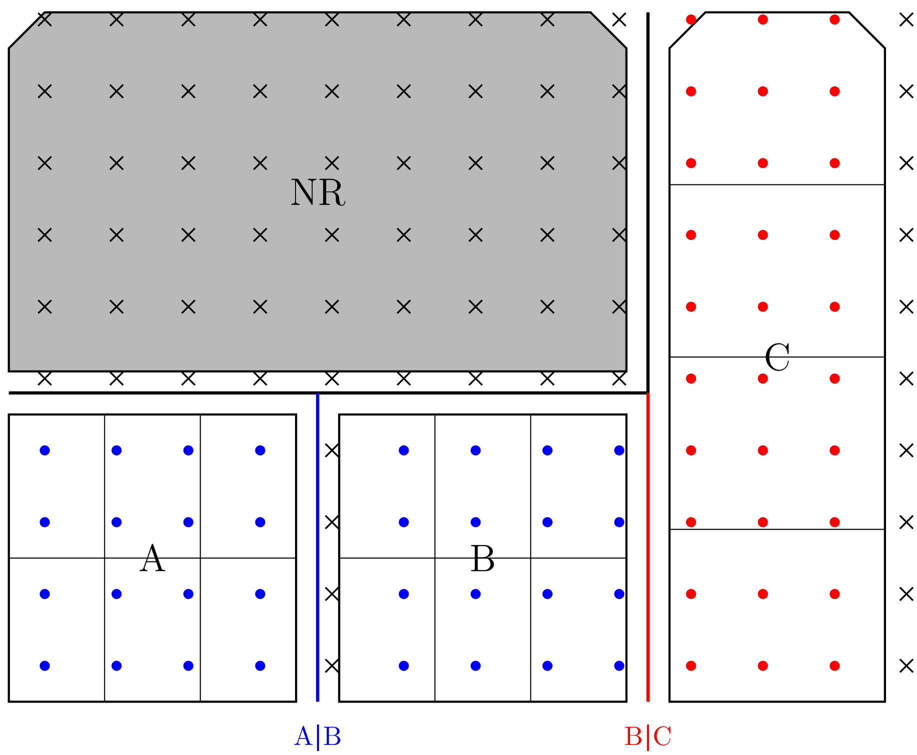


(B) Diagram from Howard (1902) "Garden Cities of To-Morrow"



Notes: This figure plots historical examples of urban planning.

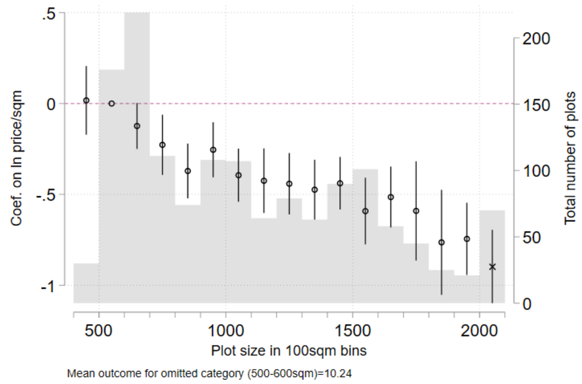
Figure A.2: Diagram of insula, plots, gridcells, and boundaries



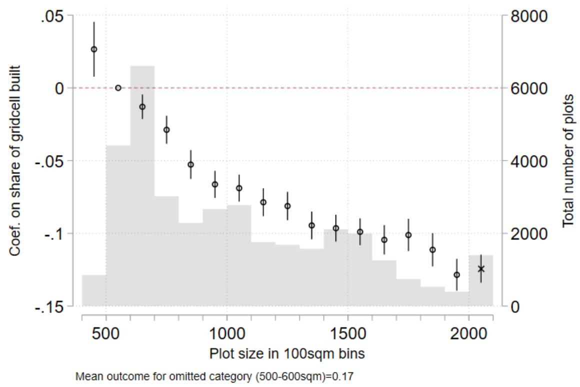
Notes: This figure provides a diagram demonstration of our data construction of insula, plots, gridcells, and boundaries. Plots are denoted by black outlines, and are colored white (residential), and grey (non-residential "NR"). Insulae are made up of the contiguous plots, each with a unique ID (A, B,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

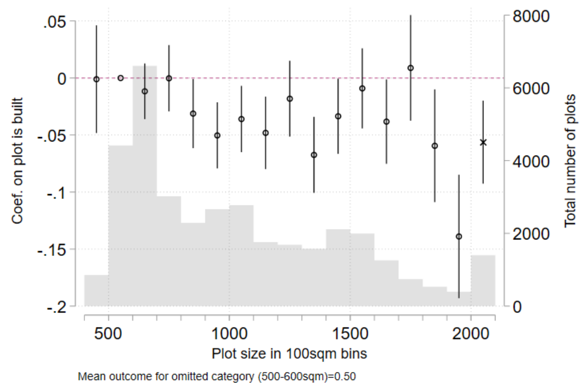
(A) Ln price per sqm



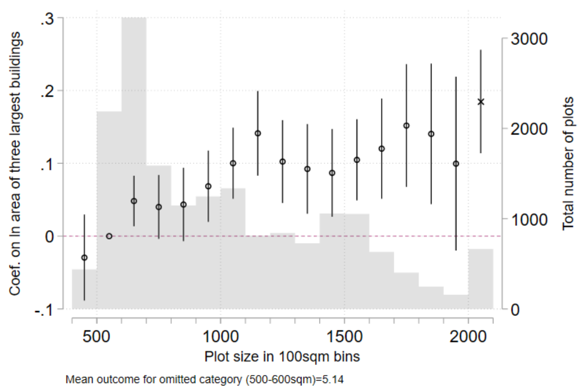
(B) Share of gridcell built



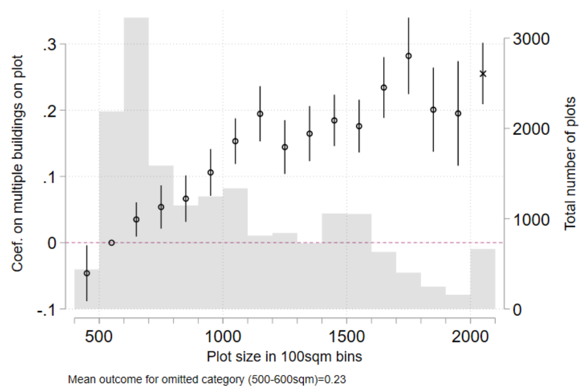
(C) Plot has a building



(D) Ln area of 3 largest buildings



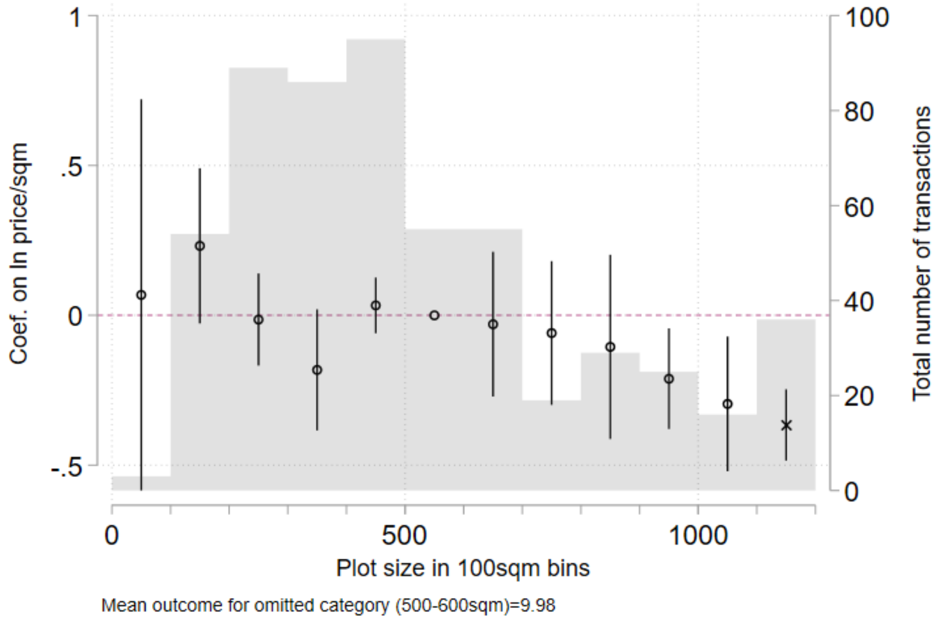
(E) Plot has multiple buildings



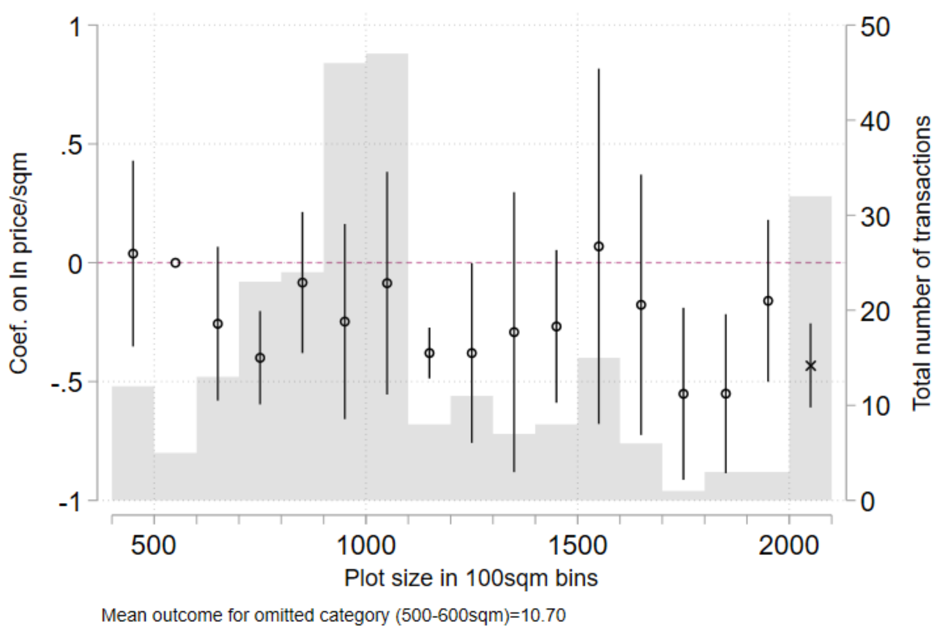
Notes: This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter and building quantity measures on 100sqm plot size bins (plots with size above 2000sqm are pooled into one bin, marked by an 'x'). The omitted bin is 500-600sqm. Outcomes vary by panel. 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)



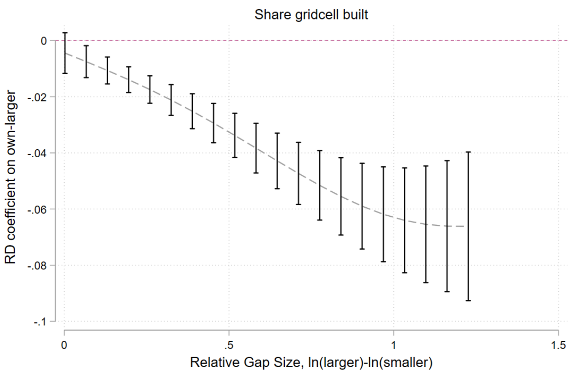
(B) Ln price per sqm (non-20k surveyed)



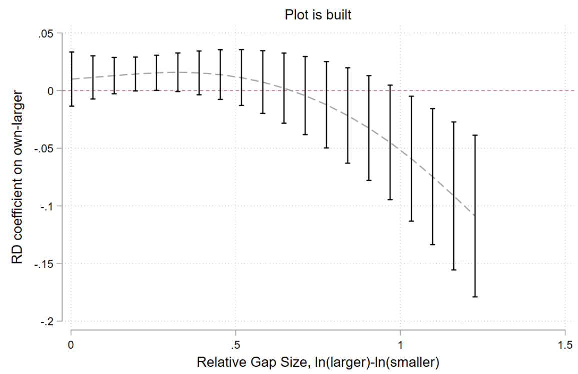
Notes: This figure plots coefficients and their 95% confidence intervals for regressions of log price per square meter in non-20k areas. The omitted bin is 500-600sqm. Samples vary by panel: unsurveyed plots in non-20k areas (a), and surveyed plots in non-20k areas (b). Controls include fixed effects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

Figure A.5: RD coefficients by gap size

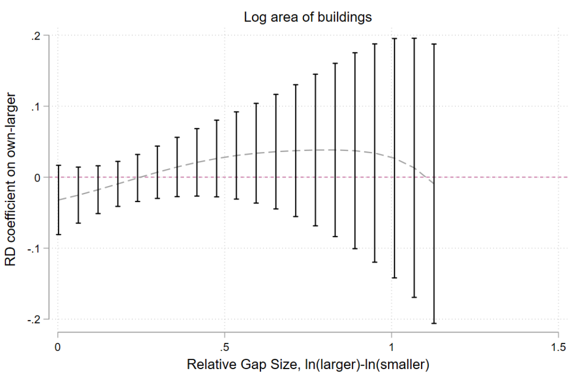
(A) Share gridcell built



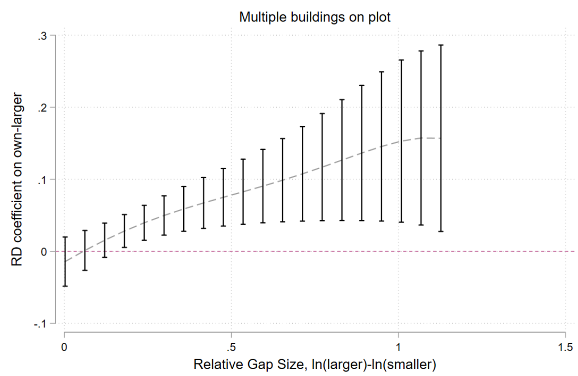
(B) Plot is built



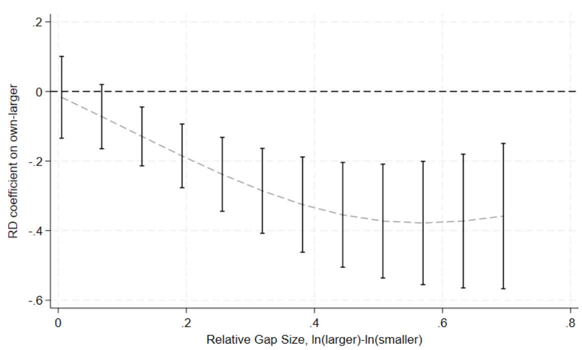
(C) Ln area of buildings



(D) Multiple buildings on plot

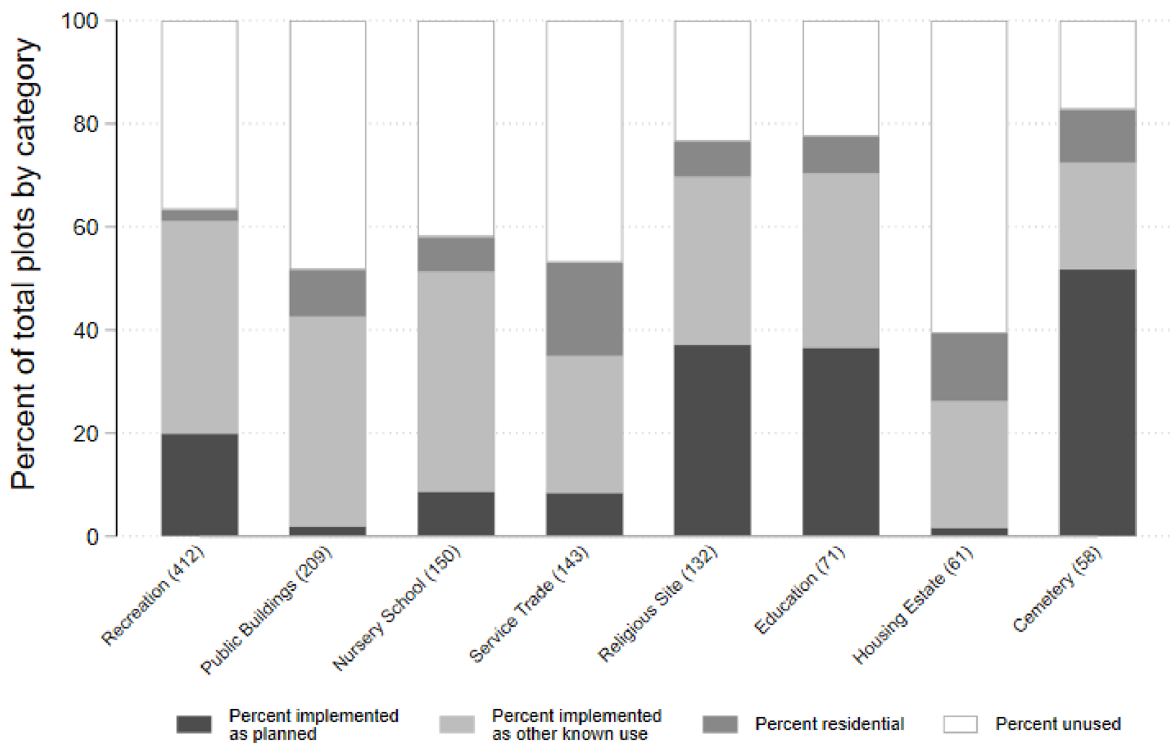


(E) Ln price per sqm



Notes: This figure plots estimated coefficients on the own plot larger dummy following the specification outlined in Eq. 2. 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 2 except the own plot larger dummy), and then run the `vc_pack` semi-parametric approach.

Figure A.6: Planned non-residential broken into implemented uses



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

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 cemetery 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 cemetery.
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 unknown 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: Homogeneity	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
Mwanagati	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: 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 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.5: Balance on natural amenities (RD)

	(1)	(2)	(3)	(4)	(5)
	Elevat.	Rugged.	River or stream in 100m	Wetland in 100m	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 less 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: elevation 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 other three (disamenities) are negative. Controls always include 20K*Mtaa Area and insula-segment FEs. Standard errors in parentheses are clustered by insula.

Table A.6: Optimal bandwidth selection (RD)

	(1) Ln Price per sqm	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) 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.7: 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, 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.8: 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.9: Built outcomes in 20k areas (OLS with amenities and implemented uses)

	(1) Ln Price per sqm	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Ln plot size	-0.51	-0.077	0.0014	0.14	0.19
	(0.055)	(0.0025)	(0.0093)	(0.018)	(0.012)
Dist (km) paved major road	-0.14	-0.015	-0.041	-0.062	-0.039
	(0.031)	(0.0016)	(0.0071)	(0.012)	(0.0087)
Elevation (m)	0.0026	0.00089	0.0028	0.0032	0.00033
	(0.0025)	(0.000099)	(0.00043)	(0.00067)	(0.00048)
Ruggedness	-0.0078	-0.0055	-0.015	-0.0097	-0.0089
	(0.022)	(0.00098)	(0.0039)	(0.0090)	(0.0053)
River/stream 100m	0.055	-0.028	-0.12	-0.067	-0.046
	(0.16)	(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	0.027	0.0029	0.016	0.0051	0.0056
	(0.028)	(0.0013)	(0.0057)	(0.010)	(0.0068)
20k edge in 100m	-0.010	-0.0033	-0.0078	-0.032	0.0096
	(0.044)	(0.0023)	(0.0095)	(0.016)	(0.011)
Impl. recreation in 100m	-0.019	0.0064	0.0026	0.034	0.0043
	(0.051)	(0.0031)	(0.011)	(0.019)	(0.013)
Impl. nursery school in 100m	0.12	0.0074	0.037	0.000049	0.018
	(0.088)	(0.0072)	(0.027)	(0.039)	(0.026)
Impl. religious site in 100m	0.094	0.011	0.028	0.012	0.013
	(0.059)	(0.0043)	(0.016)	(0.027)	(0.018)
Impl. education in 100m	0.071	0.00012	0.0018	-0.027	-0.0015
	(0.10)	(0.0044)	(0.015)	(0.027)	(0.017)
Impl. service trade in 100m	0.14	0.016	0.050	0.053	0.058
	(0.077)	(0.0053)	(0.019)	(0.029)	(0.023)
Impl. housing estate in 100m	-0.77	0.051	0.20	0.17	0.19
	(0.19)	(0.011)	(0.062)	(0.094)	(0.099)
Impl. public building in 100m	-0.17	-0.0061	-0.034	-0.013	0.083
	(0.22)	(0.010)	(0.040)	(0.060)	(0.045)
Impl. cemetery in 100m	0.20	0.0025	0.021	-0.021	0.031
	(0.17)	(0.0054)	(0.018)	(0.034)	(0.024)
Impl. other non-res in 100m	0.36	-0.0050	-0.032	0.0057	0.0089
	(0.13)	(0.0076)	(0.025)	(0.042)	(0.025)
Impl. farming in 100m	-0.019	-0.00012	0.0069	-0.011	-0.0035
	(0.045)	(0.0023)	(0.0087)	(0.015)	(0.010)
Impl. as residential in 100m	0.14	-0.0033	-0.0016	-0.013	-0.0051
	(0.052)	(0.0034)	(0.014)	(0.023)	(0.017)
Unused, kept in 100m	-0.083	0.0030	0.00053	-0.021	-0.0021
	(0.048)	(0.0026)	(0.010)	(0.016)	(0.012)
Unused, unkept in 100m	0.056	-0.0085	-0.035	-0.042	-0.023
	(0.042)	(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	10	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 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 Mtaa*20k 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. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table A.10: Prices and built outcomes on plot sizes (super-insula RD, control frac built)

	(1) Ln Price per sqm	(2) Share gridcell built	(3) Plot is built	(4) Log area of buildings	(5) Multiple buildings on plot
Own Larger	0.023	-0.0020	-0.0018	-0.00049	0.024
	(0.061)	(0.0023)	(0.0094)	(0.020)	(0.015)
Own Smaller × Dist. (km)	0.19	0.031	0.13	-0.068	-0.016
	(0.44)	(0.015)	(0.057)	(0.12)	(0.083)
Own Larger × Dist. (km)	0.54	-0.036	0.0052	0.094	-0.050
	(0.35)	(0.017)	(0.062)	(0.13)	(0.088)
Ln plot size	-0.52	-0.066	0.026	0.18	0.20
	(0.088)	(0.0031)	(0.012)	(0.027)	(0.018)
Share built in 100m	0.15	0.11	0.43	0.13	0.15
	(0.12)	(0.0060)	(0.024)	(0.050)	(0.033)
Mean Outcome	10	0.11	0.49	5.3	0.38
N (gridcells)	3962	93021	93021	45711	45711
N (plots)	1405	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 1-5 the outcomes are the same 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. Column 1 (prices) additionally controls for transaction period by source (dalali or occupier survey) FEs. Standard errors in parentheses are clustered by insula.

Table A.11: 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.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	26.50
Far, Medium	4192	R_2	3272	3271	$\%(\tau_2 < 15)$	30.75	30.74
Far, Large	938	R_3	4571	4576	$\%(\tau_3 < 15)$	27.50	27.51
Near, Small	10068	R_4	6776	4897	$\%(\tau_4 < 15)$	59.42	59.42
Near, Medium	12661	R_5	10906	8057	$\%(\tau_5 < 15)$	54.74	54.72
Near, Large	3381	R_6	12478	13148	$\%(\tau_6 < 15)$	54.63	54.62

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 equation above, take the anti-log, convert to USD, and use the real interest rate to discount from 2021 back to 2005.

Table A.13: Distribution of 20k households and consumer-surplus gains across income deciles

	Share of 20k households			Share of consumer surplus (EV)		
Decile	Baseline	W-max	LV-max	Baseline	W-max	LV-max
1	0.1%	2.5%	0.0%	0.0%	0.2%	0.0%
2	3.9%	6.4%	0.0%	0.4%	1.5%	0.0%
3	6.9%	7.2%	1.9%	1.6%	2.9%	0.2%
4	8.3%	7.5%	6.4%	3.2%	4.3%	1.2%
5	9.2%	8.0%	9.5%	5.2%	6.0%	3.2%
6	9.7%	9.9%	11.3%	7.5%	8.1%	5.7%
7	10.5%	13.1%	13.1%	10.6%	11.0%	9.4%
8	11.9%	14.6%	14.4%	14.5%	14.4%	14.5%
9	16.2%	15.0%	16.2%	20.7%	19.3%	22.7%
10	23.2%	15.8%	27.0%	36.3%	32.4%	43.1%
Total	100%	100%	100%	100%	100%	100%

Notes: Distribution of the 20k project across the city-wide income distribution, under the actual allocation (Baseline), welfare maximization (W-max), and land-value maximization (LV-max). The total number of potential households is 800,000, so each decile contains 80,000 households. The first three columns report each decile's share of all households who take a 20k plot; the last three report each decile's share of the aggregate consumer surplus (equivalent variation), which totals $249.2M under the Baseline, $277.2M under W-max, and $202.0M under LV-max. The annual income of the median household in each decile is $203, $555, $919, $1,317, $1,769, $2,299, $2,951, $3,810, $5,094, and $7,811, respectively. Gains accrue only to households that take a 20k plot; they exclude any general-equilibrium benefit to households that remain in the city (e.g. through lower rents), so the proportional gains may overstate regressivity at the bottom.

Table A.14: 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).