

# Planning Ahead for Better Neighborhoods: Long-Run Evidence from Tanzania

---

Guy Michaels

*London School of Economics*

Dzhamilya Nigmatulina

*London School of Economics*

Ferdinand Rauch

*University of Oxford and University of Graz*

Tanner Regan

*London Business School*

Neeraj Baruah

*Vivid Economics*

Amanda Dahlstrand

*London School of Economics*

Africa's demand for urban housing is soaring, even as it faces a proliferation of slums. In this setting, can modest infrastructure investments in greenfield areas where people subsequently build their own houses

We thank Lino Ferreira and Ilia Samsonov for excellent research assistance; Richard Bakubiye, Chyi-Yun Huang, Ezron Kilamhama, George Miringay, Hans Omary, Elizabeth

Electronically published May 24, 2021

[*Journal of Political Economy*, 2021, vol. 129, no. 7]

© 2021 by The University of Chicago. All rights reserved. 0022-3808/2021/12907-0005\$10.00

facilitate long-run neighborhood development? We study Sites and Services projects implemented in seven Tanzanian cities during the 1970s and 1980s, and we use a spatial regression discontinuity design to compare greenfield areas that were treated (*de novo*) with nearby greenfield areas that were not. We find that by the 2010s, *de novo* areas developed into neighborhoods with larger, more regularly laid-out buildings and better-quality housing.

## I. Introduction

Africa's cities are growing rapidly. With the continent's expanding population (UN 2015) and rising urbanization rate (Freire, Lall, and Leipziger 2014), we expect that almost a billion people will join Africa's cities by 2050. But many of these cities, especially in sub-Saharan Africa, already face problems of poor infrastructure and low-quality housing (Henderson et al. 2016; Castells-Quintana 2017). According to UN-Habitat (2012), as many as 62% of this region's urban dwellers live in slums, whose population was expected to double within 15 years. The poor living conditions in those slums have important consequences for residents' lives (Marx, Stoker, and Suri 2013).

There are various policy options for addressing the immense challenges posed by African urbanization. One option, which is often the default, is to allow neighborhoods to develop organically without much planning or infrastructure. At the other end of the spectrum, a second option is for the state to not only plan but actually build public housing. This is expensive, but has been done, for example, in South Africa (Franklin 2020). Between these two alternatives lies a third option of laying out basic infrastructure on the fringes of cities and allowing people to build their own homes; this option has been advocated by Romer (2012) and Angel (2012). A fourth option is to improve infrastructure in areas where low-quality housing develops.

---

Talbert, Office of the President of the Regional Administration and Local Government of Tanzania, and especially Charles Mariki for help in obtaining the data. For helpful and constructive comments, we thank the editor, Magne Mogstad, and six anonymous referees, as well as Shlomo Angel, Julia Bird, Paul Collier, Matt Collin, Gilles Duranton, Simon Franklin, Ed Glaeser, Michael Greenstone, Vernon Henderson, Wilbard Kombe, Sarah Kyessi, Somik Lall, Joseph Mukasa Lusugga Kironde, Amulike Mahenge, Alan Manning, Anna Mtani, Ally Hassan Namangaya, Steve Pischke, Shaaban Sheuya, Tony Venables, and Sameh Wahba. We also thank participants at the World Bank Annual Bank Conference on Africa in Oxford; World Bank Land and Poverty Conference 2016 in Washington, DC; World Bank Conference on Spatial Development of African Cities in Washington, DC; and workshop and seminar participants at the London School of Economics, Queen Mary University of London, and Stanford University. We gratefully acknowledge the generous support of a Global Research Program on Spatial Development of Cities, funded by the World Bank's Multi-Donor Trust Fund on Sustainable Urbanization and supported by the UK Department for International Development. The usual disclaimer applies. Data are provided as supplementary material online.

Understanding the implications of these options is important for current policy discussions. For example, there are debates about the respective merits of upgrading and starting anew (e.g., Duranton and Venables 2020). But we know relatively little about these options' implications for private investments and the survival of infrastructure or about their distributional consequences. One of the main contributions of this paper is to shed light on these issues. To do so, we study the long-run development of neighborhoods that were part of the Sites and Services projects (described below).<sup>1</sup> We look at neighborhoods in not only Tanzania's biggest city, Dar es Salaam, but also in six of its secondary cities.<sup>2</sup>

Our paper focuses on the long-run consequences of the third option discussed above (*de novo*) compared with the first (default) option of unregulated development. We study *de novo* neighborhoods, which were developed in greenfield areas on what were then the fringes of Tanzanian cities. The developments included the delineation of formal residential plots and the provision of basic infrastructure, consisting primarily of roads and water mains. People were then offered an opportunity to build homes on these plots in exchange for a fee. To provide a counterfactual, we use nearby control areas that were also greenfields before the Sites and Services projects began. We also provide descriptive evidence on the fourth approach discussed above by studying upgrading areas, which received infrastructure investments similar to the *de novo* areas but only after people had built low-quality housing there.<sup>3</sup> We compare these upgrading areas with nearby areas and, for Dar es Salaam only, with slums that were not upgraded.

We investigate how different neighborhoods developed over more than three decades, and we ask a number of questions. First, do *de novo* investments solve coordination failures and facilitate neighborhood development in the long run? Second, how do they shape private housing investments and the survival of public infrastructure? And, finally, what characterizes the sorting of owners and residents across neighborhoods, and to what extent can owners' sorting account for the differences in outcomes across neighborhoods?

Concretely, we study Sites and Services projects, which were cofunded by the World Bank and the Tanzanian government and were similar to projects carried out in other countries. In Tanzania, they were implemented in

<sup>1</sup> Throughout the paper, we refer interchangeably to "areas" and the "neighborhoods" that develop in them; "houses" and "housing units"; and "squatter settlements" and "slums." Finally, we refer to "owners" as those with *de facto* rights to reside in a house or rent it out. Legally, even formal ownership consists of a long and renewable lease from the state.

<sup>2</sup> This is important because Africa's secondary cities are relatively understudied, despite being home to the majority of its urban population. See, e.g., Tanzania National Bureau of Statistics (2011), NOAA (2012), and Brinkhoff (2017).

<sup>3</sup> Unlike *de novo* areas, however, upgrading areas did not receive formal plots.

two rounds: one began in the 1970s and the other in the early 1980s. Altogether, 12 *de novo* neighborhoods and 12 upgrading neighborhoods were developed in Dar es Salaam, Iringa, Morogoro, Mbeya, Mwanza, Tabora, and Tanga (World Bank 1974a, 1974b, 1977a, 1977b, 1984, 1987).<sup>4</sup>

To study the consequences of *de novo* investments, we combine high-resolution spatial imagery on all seven cities and building-level survey data on three of the cities with historical imagery and maps. We analyze these data using a spatial regression discontinuity (RD) design. We find that in *de novo* areas, houses are larger and more densely and regularly laid out, are better connected to electricity, and (in some specifications) have better sanitation. A family of outcomes index and a hedonic measure of house values show that *de novo* areas have higher-quality housing. These results, which are robust to the inclusion of various controls and robustness checks, demonstrate the crowding in of private investment in response to the public *de novo* infrastructure investments. We also find that *de novo* areas have better access to roads and water mains, reflecting the persistence of the Sites and Services infrastructure investments over several decades.

To shed light on the mechanisms that underlie our findings, we develop a simple model of owners' investment decisions that features complementarity between public and private investments. In *de novo* neighborhoods, where a sufficient fraction of owners can invest in housing quality, infrastructure investment crowds in private investment in housing quality, which in turn preserves infrastructure quality. This virtuous feedback, however, does not occur in upgrading areas, because the existing stock of (low-quality) housing disincentivizes wholesale reconstruction of housing and because owners' credit constraints prevent them from investing sufficiently, and as a result, infrastructure deteriorates. At the same time, in control areas, the infrastructure investments are lower, so no high-quality housing is built.

The model helps us interpret our empirical findings in two important ways. First, it allows us to separate the roles of owners' different credit constraints from the effect of infrastructure investments when comparing *de novo* and control areas. In practice, we find that adding owner fixed effects reduces the quality differences between *de novo* and control areas by up to one-third, but these differences remain large and precisely estimated. Second, the model allows us to infer land value differences across neighborhoods from differences in housing quality. Our calculations suggest that the local gains in land value from *de novo* were, at least in Dar es Salaam, no less than \$75–\$100 per square meter plot

<sup>4</sup> Until recent years, the government maintained sole authority for creating new formal plots in Tanzania, so we cannot study the long-run consequences of privately provided plots.

(in 2017 prices). These gains far exceed the costs of the project, which amounted to no more than \$8–\$13.

In our empirical analysis, we also use census microdata to characterize the sorting of residents across neighborhoods. We find that as of 2012, *de novo* neighborhoods attracted better-educated residents, who likely had higher incomes to pay for better amenities. The sorting on education across neighborhoods is, however, only partial: about 45% of adults in *de novo* areas had no more than a primary school education. Furthermore, even less educated people who initially owned *de novo* plots and eventually sold them likely gained from some of the land value appreciation.<sup>5</sup>

In contrast to our findings on *de novo* areas and in line with our model, our descriptive analysis of upgrading areas suggests that their housing quality is either similar to that of nearby areas or nonupgraded slums or, in some cases, even worse. Our findings also suggest that upgrading areas do not enjoy better access to water mains or roads than the control areas, so the Sites and Services investments in these areas likely deteriorated. These results should be interpreted cautiously, however, since it is harder to find a clean counterfactual for upgrading areas (which were populated to begin with) than for *de novo* areas.

The economic evaluation of *de novo* Sites and Services areas is thus the focal point of our paper. Previous studies of Sites and Services around the world include surveys (e.g., Laquian 1983) and critical discussions (e.g., Mayo and Gross 1987; Buckley and Kalarickal 2006). In the Tanzanian context, there are descriptive studies of Sites and Services in Dar es Salaam (Kironde 1991, 1992; Owens 2012). Other work on Dar es Salaam studies different interventions, including the short-term impact of more recent slum upgrading projects on health, schooling, and income (Coville and Su 2014); descriptive analyses of a more recent episode of serviced plot provision, known as the 20,000 Plots project, which suggests sizable short-run gains in land values (Tiba et al. 2005; Kironde 2015); and willingness to pay for land titling in poor neighborhoods (Ali et al. 2016; Manara and Regan 2019). But as far as we are aware, ours is the first long-run econometric evaluation of *de novo* Sites and Services areas.

Our study is related to research on the role of coordinating land institutions (Libecap and Lueck 2011)—in our case, formal plots—in underpinning economic development. It is also related to studies of housing externalities in cities (Rossi-Hansberg, Sarte, and Owens 2010; Hornbeck and Keniston 2017). Another recent and related paper—on Indonesia rather than Tanzania—is by Harari and Wong (2017). They, like us, find that upgraded slums do not perform well economically in the long run.

<sup>5</sup> As we discuss below, a few years after Sites and Services were implemented, most of the residents in *de novo* neighborhoods in Dar es Salaam were still those targeted by the policy, many of whom were poor.

Our paper, however, differs from theirs, since we focus on *de novo* neighborhoods, which are not part of the context they study.

Our paper is also related to the literature on the economics of African cities (Freire, Lall, and Leipziger 2014). Like Gollin, Jedwab, and Vollrath (2016), we study not only the largest African cities (e.g., Dar es Salaam in Tanzania) but secondary cities as well. Our contribution to this literature comes from studying these cities at a fine spatial scale, examining individual neighborhoods and buildings, using a combination of very high-resolution daylight satellite images, building-level survey data, and precisely georeferenced census data.

A few recent papers study outcomes not only across African cities but also within them (see, e.g., Henderson et al. 2016). Our study differs not only in our focus on secondary African cities but also in the longer time horizon we cover. We use historical satellite images and highly detailed maps going back more than 50 years, which allow us to evaluate long-run changes on historically undeveloped land in response to specific infrastructure investments. By combining these with data on individuals, we also provide more evidence on sorting across neighborhoods.

Also related to our paper is a broader literature on the economics of slums (e.g., Castells-Quintana 2017; Marx, Stoker, and Suri 2019). Our contribution to this literature is to illustrate conditions under which housing of better quality forms and persists, and the limitations of upgrading existing slums. Poor neighborhoods have also been studied in other settings, especially in Latin America and South Asia. For example, Field (2005) and Galiani and Schargrodsky (2010) find that providing more secure property rights to slum dwellers in Latin America increases their investments in residential quality.<sup>6</sup> Our paper differs in its setting (Tanzania is considerably poorer than Latin America) and its focus on early infrastructure provision.

While our paper's focus is on new neighborhoods rather than new cities, it is also related to the work of Romer (2010), who investigates the potential for new charter cities as pathways for urban development in poor countries. Our work is also related to the position advocated by Angel (2012), that Sites and Services may be a relevant model for residential development in some circumstances.<sup>7</sup>

Methodologically, we contribute to the nascent literature using very high-resolution daylight images (e.g., Jean et al. 2016). Like Marx, Stoker, and Suri (2019), we study roof quality as a measure of residential quality. Our

<sup>6</sup> In another paper, Galiani et al. (2013) study an intervention that provides prefabricated homes at a cost of around \$1,000 each in Latin America but that come without any infrastructure.

<sup>7</sup> For a discussion of the idea, see, e.g., this interview with Shlomo Angel: <http://www.smartcitiesdive.com/ex/sustainablecitiescollective/conversation-dr-shlomo-angel/216636/>.

measure of quality differs, however; instead of measuring luminosity, we assess whether roofs are painted, since paint protects roofs from rust. We also use the imagery data to develop a set of measures of residential quality, including building size, access to roads, and a measure of regularity of neighborhood layout, which we combine with survey data on building quality.

The remainder of our paper is organized as follows. Section II discusses the institutional background and data we use; section III presents the research design and our empirical findings; section IV contains a model of investments in infrastructure and housing in different neighborhoods; and section V concludes.

## II. Institutional Background and Data

### A. *Institutional Background*

#### 1. What Were Sites and Services Projects?

This paper studies the long-term consequences of ambitious projects that were designed to improve the quality of residential neighborhoods in Tanzania. These projects, called Sites and Services, formed an important part of the World Bank's urban development strategy during the 1970s and 1980s. Sites and Services projects were implemented not only in Tanzania but also in other countries such as Senegal, Jamaica, Zambia, El Salvador, Peru, Thailand, and Brazil (Cohen, Madavo, and Dunkerley 1983). Of the World Bank's total shelter lending of \$4.4 billion (in 2001 US dollars) from 1972 to 1986, Sites and Services accounted for almost 50%, and separate slum upgrading accounted for more than 20%.

In Tanzania, Sites and Services were implemented in two rounds—the first began in the 1970s (World Bank 1974b, 1984) and the second in the 1980s (World Bank 1977b, 1987). These projects were cofinanced by the World Bank and the Tanzanian government (World Bank 1974a, 1977a).

Sites and Services projects in Tanzania fell into two broad classes. The first involved de novo development of previously unpopulated areas. The second involved upgrading of preexisting squatter settlements (slum upgrading). In total, across both rounds, the program laid the groundwork for 12 de novo neighborhoods and 12 upgrading neighborhoods spread across seven cities (World Bank 1974b, 1977b, 1984, 1987).

The overall cost of the Sites and Services projects in Tanzania was approximately \$130 million (in 2017 US dollars), of which \$83 million were direct costs, covering infrastructure, land compensation, equipment, and consultancy (World Bank 1974b, 1977b, 1984, 1987).<sup>8</sup>

<sup>8</sup> The remainder of the costs covered a loan scheme and community buildings.

The direct costs per square meter in the first round in *de novo* (\$2.20) and upgrading (\$2.37) were similar (World Bank 1974b, 1977b, 1984, 1987). To compare these costs with present-day land values (see below), we focus on costs per square meter of plot, excluding public areas. As we explain in the appendix, we estimate that the direct costs per square meter of plot were no more than \$8, and the total costs were no more than \$13 per square meter.

## 2. What Were the Treatment and Counterfactual?

Our main empirical analysis compares *de novo* (our main treatment) with nearby control areas (our counterfactual). As we explain in section III, we implement a spatial RD design, focusing on the difference in outcomes close to the boundary of *de novo* areas and adjacent control areas, which were (like *de novo*) unbuilt before the Sites and Services projects began. In section III, we also discuss and address potential threats to our identification strategy. Here we explain why we focus on the comparison between *de novo* and control areas and what we learn from it.

*De novo* areas received roads, which were mostly unpaved, and water mains, as well as formal plots.<sup>9</sup> The combination of these three infrastructure elements (formal plots, roads, and water mains) constitutes the main treatment for *de novo* areas.<sup>10</sup> Roads reduce travel costs for both work and leisure for residents, customers, and visitors. Water mains may improve the quality and reliability of water consumed and reduce the transaction costs of purchasing water (e.g., from water trucks). They may also improve the residents' health and help them grow food. Formal plots reduce the risk of full expropriation and infringements onto parts of owners' plots and public spaces (e.g., roads and areas required to maintain water mains). They may also reduce conflicts over ownership and the need to engage in costly defensive actions (e.g., building fences or walls). Moreover, formal and regular plots may mitigate coordination problems, lead to easier access and better use of space, and make plots more easily tradeable, increasing the incentives to invest in them. Long-term gains in land values from a regular grid of plots (vs. a decentralized and irregular

<sup>9</sup> Formal plots are delineations of land, which meet local surveying and town planning standards. They increase tenure security and are a prerequisite in any application for a Certificate of Right of Occupancy (the highest land tenure document in Tanzania).

<sup>10</sup> Upgrading areas also received roads and water mains but no formal plots. The appendix contains more information about the precise timing and details the investments' cost breakdown. The second round investments were generally lower. In some cases, they may have excluded water mains; and for one of the *de novo* areas (the one in Tanga), we have some uncertainty as to the extent of infrastructure that was actually provided (World Bank 1987). Most of the *de novo* plots were, however, laid out in the first round.



system) have been documented in the US context (Libecap and Lueck 2011).<sup>11</sup>

In addition to the main treatment components, both *de novo* and upgrading areas received a small number of public buildings, which were designated as schools, health clinics, and markets.<sup>12</sup> While these could have had an impact, we think that they matter less than the plots, the roads, and the water mains. First, the total cost of the public buildings was lower than either the roads or the water mains; and second, even if Sites and Services areas received more buildings than other areas (and we do not know that they did), there is no evidence that access to them ends discontinuously at the project boundaries. In addition to the infrastructure investments, some Sites and Services residents were offered loans, which were not fully repaid. We think of these loans as relaxing some owners' budget constraints, and below we explain our strategy for studying the implications of differences across neighborhoods in owners' credit constraints.

As we discuss in more detail in the appendix, control areas appear to have received significantly less infrastructure investments, although our data do suggest that they have some roads and connections to water mains.

For upgrading areas, we do not have a clean counterfactual, because those areas were built on by squatters before Sites and Services began. Thus, any present-day differences between them and other areas may reflect a combination of preexisting differences and the treatment effect of upgrading. In section III, we explain what we can nevertheless learn about those areas, at least descriptively, by comparing them with nearby areas or other preexisting squatted areas that were not treated by Sites and Services.

### 3. How Were Treatment Areas Selected?

While our RD design helps to mitigate concerns about selection of areas, it is nonetheless important to explain how the locations of the treatment and control areas were selected. For *de novo* neighborhoods, the planners intended to purchase mostly empty (greenfield) land parcels measuring at least 50 ha each, although in practice this criterion appears to have been met only for seven of the 12 *de novo* areas. The planners also sought land suitable for construction (e.g., with natural drainage) with access to off-site water mains, trunk roads, and employment opportunities. For upgrading, the planners looked for squatter settlements that were large, well defined, hazard safe, and suitable for infrastructure investments (World Bank 1974a, 1977b). In most but not all cities, *de novo* and upgrading areas were

<sup>11</sup> Hornbeck and Keniston (2017) similarly emphasize that starting afresh can lead to higher local land values in an urban setting.

<sup>12</sup> The first-round public buildings were also surrounded by street lighting.

adjacent to each other. We discuss our selection of the control areas in more detail below. All the areas are depicted in figure A1 (figs. A1–A3 are available online).

#### 4. Who Took Part in the Program?

Another important aspect of the Sites and Services projects was the characteristics of the population they targeted. The planners had intended for the plots to be allocated following a point system, which prioritized applicants who met certain criteria. Different sources do not agree precisely on the criteria used, although it seems that a preference was given to the poor—but not the poorest—urban residents (World Bank 1974a, 1977b). Laquian (1983) explains that the *de novo* projects in Tanzania were intended for income groups between the 20th and 60th income percentile of a country. In a similar vein, Kironde (1991) argues that eligibility for *de novo* sites in Dar es Salaam excluded the poorest and richest households but targeted an intermediate range of earners that covered more than 60% of all urban households. It seems that the opportunity to purchase *de novo* plots was initially given to low-income households, including those displaced from upgrading areas, presumably as a result of building new infrastructure (World Bank 1984; Kironde 1991).<sup>13</sup>

There is some disagreement as to how this process was implemented. One report (World Bank 1984) argues that there were irregularities that allowed some richer households to sort into *de novo* neighborhoods. But in discussing the *de novo* sites in Dar es Salaam in the late 1980s, Kironde (1991) argues that most plots were awarded to the targeted income groups. As of the late 1980s, he wrote, “The majority of the occupants (57.9%) are still the original inhabitants but there are many ‘new’ ones who were either given plots after the original awardees had failed to develop them, or who were given ‘created’ plots. A few, however, obtained plots through purchase or bequeathment” (36). Taken together, the evidence suggests that *de novo* locations attracted some households with modest means but also richer ones over time. This type of sorting would likely have occurred even if the project had been administered flawlessly.

#### 5. How Relevant Are Sites and Services Today?

The difficulty of recouping Sites and Services costs and criticism that they excluded the poorest urban population appear to have motivated

<sup>13</sup> The planners had intended for the plots to be allocated following a point system, which prioritized applicants who met certain criteria. But different sources (e.g., World Bank 1974a, 1977b; Kironde 1991) differ in their accounts of what these precise criteria were.

a shift away from them during the 1980s (Mayo and Gross 1987; World Bank 1987; Buckley and Kalarickal 2006). As a result, the share of Sites and Services (including slum upgrading) in the World Bank's shelter lending fell from around 70% in 1972–86 to around 15% in 1987–2005 (Buckley and Kalarickal 2006).

Nevertheless, Sites and Services projects deserve renewed attention for several reasons. First, as mentioned above, Africa's urban population is growing rapidly, adding pressure to its congested cities. Second, Africa's gross domestic product per capita has grown in recent decades, so more Africans can now afford better housing, so an important question is how to deliver this. Alternative solutions, such as government provision of public housing, are considerably more expensive than a *de novo* approach of the type we study.<sup>14</sup> Third, cost recoupment and administration have since improved through increased use of digital record keeping, as evidenced by the Tanzanian Strategic Cities Project (TSCP; World Bank 2013).<sup>15</sup> For example, the 20,000 Plots project, a *de novo* program implemented in Tanzania in the early 2000s, appears to have reduced the cost per plot by about half compared with the historical Sites and Services projects, even though the new plots were bigger (Tiba et al. 2005). Finally, land on the fringes of Tanzanian cities remains inexpensive (Tanzania Ministry of Lands 2012), so there are still opportunities for more *de novo* developments.<sup>16</sup>

To shed light on the motivations of urban planners in considering *de novo* projects, we turn to the abovementioned 20,000 Plots project. Among the concerns backgrounding this program were the ongoing expansion of unplanned squatter areas, which suffer from poor waste management, an inadequate supply of urban services and infrastructure, and transportation problems. These unplanned areas also hamper the government's ability to collect tax revenues (Tiba et al. 2005). It is in this context that the 20,000 Plots project aimed to alleviate the shortage of surveyed and serviced plots and to reduce the rapid increase of informal settlements, as well as to restrict land speculation and corruption (Tiba et al. 2005). At the same time, distributional concerns regarding *de novo* projects remain relevant (Kironde 2015), and we revisit those in section V.

<sup>14</sup> According to correspondence with Simon Franklin, from the experience of housing programs in cities such as Addis-Ababa, four-room apartments (with a bathroom) in five-story buildings entail construction costs of around \$10,000, plus a further \$3,000–\$4,000 for infrastructure and administration. This figure excludes land costs.

<sup>15</sup> The TSCP was approved by the World Bank in May 2010 (see <http://projects.worldbank.org/P111153/tanzania-strategic-cities-project?lang=en>).

<sup>16</sup> Even cheap land on the city fringes is likely to have some residents, however, and ensuring that *de novo* programs treat them inclusively is an important issue, which we revisit in our conclusion.

### *B. Data Description*

This section outlines how we construct the data sets that we use in our empirical analysis, leaving further details to the appendix. First, we explain how we measure the treatment and control areas. Second, we explain our choice of units of analysis. Third, we explain how we construct the variables that we use in our analysis. Finally, we discuss summary statistics for key outcomes.

#### 1. How Do We Measure Treatment and Control Areas?

For five of the seven Sites and Services cities (Dar es Salaam, Iringa, Tabora, Tanga, and Morogoro), we have maps showing the program area boundaries (World Bank 1974a, 1974b, 1977a, 1977b, 1984, 1987). For the two remaining cities, we use information from local experts (for Mbeya) and other historical maps (for Mwanza), as we explain in the appendix. Tables A1 and A2 (tables A1–A17 are available online) list all 24 areas (12 de novo and 12 upgrading) with some information on the data we have on each.

Having defined the treated areas, we now explain how we construct our control areas. In much of our analysis, we use as control areas all initially unbuilt (greenfield) land within 500 m of the boundary of de novo.<sup>17</sup> We exclude areas that were uninhabitable (e.g., off the coast), built up, or designated for nonresidential use prior to the start of the Sites and Services projects. In order to infer what had been previously built up, we use historical maps and imagery collected as close as possible to the start of the Sites and Services project and, where possible before its start date, as discussed in the appendix.<sup>18</sup>

To construct control areas for the upgrading areas, we similarly use greenfield areas within 500 m of upgrading or, alternatively, 21 slums that were delineated in the 1979 Dar es Salaam master plan (MMM 1979) and were not upgraded as part of Sites and Services. Comparisons across slums should be taken with caution, since in accordance with the planners' intention to target larger slums (see sec. II), the upgraded slums covered an average area about four times larger than the control slums. Both upgraded and nonupgraded slums, however, had similar initial population densities (195 people per hectare in the upgraded slums and 234 in nonupgraded

<sup>17</sup> Note that throughout our paper, the control areas always exclude de novo and upgrading areas.

<sup>18</sup> For some of the analysis, we also study untreated areas farther than 500 m from the treatment areas, in which case, we again excluded areas that were built up before Sites and Services began.

slums in 1979). Figure A1 shows the de novo, upgrading, and control areas in all seven cities.<sup>19</sup>

Our empirical approach described below assumes that both the de novo and the control areas were unbuilt (greenfields) before the onset of Sites and Services. To provide evidence that this was indeed the case, we use a subsample of the TSCP survey data, which provides construction years for buildings in Mbeya and Mwanza (see appendix). We report results from using these data cautiously, since they involve a fairly small sample and a variable (construction year) that is measured with noise and observed only for surviving houses. With these caveats in mind, we note that, of the housing units whose construction date is known, only about 0.5% in de novo areas and about 1.3% in the nearby control areas were built before the start of Sites and Services, suggesting that the control and de novo areas were probably very sparsely populated.

## 2. How Do We Construct the Units of Analysis?

Our research design (discussed below) uses as its main units of analysis a grid of  $50 \times 50$ -m blocks, each of which is assigned to de novo, upgrading, or control area, depending on where its centroid falls. This creates a fine partition of our study area, which allows us to account for empty areas at the block level and within blocks. As we explain below, however, data constraints compel us to conduct some of the analysis at the level of individual housing units, or at the level of 2012 census enumeration areas (EAs) or subunits of EAs (Tanzania National Bureau of Statistics 2014, 2017).

## 3. What Are the Key Variables We Measure?

To study the quality of housing across all 24 Sites and Services locations, we use high-resolution WorldView satellite images (DigitalGlobe 2016).<sup>20</sup> We employed a company (Ramani Geosystems) to trace out the building footprints from these data for six of the seven cities. For the final city, Dar es Salaam, we used separate building outlines from a freely available source—Dar Ramani Huria (2016). For all seven cities, we then assembled more information on outcomes and control variables, as we explain in the appendix. Here we describe some of the key variables.

For the purpose of measuring the quality of private housing using imagery data, we think of slums as typically containing small and irregularly laid-out buildings made of low-quality materials and with poor access to

<sup>19</sup> To keep the maps on a fixed and legible scale, we do not show the locations of the nonupgraded slums in Dar es Salaam.

<sup>20</sup> The resolution of the images is  $50 \times 50$  cm for grayscale and a little coarser for color.

roads. We therefore define as positive outcomes the opposite of this image of slums: buildings with large footprints that are regularly laid out and have good roofs and access to roads. We use three outcomes that we think of as largely reflecting private complementary investments. First is the logarithm of building footprint size, derived directly from the processed imagery. Second, we use the color satellite imagery to assess whether each roof is likely painted and therefore less prone to rust. Third, we calculate the orientation of each building using the main axis of the minimum bounding rectangle that contains it. We then calculate the difference in orientation between each building and its nearest neighboring building, modulo 90°, with more similar orientations representing a more regular layout.<sup>21</sup> Finally, we construct an indicator for buildings that are within no more than 10 m of the nearest road. Unlike the three previous measures, we think of this measure of road access as largely representing persistence of Sites and Services infrastructure investments.

While the imagery and the outcomes we derive from it have the advantage of broad coverage, we complement them with detailed survey data on all the buildings in three of the Sites and Services cities, Mbeya (in southwest Tanzania), Tanga (in northeast Tanzania), and Mwanza (in northwest Tanzania). These data are derived from the TSCP survey, which was conducted from 2010 to 2013 (World Bank 2010). We use these data to build a more detailed picture of building quality in the areas we study. The TSCP data allow us to identify outbuildings (e.g., sheds, garages, and animal pens) that are generally smaller and which we exclude from the analysis.<sup>22</sup> This leaves us with a sample of buildings that are used mostly for residential purposes, although a small fraction may also serve commercial or public uses.

We use the TSCP data to measure the logarithm of building footprint and create indicators for buildings that have more than one story, good (durable) roof materials, connection to electricity, and at least basic sanitation.<sup>23</sup> These measures likely reflect private investments, since they were not part of the Sites and Services investments. In addition, we measure connection to water mains and having road access as largely reflecting

<sup>21</sup> When we regress the log hedonic price index (discussed below) on the three imagery measures using a block-level regression, the coefficients on each of the three measures is positive and significant. This provides further support for our use of these measures of housing quality. Where applicable, we standardize and pool the three quality measures together to construct a “family of outcomes” z-score (Kling, Liebman, and Katz 2007; Banerjee et al. 2015).

<sup>22</sup> Outbuildings account for around 10%–30% of buildings in the areas we consider, where the fraction varies by city. Their mean size is typically around one-third that of the average regular building size.

<sup>23</sup> In the de novo, upgrading, and control areas, we classify as “basic sanitation” having either a septic tank (30% of buildings) or sewerage connection (0.5%). Not having basic sanitation usually means a pit latrine (67%) or “other” or none. As before, we construct a “family of outcomes” measure based on nonmissing observations for each variable.

persistence of Sites and Services investments. The TSCP data also provide the full names of owners of housing units, which we use as explained below.

We also use separate TSCP (World Bank 2013) valuation data for Arusha, a city where Sites and Services were not implemented, to construct a hedonic measure of building quality, as we explain in the appendix.<sup>24</sup> Another separate data source (Tanzanian Ministry of Lands 2012) provides us information about land values in Dar es Salaam, although at a coarser level.

In addition to these variables, we construct geographic variables (distance to the nearest shore, an indicator for rivers or streams, and a measure of ruggedness) and other variables, which we use in our analysis below. All these are again explained in the appendix.

We complement all these measures of the physical environment, with some data on people, including indicators for owners (identified by their full name and city), taken from the TSCP survey, as well as population density and measures of schooling and literacy, which we calculate from the 2002 and 2012 censuses at the level of enumeration areas. We sometimes split enumeration areas to allocate them across treatment and control areas, as we explain in the appendix.

#### 4. How Do the Different Areas Compare Using Raw Data?

Table A3 summarizes information on the number of plots and the population density, as of 2002, in *de novo* and upgrading areas, and their respective control areas. As the table shows, *de novo* areas were more densely populated than nearby control areas. Upgrading areas were very densely populated, and again denser than control areas near them. As we shall see below, the higher density in upgrading areas did not correspond to more multistory buildings but, in fact, the opposite.

Figure A2 shows visual examples of parts of a *de novo* area, a control area near *de novo*, and an upgrading area, all in the same district of Dar es Salaam. The differences between the most orderly location (*de novo*) area and the least orderly one (upgrading) are visibly clear, and the control area lies somewhere in between.

The impression that *de novo* areas have higher-quality housing is corroborated in the summary statistics table (table A4). The imagery data shows that compared with the control areas, *de novo* areas have buildings with larger footprints, a higher fraction of painted roofs, more regularly

<sup>24</sup> Our approach of using characteristics linearly in a hedonic regression follows that of Giglio, Maggiori, and Stroebe (2014). There is also some evidence that in the case of housing, using the imputed hedonic values as dependent variables does not lead to much bias in the inference (McMillen and Redfearn 2010; Diewert, de Haan, and Hendriks 2015).

laid-out buildings, and better access to roads. The survey data shows that de novo areas are also more likely to have multiple stories, good roof materials, connection to electricity, basic sanitation, and connection to water mains, as well as a much higher hedonic value. On almost all these measures, including the fraction of buildings with multiple stories, upgrading areas look worse, and control areas are somewhere in between de novo and upgrading. The log hedonic price differences suggest that, on average, de novo housing units are about 63% more valuable than those in control areas and about 92% more valuable than those in upgrading areas.

### III. Research Design and Empirical Findings

#### A. Research Design

The differences in outcomes described in table A4 suggest that housing quality in de novo areas is considerably better than in control areas. The higher quality of housing in de novo areas reflects both elements that Sites and Services invested in directly, such as roads and water, and elements that it did not, such as electricity. But in order to study whether the de novo investments did in fact crowd in private investments (and, if so, by how much), we need to move beyond the descriptive statistics, as this section explains.

Our identification strategy compares de novo areas with nearby control areas, which (like de novo areas) were largely empty before the onset of Sites and Services. In our main analysis, we follow Gelman and Imbens (2017) by implementing a semiparametric RD design:

$$y_i = \beta_0 + \beta_1 \text{Denovo}_i + \beta_2 \text{Dist}_i + \beta_3 \text{Dist}_i \times \text{Denovo}_i + \beta_4 \text{Nearest\_Denovo}_i + \beta_5 \text{Dist\_CBD}_i + \beta_6 \text{Controls}_i + \epsilon_i, \quad (1)$$

where  $y_i$  measures various outcomes, as described in section II and the appendix;  $\text{Denovo}_i$  is the main regressor of interest, which indicates whether the centroid of  $i$  is in de novo areas, where control areas are the omitted category;  $\text{Dist}_i$  is the distance in kilometers to the boundary between de novo and control areas; **Nearest\_Denovo** $_i$  is a vector of fixed effects for the nearest de novo areas;  $\text{Dist\_CBD}_i$  measures the distance in kilometers of unit  $i$  from the central business district (CBD) of the city in which it is located; **Controls** $_i$  is a vector of additional controls, which we discuss below; and  $\epsilon_i$  denotes the error term. The role of distance to the CBD is emphasized in many urban economics models (for an overview, see Duranton and Puga 2015), and adding **Nearest\_Denovo** $_i$  ensures that we only compare control areas with their nearest de novo area. In



our baseline specification, each observation is a  $50 \times 50$ -m block, but later on, as we explain, we also use housing units within buildings and enumeration areas as units of analysis.

Our baseline analysis uses data from within 500 m of the boundary between *de novo* and control areas. Using this fixed distance allows us to analyze all our outcomes across imagery and TSCP data (World Bank 2013) consistently. As we discuss further below, 500 m also turns out to be fairly close to the optimal bandwidth we find for our key outcomes using the survey data. Finally, we also present below alternative specifications using more—and less—data.

In our baseline estimates, we cluster the standard errors on  $850 \times 850$ -m blocks, following the approach of Bester, Conley, and Hansen (2011) and Bleakley and Lin (2012). The size of the blocks on which we cluster reflects the size of the Sites and Services neighborhoods. The median size of the 12 *de novo* neighborhoods was approximately 0.538 sq. km, and the median size of all 24 neighborhoods was around 0.718 sq. km. This last figure is just a little smaller than the area of a square whose sides are 850 m, which we chose as a conservative benchmark for clustering.<sup>25</sup>

## 1. Addressing Threats to Identification

Our identification strategy assumes that conditional on the controls in specification (1), the potential expected outcome functions are continuous at the discontinuity threshold. Our spatial RD approach is similar to that of Dell (2010), and much of our analysis likewise applies a semiparametric RD, which combines both controls, as in equation (1), and a focus on areas that are close to (within 500 m of) the boundary of *de novo* and control areas.<sup>26</sup>

One potential concern is that the areas selected for *de novo* differed in their “first nature” location fundamentals. But in our setting, the geographic distances are much smaller than in most other settings, so we are less concerned with larger-scale changes in geography, such as climate or soil fertility. Further, in our empirical analysis, we report balancing tests, which use specification (1) to compare the geographic variables as

<sup>25</sup> In earlier versions of this paper, we also reported specifications using Conley (1999) standard errors with a decay area equal to the size of the abovementioned blocks, and the results were similar. To mitigate concerns about the variation in neighborhood size, we also experimented with modifying our baseline clustering blocks to treat each Sites and Services neighborhood as a separate clustering unit, with the remainder of the cluster units based on the grid (cut where necessary by the Sites and Services neighborhoods). Once again, the estimated standard errors were quite similar.

<sup>26</sup> Since we have variation within several cities, we use functions of distance to the *de novo* boundary in our main specification and functions of longitude and latitude only in our robustness checks, as we discuss below.

outcomes as we cross the *de novo*–control boundary. Some of the geographic variables—the land’s ruggedness and the presence of rivers or streams—may be endogenous to housing development. Therefore, below we report estimates both with and without the geographic controls.

Our identification strategy also assumes that both *de novo* and control areas were essentially empty (greenfields) before the start of Sites and Services. As we discuss in the appendix, our classification of areas relies on historical aerial images and topographic maps, which allow us to detect preexisting buildings. And in section II, we provide support for this assumption using a subsample of buildings for which we have construction dates.

Another relevant question is whether administrative boundaries correspond to some of the *de novo*–control boundaries, leading to different municipal policies on either side of the boundary. To address this question, we verified that in none of the cases do the boundaries between any treatment areas and the control areas coincide with the ward or district boundaries.<sup>27</sup>

A different type of concern is that there may be spillovers across neighborhoods.<sup>28</sup> So, for example, it is possible that proximity to *de novo* areas improves nearby control areas or that proximity to control areas worsens *de novo* areas; both would attenuate our estimates. To mitigate this concern, we report donut RD specifications, which exclude bands of 100 m around the boundary between *de novo* and control areas. To mitigate a related concern that upgrading areas may be affecting our estimates, we also report specifications that exclude all blocks within 100 m of upgrading areas. In a similar vein, since the TSCP but not the imagery data cover entire cities, we also report specifications that use wider control areas rather than only those near *de novo* areas. In those cases, we use the same specification as in the baseline but also report some results using second- and third-order polynomials in distance to the boundary.<sup>29</sup>

A related concern is that Sites and Services may have reshaped cities and even affected the location of their CBD and the distance to it. To address this concern, we report robustness checks, which use distances to historically central locations—mostly railway stations, as discussed in section II and the appendix.

<sup>27</sup> The closest case is that of Mwanza, where one district (Nyamangana) cuts into less than a quarter of the control area, while another (Ilemela) contains all of the treatment and most of the control area. However, this boundary was only observed in the 2012 census and not in the 2002 census, so it is almost certainly either unrelated to the Sites and Services project or an indirect outcome of it. In the 2002 census, the Ilemela district fully contained the Mwanza treatment and control areas.

<sup>28</sup> See related discussions by Turner, Haughwout, and Van Der Klaauw (2014), Redding and Sturm (2016), and Hornbeck and Keniston (2017).

<sup>29</sup> The full city data also allow us to estimate regressions using an optimal bandwidth (Imbens and Kalyanaraman 2012), which we also report.

## 2. Did Sites and Services Create or Displace Value?

Another question that we consider is whether Sites and Services created value or merely displaced it. Like many studies of place-based policies, it is difficult for us to answer this question definitively, since we do not have counterfactual cities of similar size, which were untreated by Sites and Services. And even if such cities had existed, one might still have worried about displacement of activity across cities. Nevertheless, our findings below suggest that *de novo* areas are relatively regularly laid out and preserve good access to roads. It therefore seems likely that by solving coordination failures, they created value and not merely displaced it.<sup>30</sup>

## 3. Exploring Mechanisms: Sorting across Neighborhoods and Infrastructure Persistence

Our setting allows us to explore another important issue—the sorting of owners across neighborhoods. As we discuss above, initial ownership criteria in *de novo* areas excluded the poorest, and program loans may have further alleviated credit constraints for some of these owners (as well as for some of the owners in upgrading areas). The model characterizes sufficient conditions under which including owner fixed effects overcomes the potential differences in credit constraints of owners who rent out multiple housing units.<sup>31</sup> We note that renting is fairly common in our setting: as of 2007, renters accounted for a small majority of Dar es Salaam's residents and more than a third of residents in other urban areas; back in 1992, the share of renters was even higher (Komu 2013).

To shed light on the sorting of residents across neighborhoods, we also use census data to characterize residents by measures of education, which are the best proxies we have for lifetime earnings.

Our model also highlights the role of persistently better infrastructure in *de novo* neighborhoods as a mechanism for crowding in investments in housing quality. Empirically, we estimate regressions of the same form of as equation (1), using measures of water connection and access to roads as outcomes, since these closely relate to the investments made in the Sites and Services projects.

<sup>30</sup> As we also discuss below, our findings suggest that Sites and Services not only had positive effects on local land values, they may also have generated positive spillovers on nearby areas, an issue that we revisit in sec. III.

<sup>31</sup> To be precise, we consider a full name as different if it appears in more than one city. In practice, this does not seem to make much difference. Since this strategy uses variation within owners, it employs only part of the data, so in this case, we need to use control areas from the rest of the city to ensure sufficient variation. We also acknowledge that some units may be owner occupied, while others may be rented out, but we cannot separate the two with our data.

#### 4. Studying Upgrading Areas

Finally, we repeat our analysis for upgrading areas, comparing them with proximate control areas, following the procedure outlined above.<sup>32</sup> Finding appropriate counterfactuals for upgrading areas (which were populated before the program began) is harder than for de novo areas (which were essentially empty). To mitigate concerns about different starting conditions, we also report regressions that compare upgrading areas with 21 other slums that existed in Dar es Salaam in 1979 but were not upgraded as part of Sites and Services. The slums that were not upgraded were, on average, smaller in area (see sec. II) but had similar, or even slightly higher, population density in 1979. The comparisons of upgrading areas with non-upgraded slums come with two caveats: first, this analysis is not a spatial RD, since the nonupgraded slums were not adjacent to the upgraded ones, although for consistency, we still use specification (1); and second, these comparisons are only possible for the imagery data, since Dar es Salaam is not covered by the TSCP data.

### B. *Empirical Findings*

#### 1. Balancing Tests

We begin the discussion of our findings by reporting balancing tests on geographic characteristics. As table A5 shows, when we compare geographic characteristics in de novo areas with nearby control areas, both distance to the shore and ruggedness differ in de novo areas (panel A), but after including our baseline controls as in equation (1) (panel B), de novo and control areas look balanced. We also report balancing tests using TSCP data, which also look balanced (with the exception of rivers and streams in the sample adjacent to the de novo areas). We note, however, that rivers and ruggedness may be endogenous to the de novo development, which may have flattened the soil and buried or diverted some streams. For completeness, we report below estimates both with and without the geographic controls.

#### 2. Crowding In of Private Investments

We now turn to our main results. In table 1, we report estimates using specification (1) and our imagery sample. Panel A shows that de novo areas have footprints that are roughly 12% larger and have more regular layout, but their roof quality is not better. The *z*-score aggregating all three measures indicates that de novo areas have higher-quality housing than nearby areas, and other estimates show that they have fewer empty

<sup>32</sup> In upgrading area regressions, we measure distance to the upgrading (vs. de novo)–control boundary and fixed effects for the nearest upgrading (vs. de novo) area.

TABLE 1  
DE NOVO REGRESSIONS USING IMAGERY DATA FOR ALL SEVEN CITIES

	(1)	(2)	(3)	(4)	(5)	(6)
A. 500-m Bandwidth						
	Mean Log Building Footprint Area	Share of Buildings with Painted Roof	Mean Similarity of Building Orientation	Mean zScore	Share of Empty Blocks	Share of Area Built Up
De novo	.114 (.051)	−.013 (.012)	2.821 (.722)	.168 (.057)	−.152 (.037)	.094 (.013)
Observations	6,562	6,500	6,562	6,562	8,440	8,440
Mean (control)	4.457	.184	−8.669	.042	.306	.155
B. Robustness (Mean zScore Only as Outcome)						
	Geography	Lat-Long. Second Polynomial	Historical CBD	Donut 100 m	Exclude 100 m to Upgrade	
De novo	.143 (.053)	.156 (.057)	.168 (.057)	.241 (.100)	.175 (.059)	
Observations	6,562	6,562	6,562	4,568	6,158	
Mean (control)	.042	.042	.042	.015	.047	

NOTE.—Estimates from regressions use specification (1) and block-level observations with outcomes derived from imagery for all seven Sites and Services cities. The sample includes de novo areas and control areas within 500 m of their boundary. The outcomes are measures of housing quality that do not reflect direct investments in de novo areas. Each observation is a block based on an arbitrary grid of 50 × 50-m blocks. Blocks are assigned to de novo or control areas based on where their centroid falls. Outcomes are derived from the set of buildings with a centroid in the block (see appendix for details). In panel A, the outcomes vary, while in panel B, the dependent variable in all columns is the z-score (composed of all outcomes in cols. 1–3 in panel A). In each specification, the regressor of interest is de novo, and the control variables include a linear control in distance to the de novo–control area boundary interacted with the de novo indicator, fixed effects for the nearest de novo area, and distance to the central business district (CBD) of each city. In addition, in panel B, col. 1 includes geographic controls, col. 2 includes a second-order polynomial in longitude and latitude, col. 3 uses distance to historical (vs. contemporary) CBDs, col. 4 excludes areas within 100 m of the boundary between de novo and control areas, and col. 5 excludes areas within 100 m of the boundary between upgrade and control areas. Standard errors, in parentheses, are clustered by arbitrary 850 × 850-m grid squares, corresponding to the median size of Sites and Services areas. There are 90 clusters, except in cols. 5 and 6 of panel A, which have 92 clusters; col. 4 of panel B, which has 89 clusters; and col. 5 of panel B, which has 88 clusters.

blocks and that a higher fraction of their area is built up.<sup>33</sup> Panel B reports robustness checks for the z-score using geographic controls, longitude and latitude polynomials, an alternative measure of CBDs that predates Sites and Services, and excluding blocks near upgrading areas—all are similar to our baseline estimate. When we use donut RD specifications

<sup>33</sup> To visualize our results, fig. A3a shows a regression discontinuity plot of binned values of the z-score.

to exclude areas near the boundary of de novo and controls, the estimates increase somewhat, suggesting that our baseline estimates may be a little attenuated due to spillovers (positive ones from de novo to controls, negative ones from controls to de novo, or both). This finding also suggests that the higher-quality housing in de novo areas may generate positive spillovers in neighboring areas (for related discussions of local spillovers, see Turner, Haughwout, and Van Der Klaauw 2014 and Hornbeck and Keniston 2017).

In sum, results for all seven cities using the satellite image data suggest that de novo areas have larger and more regularly oriented buildings. To get a more detailed picture of the differences in residential quality, we turn to the TSCP data for Mbeya, Mwanza, and Tanga. In panel A of table 2, we report results again using specification (1). One advantage of the survey data is that, unlike the imagery data, they allow us to focus on residential buildings by excluding outbuildings, which we do. As panel A shows, buildings in de novo areas have footprints that are about 50% (or 0.41 log points) larger than the control areas. They are also about 23 percentage points (or 48%) more likely to be connected to electricity. The regressions also show economically large but statistically imprecise differences in favor of de novo areas in the share of buildings with multiple stories and with at least basic sanitation but, again, almost no difference in roof quality.

We aggregate the measures of quality in the survey data in two ways: first, using a z-score and, second, using the predicted log hedonic value. Regressions using either as an outcome indicate significantly higher residential quality in de novo areas than in control areas.<sup>34</sup> Specifically, the regressions suggest that the hedonic price is around 56% higher in de novo areas. This may understate the actual differences in house values, since the hedonics do not directly account for either all housing characteristics or the full impact of local neighborhoods' infrastructure. Noting this caveat, a result from the model in section IV.D below suggests that land value differences in de novo (vs. control areas) are about 50% larger than house price differences. This result, combined with our hedonic estimates, suggests that land values in de novo areas are at least 86% higher than in control areas. To interpret this difference, we note that in Dar es Salaam, land values in de novo neighborhoods are in the range of \$160–\$220 per square meter (in 2017 prices).<sup>35</sup> Combined with our estimates above, this suggest that de novo may have increased local land values by at least \$75–\$100 per square meter.

These values are high compared with the cost of investments per unit of treated plot area, which we estimate above to be no more than \$8 per

<sup>34</sup> Panels *b* and *c* of fig. A3 show regression discontinuity plots for the z-score and log hedonic prices.

<sup>35</sup> The coarse data we have on land values do not separately identify the control areas near de novo.

TABLE 2  
DE NOVO REGRESSIONS USING TSCP SURVEY DATA FOR MBEYA, MWANZA, AND TANGA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean Log Building Footprint Area	Share of Buildings with Multiple Stories	Share of Buildings with Good Roof	Share of Buildings Connected to Electricity	Share of Buildings with Sewerage or Septic Tank	Mean z-Score	Mean Log Hedonic Value
A. 500-m Bandwidth							
De novo	.405	.081	-.010	.226	.142	.342	.446
Observations	(.070)	(.066)	(.008)	(.039)	(.091)	(.091)	(.081)
Mean (control)	2,009	1,975	2,009	2,009	2,008	2,009	2,009
	4,739	.096	.984	.466	.381	.033	17,234
B. Robustness (Mean z-Score Only as Outcome)							
	Geography	Lat-Long. Second Polynomial	Historical CBD	Donut 100 m	Exclude 100 m to Upgrade	Full City	Optimal Bandwidth
De novo	.263	.323	.342	.408	.375	.588	.312
Observations	(.091)	(.076)	(.090)	(.190)	(.093)	(.079)	(.082)
Mean (control)	2,009	2,009	2,009	1,410	1,887	34,602	34,602
	.033	.033	.033	.001	.022	-.149	.038

C. Robustness (Mean Log Hedonic Value Only as Outcome)							
De novo	.329 (.081)	.431 (.059)	.446 (.077)	.541 (.190)	.427 (.077)	.505 (.089)	.411 (.063)
Observations	2,009	2,009	2,009	1,410	1,887	34,602	34,602
Mean (control)	17,234	17,234	17,234	17,231	17,229	17,113	17,239

NOTE.—Estimates from regressions use specification (1) and block-level observations with outcomes derived from TSCP survey data for the three cities where these data exist: Mbeya, Mwanza, and Tanga. The sample includes the de novo areas and control areas within 500 m of their boundary. The outcomes are measures of housing quality that do not reflect direct investments in de novo areas. Each observation is a block based on an arbitrary grid of  $50 \times 50$ -m blocks. Blocks are assigned to de novo or control areas based on where their centroid falls. Outcomes are derived from the set of buildings with a centroid in the block (see appendix for details). In panel A, the outcomes vary, while in panel B, the dependent variable in all columns is the *z*-score (composed of all outcomes in cols. 1–5 of panel A), and in panel C, the dependent variable is the predicted log value from hedonic regressions. In each specification, the regressor of interest is de novo, and the control variables include a linear control in distance to the de novo–control area boundary interacted with the de novo indicator; fixed effects for the nearest de novo area, and distance to the central business district (CBD) of each city. In addition, in panels B and C, col. 1 includes geographic controls, col. 2 includes a second-order polynomial in longitude and latitude, col. 3 uses distance to historical (vs. contemporary) CBDs, col. 4 excludes areas within 100 m of the boundary between de novo and control areas, col. 5 excludes areas within 100 m of the boundary between upgrade and control areas, col. 6 changes the control area to the sample of blocks covering the whole city excluding de novo areas, and col. 7 uses 2,033 observations inside the optimal bandwidth for panel B and 1,882 observations inside the optimal bandwidth for panel C, based on the work of Imbens and Kalyanaraman (2012). “Full City” in col. 6 is robust to higher-order polynomials in distance to boundary. In panel B, second-order polynomial gives an estimate of 0.571 and standard error (SE) of 0.087, and third-order polynomial gives an estimate of 0.381 and SE of 0.098. In panel C, second-order polynomial gives an estimate of 0.498 and SE of 0.104, and third-order polynomial gives an estimate of 0.296 and SE of 0.117. The control mean in col. 7 reports the mean for the control areas inside the optimal bandwidth (Imbens and Kalyanaraman 2012). Standard errors, in parentheses, are clustered by arbitrary  $850 \times 850$ -m grid squares, corresponding to the median size of Sites and Services areas. There are 29 clusters, except in col. 5 of panels B and C, which has 28 clusters, and cols. 6 and 7 of panels B and C, which has 439 clusters.



square meter of plot area or no more than \$13 per square meter, if we include indirect costs (in 2017 US dollars). While these estimates should be interpreted with caution, they suggest that the gains from *de novo* investments were large, at least in Dar es Salaam. That said, we acknowledge that the gains in other cities, where prices are lower, may not be quite as high.<sup>36</sup>

In panel B of table 2, we report results from a series of robustness checks, focusing for brevity on the *z*-score and the log hedonic price. The estimates with geographic controls in column 1 are a little lower than the baseline; this could be because either the baseline regressions overstate the difference due to better geographic fundamentals in *de novo* location or the geographic controls are themselves outcomes, and adding them understates the impact of *de novo*. Columns 2 and 3 show that controlling for the polynomial of longitude and latitude or using distance to historical (vs. contemporary) CBDs makes little difference compared with panel A. The donut specification in column 4 is larger than the baseline, suggesting (as in table 1) that the baseline estimates may be too small due to positive spillovers from *de novo* to controls (or negative ones going the other way). Column 5 excludes blocks near upgrading areas, and the results are similar to the baseline. Column 6 uses control areas from the rest of the city, and the estimates are again larger, possibly because we are comparing *de novo* areas with a control group that is, on average, farther away and less affected by local spillovers.<sup>37</sup> Finally, column 7 uses an optimal bandwidth, following Imbens and Kalyanaraman (2012), and the estimate is again quite similar to the baseline.

The results using hedonic values as outcomes in panel C follow a similar pattern, where adding geographic controls reduces the estimate a little, and excluding areas near the boundary increases them a little. The main message, however, is that our baseline estimates are robust to using different specifications.

### 3. The Role of Sorting

The results discussed so far are silent on the respective role of the *de novo* treatment and the endogenous sorting across neighborhoods of owners with different levels of credit constraints. As our model below (in sec. IV) shows, we can account for differences across areas in owners' credit constraints by adding owner fixed effects, which allow us to isolate the impact of *de novo* areas compared with control areas for owners with multiple housing units. The units of analysis used in these regressions are individual

<sup>36</sup> Unfortunately, our land value data for other cities are either missing or not detailed enough to give a credible picture.

<sup>37</sup> The estimates are robust to using second- and third-order polynomials, although in the latter case they are smaller.

housing units, since this is the level at which ownership is defined. The housing units we focus on are those owned by owners of multiple units, which account for about 13% of all housing units. To ensure a sufficiently large sample, we reestimate specifications as in (1) for the full city TSCP sample, but now focusing on housing units whose owners have more than one unit. Table 3 reports estimates of these regressions with owner fixed effects (panel A) and without them (panel B). The estimates show that in this sample, housing units in *de novo* areas are considerably larger and much more likely to have electricity and basic sanitation. Without owner fixed effects, they also are more likely to be in multistory buildings, although this difference becomes smaller and only marginally significant once we control for owner fixed effects. As reported previously, *de novo* housing units do not have better roof materials. The difference in quality between *de novo* and control areas, as reflected in the *z*-score and the hedonic value, suggests that *de novo* areas may be about 60 log points (or about 83%) more valuable; as discussed above, this may understate the actual difference, since it is unlikely to reflect all the amenity differences. Panels C and D of the table report robustness checks for the specifications with and without fixed effects, using the *z*-score as an outcome. Across a range of specifications reported in table 3, roughly a third of the quality advantage of *de novo* areas is accounted for by the different ownership, and the rest likely reflects the impact of *de novo* on quality for owners who are relatively unconstrained in terms of investment.<sup>38</sup>

The characteristics of residents in *de novo* areas, compared with control areas, likely reflect their willingness to pay for higher-quality housing. In table A6, we report regressions using 2012 census data with cut enumeration areas as units of analysis (see sec. II and appendix for details).<sup>39</sup> Consistent with the results discussed above, residents in *de novo* areas are better educated and more likely to be literate in English. The higher schooling level of *de novo* residents is consistent with sorting across neighborhoods and a higher willingness of the more educated to pay for better housing quality, although it is also possible that some of it is the result of better access to schooling of existing residents. Still, as table A6 shows, only about 55% of adults in *de novo* areas had more than primary school education, so the other 45% had no more than primary school education. This means that many less educated Tanzanians are still benefitting from *de novo* amenities.

<sup>38</sup> When we use the hedonic measure as an outcome, the regression estimates with and without owner fixed effects are more similar to each other (results available on request).

<sup>39</sup> In this case, the number of units of analysis is small and they are uneven (some are whole EAs and some are cut), which makes it difficult to get a good measure of distance to the boundary. Therefore, in these specifications, we use nonparametric RD without controls for distance to the boundary.

TABLE 3  
DE NOVO REGRESSIONS USING TSCP SURVEY DATA FOR MBAYA, MWANZA, AND TANGA WITH OWNER NAME FIXED EFFECTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Building Footprint Area		Multistory Building	Good Roof	Connected to Electricity	Sewerage or Septic Tank	zScore	Log Hedonic Value
	A. Full City, Owner Fixed Effects						
De novo	.553	.119	-.002	.417	.123	.447	.604
Observations	(.145)	(.062)	(.038)	(.078)	(.091)	(.088)	(.133)
Mean (control)	20,177	16,605	20,054	20,139	19,595	20,177	20,177
	4.573	.164	.968	.404	.249	-.016	17.016
	B. Full City, No Owner Fixed Effects, Same Sample as A						
De novo	.594	.514	-.010	.405	.122	.642	.612
Observations	(.177)	(.138)	(.015)	(.066)	(.069)	(.086)	(.185)
Mean (control)	20,177	16,605	20,054	20,139	19,595	20,177	20,177
	4.573	.164	.968	.404	.249	-.016	17.016
	C. Robustness Owner Fixed Effects (zScore Only as Outcome)						
Geography	Lat.-Long. Second Polynomial	Historical CBD	Donut 100 m	Exclude 100 m to Upgrade	Second-Order Polynomial	Third-Order Polynomial	
De novo	.422	.411	.434	.336	.471	.431	.372
Observations	(.085)	(.092)	(.087)	(.166)	(.093)	(.112)	(.140)
Mean (control)	20,177	20,177	20,177	19,694	19,729	20,177	20,177
	-.016	-.016	-.016	-.019	-.018	-.016	-.016



#### 4. The Persistence Infrastructure

To conclude our empirical analysis of the *de novo* areas, we explore whether their better housing quality corresponds to persistently better infrastructure. Here we focus on two of the main investments in Sites and Services, roads and water mains, and we again use specification (1). As panel A of table 4 shows, across both our imagery and TSCP data, *de novo* areas enjoy better access to roads, and the TSCP data also show that they are more likely to be connected to water mains.<sup>40</sup> Panels B–D report robustness checks using the same specifications as in table 2. Again, the estimates are a little smaller when we control for geographic covariates and a little larger when we focus on control areas that are further from *de novo*, with our main estimates in between. All the estimates are positive and statistically significant, showing that *de novo* investments translated into better infrastructure in the long run.

#### 5. Upgrading Areas

Having discussed the *de novo* areas, we now briefly discuss what we can learn from similar regressions for upgrading areas. As table A7 suggests, upgrading areas look fairly similar to nearby control areas in terms of the geographic controls, except that in some specifications, they are less rugged and less likely to have rivers or streams. When compared with the non-upgraded slums, and conditional on our baseline controls, the upgrading areas are closer to the shore but not significantly different in the other two geographic controls (results available on request).

Table A8 reports estimates using imagery data for all seven cities. Panel A suggests that housing quality in upgrading areas is similar to that of nearby control areas. The only significant difference is that upgrading areas have fewer empty areas and are more densely built up. Panel B shows that this conclusion is robust to a range of different specifications.

In panels C and D, we compare upgrading areas in Dar es Salaam only with the preexisting (old) slums that were not upgraded as part of Sites and Services. Once again, the results suggest that upgrading areas are quite similar to control areas, except that their buildings are a bit smaller and more regularly laid out. Upgrading areas also seem to have fewer empty blocks and a larger fraction of built-up area.

Next, in table A9, we use TSCP survey data outcomes. Here the upgrading areas look somewhat worse than nearby control areas: they have fewer multistory buildings and worse roofs, and their overall quality seems

<sup>40</sup> This last result is robust to excluding Tanga, where we have some uncertainty about the nature of *de novo* investments.

lower. This conclusion is reinforced in most of the robustness checks in panels B and C, although not all the estimates are precise.

In table A10, we examine the role of ownership in accounting for the worse quality in upgrading areas. The results suggest that ownership differences may partially explain the worse housing quality in upgrading areas, since controlling for owner fixed effects results in estimates that are small and, in most cases, imprecise.

A comparison of infrastructure persistence measures in upgrading areas may also help to explain why their housing is no better than that of nearby control areas. As table A11 shows, upgrading areas look similar to nearby areas in their access to roads and water; the coefficients on upgrading areas are small, imprecise, and mostly negative. By adding the coefficients and the control means and then comparing them with the estimates in de novo areas (table 4), it appears that upgrading areas have worse infrastructure than de novo areas. As we discussed in section II, upgrading areas did receive roads and water mains, and investments measured in dollars per square meter were similar to those of de novo areas. A likely explanation for the poor state of upgrading areas' infrastructure today is that those areas' infrastructure deteriorated more than that of de novo areas. Kironde (1994, 464) and Theodory and Malipula (2012) discuss evidence that infrastructure did in fact deteriorate in upgrading slums in Dar es Salaam. Kironde (1994) mentions, for example, the deterioration of roadside drainage due to lack of maintenance, private construction on land that was intended for public use, and the degradation of water provision infrastructure.

Finally, table A12 shows that residents of upgrading areas are less educated than those of nearby areas, consistent with the lower housing quality in these neighborhoods.

#### IV. Model

##### A. *Assumptions and Their Relationship to the Institutional Setting*

To frame our empirical analysis, we present a model that characterizes conditions under which investment in infrastructure (as defined below) incentivizes owners to build higher-quality housing. The model captures key aspects of our description of the Sites and Services projects in section II. It also connects to our econometric analysis in section III by relating gains in house values to gains in land values and motivating our use of owner fixed effects to account for owner sorting across neighborhoods.<sup>41</sup>

<sup>41</sup> Our model builds on the work of Hornbeck and Keniston (2017) but differs from theirs in several ways. We add to the model infrastructure and variation across owners in credit constraints, and we derive new analytical results. We also model spillovers across houses differently, and for simplicity, we exclude the exogenous time trends.

TABLE 4  
DE NOVO REGRESSIONS ON PERSISTENCE MEASURES USING IMAGERY AND TSCP SURVEY DATA

	(1)	(2)	(3)	(4)	(5)	(6)
	Share of Buildings with Road within 10 m (All Seven Cities) <sup>a</sup>	Share of Buildings with Road Access (Mbeya, Mwanza, and Tanga) <sup>b</sup>	Share of Buildings Connected to Water Mains (Mbeya, Mwanza, and Tanga) <sup>b</sup>	Share of Buildings Connected to Water Mains (Mbeya and Mwanza) <sup>b</sup>		
A. 500-m Bandwidth						
De novo	.141	.197	.211	.233		
Observations	(.028)	(.050)	(.057)	(.060)		
Mean	6,562	2,008	2,009	1,952		
(control)	.202	.477	.547	.547		
B. Robustness for Share of Buildings with Road within 10 m (Imagery)						
Geography		Lat-Long, Second Polynomial	Historical CBD	Donut 100 m	Exclude 100 m to Upgrade	Full City
B. Robustness for Share of Buildings with Road within 10 m (Imagery)						
De novo	.129	.142	.142	.185	.150	
Observations	(.025)	(.029)	(.028)	(.056)	(.029)	
Mean	6,562	6,562	6,562	4,568	6,158	
(control)	.202	.202	.202	.205	.197	

C. Robustness for Share of Buildings with Road Access (TSCP)					
De novo	.134	.190	.199	.191	.206
Observations	(.039)	(.049)	(.050)	(.159)	(.056)
Mean (control)	2,008	2,008	2,008	1,409	1,886
	.477	.477	.477	.485	.449
					.573
D. Robustness for Share of Buildings Connected to Water Mains (TSCP)					
De novo	.164	.188	.209	.319	.204
Observations	(.060)	(.051)	(.057)	(.128)	(.042)
Mean (control)	2,009	2,009	2,009	1,410	1,887
	.547	.547	.547	.535	.534
					.433

NOTE.—Estimates from regressions use specification (1) and block-level observations with outcomes derived from imagery for all seven Sites and Services cities (road within 10 m) and TSCP survey data for Mbeya, Mwanza, and Tanga (road access and connection to water mains). The sample includes the de novo areas and control areas within 500 m of their boundary. The outcomes are measures of persistence of infrastructure treatment. Each observation is a block based on an arbitrary grid of 50 × 50-m blocks. Blocks are assigned to de novo or control areas based on where their centroid falls. Outcomes are derived from the set of buildings with a centroid in the block (see appendix for details). In panel A, the outcomes vary, while in panel B, the dependent variable in all columns is the share of buildings with a road within 10 m (from imagery data), in panel C, the dependent variable in all columns is the share of buildings with road access (from TSCP data), and in panel D, the dependent variable is the share of buildings connected to water mains (from TSCP data). In each specification, the regressor of interest is de novo, and the control variables include a linear control in distance to the de novo–control area boundary interacted with the de novo indicator, fixed effects for the nearest de novo area, and distance to the central business district (CBD) of each city. In addition, in panels B–D, col. 1 includes geographic controls, col. 2 includes a second-order polynomial in longitude and latitude, col. 3 uses distance to historical (vs. contemporary) CBDs, col. 4 excludes areas within 100 m of the boundary between de novo and control areas, and col. 5 excludes areas within 100 m of the boundary between upgrade and control areas. Moreover, in panels C and D, col. 6 changes the control area to the sample of blocks covering the whole city excluding upgrade areas. Standard errors, in parentheses, are clustered by arbitrary 850 × 850-m grid squares, corresponding to the median size of Sites and Services area. There are 29 clusters in TSCP data, except in col. 5 of panels C and D, which have 28 clusters, and in col. 6 of panels C and D, which have 439 clusters. There are 90 clusters in imagery data, except in col. 4 of panel B, which has 88 clusters, and col. 5 of panel B, which has 89 clusters.

<sup>a</sup> Derived from imagery for all seven Sites and Services cities.

<sup>b</sup> Derived from TSCP survey data.



We consider a population of infinitely lived, profit-maximizing owners with formal or informal rights to build on their plots, which are organized into neighborhoods (areas). In each plot, the owner can build a house and rent it out.<sup>42</sup> The model is in discrete time, and in each period  $t \geq 1$ , owners maximize their expected present discounted stream of rents, net of house construction costs, on each plot they own:

$$E \left[ \sum_{s=t}^{\infty} \delta^s [r(q_s, I_s) - B_s c(q(I_s))] \right], \text{ s.t.} \quad (2)$$

$$\Pr(q_{t+1} = q_t) = 1 - d, \Pr(q_{t+1} = 0) = d.$$

The expectations are defined over the exogenous destruction probability of houses in each period, as discussed below. Owners are assumed to have a time preference  $\delta \in (0, 1)$ . The rent that each owner receives on each house in each period is  $r(q, I) = q^\alpha I^{1-\alpha}$ , where  $q$  and  $I$  denote the quality of the house and the neighborhood infrastructure, and  $\alpha \in (0, 1)$ . The indicator  $B_t$  equals one if a house is built in period  $t$  and zero otherwise. The construction costs of a house of quality  $q$  are  $c(q) = cq^\gamma$ , where  $c > 0$ ,  $\gamma > 1$ . This convex cost function generalizes the work of Hornbeck and Keniston (2017), who assume  $\gamma = 2$ . In a different context, Combes, Duranton, and Gobillon (2016) find that the production function for housing can be approximated by a constant returns-to-scale Cobb-Douglas function using land and other inputs, where the coefficient on nonland inputs is approximately 0.65. Holding land constant, this production function is consistent with a cost function  $c(q) = cq^\gamma = cq^{1/0.65}$ , or  $\gamma \approx 1.54$ .

Infrastructure captures a broad set of neighborhood characteristics, including formal and regularly laid-out plots, which reduce coordination failures and protect owners' property rights; roads, which reduce the cost of travel and trade; and water mains, which contribute to living standards and health.<sup>43</sup> Infrastructure also reflects other neighborhood-level effects.<sup>44</sup> For tractability, we consider three types of infrastructure: high quality ( $I_H$ ),

<sup>42</sup> A "house" in the model is shorthand for a housing unit that we consider in the empirical analysis. Unlike Bayer, Ferreira, and McMillan (2007), we do not account for renter heterogeneity in our model because we have no data on the rents paid and little information on the residents. Knowing more about renters would have allowed us to build a better picture of the welfare gains from de novo areas.

<sup>43</sup> Property rights protection may reduce the risk of outright expropriation, as we discuss below, as well as the risk of partial expropriation, when part of an owner's plot is built without authorization, which we do not model explicitly.

<sup>44</sup> In practice, other types of neighborhood effects may also matter. For example, the absence of proper sewerage may increase the risk of contagious diseases. Consistent with this, Picarelli, Jaupart, and Chen (2017) show that cholera outbreaks in Dar es Salaam were much more severe in slum areas with poor infrastructure. Another possibility is that neighborhoods with poor electrification and lighting (Painter and Farrington 1997) and high population density (Gollin, Kirchberger, and Lagakos 2017) may attract crime. While we think that both of these channels could amplify the land value differentials between neighborhoods, we do not have the data to study them in our context.

medium quality ( $I_M$ ), and low quality ( $I_L$ ), where  $I_H > I_M > I_L > 0$ . High quality describes the bundle that Sites and Services offered—mostly formal plots, roads, and water mains. We assume that high-quality infrastructure deteriorates to medium quality, unless the fraction of high-quality housing is larger than a constant  $\phi > 0$ .<sup>45</sup> Medium-quality infrastructure is basic and unmaintained (e.g., bumpy dirt roads). It may be either high-quality infrastructure that has deteriorated or it may start out as medium quality. We assume that medium-quality infrastructure does not deteriorate.<sup>46</sup> Low-quality infrastructure corresponds to the level that prevails without any infrastructure investments in the neighborhood.

There are two types of owners in the model. Unconstrained owners may each own any finite number of plots and afford any level of investment in each plot, while constrained owners may own no more than a single plot and may afford to build only low-quality housing  $q_L = q(I_L)$ , as defined below.<sup>47</sup> Consistent with our setting, we assume that no single owner has a sufficiently large number of plots to exert market power or to solve coordination problems that arise from neighborhood-level externalities.<sup>48</sup>

We consider three types of areas (neighborhoods), each with a continuum of plots.<sup>49</sup> De novo areas start with empty plots ( $q = 0$ ) and high-quality infrastructure ( $I_H$ ); control areas start with empty plots and medium-quality infrastructure ( $I_M$ ); and upgrading areas start out with low-quality housing ( $q_L$ ) and high-quality infrastructure.<sup>50</sup> This reflects the situation at the time when Sites and Services was implemented.<sup>51</sup>

<sup>45</sup> High-quality housing is  $q_H = q(I_H)$ , as defined below. The potential for infrastructure deterioration means that housing quality can be indirectly affected by neighboring houses through the effect on infrastructure. This mechanism is different from the direct impact of neighbors' housing quality in the work of Hornbeck and Keniston (2017).

<sup>46</sup> Our assumption that medium infrastructure and deteriorated infrastructure are equal in quality is a simplifying assumption, motivated by our empirical finding that upgrading areas are no better than nearby control areas in terms of access to roads and water. Adding further parameters for deteriorated high-quality and deteriorated medium-quality infrastructure would not add much insight to the model.

<sup>47</sup> The distinction between two types of owners allows us to analyze owner sorting in a simple way. The results would have been similar had we assumed that constrained owners could build up to any quality that is strictly lower than  $q(I_M)$ , as defined below.

<sup>48</sup> Our TSCP data indicate that only a small share of housing units are owned by those with more than a handful of plots. It is true that, in principle, a rich individual or a firm could buy up an entire neighborhood and internalize the externalities involved. But until recent years, the Tanzanian government exerted strict control that prevented the concentration of neighborhood ownership.

<sup>49</sup> Our model does not account for other types of neighborhoods, e.g., former colonial areas (which typically constitute a small and wealthy part of cities), nor do we consider movements between different neighborhoods within the city.

<sup>50</sup> In sec. II, we discuss the investments that were made as part of the Sites and Services projects. These suggest that investments per total land area in de novo and upgrading areas were similar.

<sup>51</sup> We also note that while the control areas we use were, by definition, empty to begin with, other areas looked like control areas but had a stock of low-quality housing by the time they received infrastructure  $I_M$ .

The initial fractions of unconstrained owners are  $\theta_D$  in de novo areas,  $\theta_C$  in control areas, and  $\theta_U$  in upgrading areas. We assume that (as in the real world) upgrading areas are targeted for their relatively poor population, so they have few unconstrained owners, and therefore  $\theta_U < \phi$ .

In every period, the following sequence of events takes place. First, each owner decides whether to build (or rebuild) a house on each plot they own.<sup>52</sup> Second, if the neighborhood's housing quality is insufficiently high, infrastructure quality deteriorates, as we discuss below. Third, each owner collects the rent on each house they own. Finally, there is an exogenous probability  $d > 0$  that each house is destroyed, resetting housing quality to zero.<sup>53</sup>

We assume that the risk that houses are destroyed and the fraction of owners of each type in each neighborhood are common knowledge, as is the understanding that all unconstrained owners build high-quality housing if the share of unconstrained owners is at least  $\phi$ . In Nash equilibrium, each owner solves their own maximization problem in each period, assuming that all other owners do the same.

### B. Solving the Model

This section characterizes the optimal level of investment by owners, beginning with unconstrained owners.

Unconstrained owners maximize profits on each plot they own by solving the following Bellman equation:

$$V(q, I) = \max \begin{cases} r(q, I) + \delta E[V(q, I)] \\ r(q(I), I) + \delta E[V(q(I), I)] - c(q(I)), \end{cases} \quad (3)$$

where  $r$  is return on a house (e.g., rent);  $q \geq 0$  is the house quality;  $I \geq 0$  is the infrastructure quality that is expected when rents are collected and from that point onward;  $q(I)$  is the optimal house quality; and  $c(q(I))$  is the cost of building a house of quality  $q(I)$ .<sup>54</sup>

<sup>52</sup> Following Henderson, Regan, and Venables (2017) and Hornbeck and Keniston (2017), we assume that owners cannot renovate incrementally and that houses do not depreciate. The assumption that rebuilding a higher-quality house requires a fresh start is particularly relevant for low-quality housing that characterizes poorer neighborhoods in East African cities. It may be possible to make minor improvements to a house built of tin or mud. However, demolition and construction from scratch is required to make meaningful improvements e.g., adding brick walls, multiple stories, or plumbing. For simplicity, we maintain the assumption that no incremental improvement is possible. Relaxing this would reduce the benefit of early (de novo) investments.

<sup>53</sup> If a house is destroyed, the owner retains their plot. Given the paucity of construction dates in our data, it is difficult to assess  $d$ , but Henderson, Regan, and Venables (2017) estimate it at 3.2% per year using data from Tanzania's neighbor, Kenya.

<sup>54</sup> We could have included a probability  $(1 - \psi)$  that a plot is fully expropriated at the end of each period. If that were the case, we would need to substitute  $\psi\delta$  instead of  $\delta$  throughout the analysis, but for simplicity, we focus on the case without expropriation,

The infrastructure quality that is anticipated when rents are collected and from that point onward is equal to the existing level, except where infrastructure of quality  $I_H$  deteriorates to  $I_M$ . This deterioration happens when the fraction of high-quality housing ( $q_H = q(I_H)$ ), as described below) is strictly lower than  $\phi$ .

The model reflects a trade-off between keeping the current house quality  $q$  and improving it to  $q(I)$ . But if an unconstrained owner's house is exogenously destroyed, it is always rebuilt at the optimal quality  $q(I)$ . Starting from an empty plot, the optimal house quality for an unconstrained owner anticipating infrastructure  $I$  at the time of rent collection is

$$q(I) = \left[ \frac{\alpha I^{1-\alpha}}{\gamma c(1 - \delta + d\delta)} \right]^{1/(\gamma-\alpha)}. \quad (4)$$

The quality of housing is characterized by the following comparative statistics. First,  $\partial q(I)/\partial \delta > 0$ , so more patient people invest more. Second,  $\partial q(I)/\partial d < 0$ , so a higher probability of house destruction leads to lower-quality housing. And finally,  $\partial q(I)/\partial c < 0$ , so a higher construction cost reduces housing quality.

If an unconstrained owner starts with housing  $q_1 \equiv q(I_1)$  but with infrastructure  $I_2$  (where  $I_2 > I_1$ ), they choose between two options.<sup>55</sup> They can replace their house with a higher-quality house, in which case their expected payoff is equal to the expected value of an unbuilt plot of land:

$$\pi(0, I_2) = \frac{q_2^\alpha I_2^{1-\alpha} - cdq_2^\gamma}{1 - \delta} - (1 - d)cq_2^\gamma, \quad (5)$$

where  $\pi(q, I)$  is the maximized expected payoff from an existing house of quality  $q$  and infrastructure quality  $I$ . Alternatively, they can keep the current quality  $q_1$  and build a better house only when their house needs rebuilding. In this case, their expected payoff is

$$\pi(q_1, I_2) = q_1^\alpha I_2^{1-\alpha} + \delta[(1 - d)\pi(q_1, I_2) + d\pi(0, I_2)]. \quad (6)$$

Solving this expression, we get

$$\pi(q_1, I_2) = \frac{q_1^\alpha I_2^{1-\alpha} + d\delta\pi(0, I_2)}{1 - \delta + d\delta}. \quad (7)$$

**Proposition 1.** For each level of infrastructure  $I_1 > 0$ , there exists a unique value  $I_1^{\text{crit}} = (\gamma/\gamma - \alpha)^{(\gamma-\alpha)/(\alpha(\alpha-1))} I_1$ , such that unconstrained owners starting with  $q_1 = q(I_1)$  and infrastructure  $I_2 = I_1^{\text{crit}}$  are indifferent

namely,  $\psi = 1$ . Higher patience may reflect, at least in part, a lower risk of expropriation. Collin, Sandefur, and Zeitlin (2015) elicit owners' perceived expropriation risk in Temeke, an informal area close to the CBD of Dar es Salaam, which implies a risk of around 8% per year. Given the setting, this is likely an upper bound to the perceived expropriation risk in the locations we study.

<sup>55</sup> We assume that if owners are indifferent, they do not improve their houses.

between rebuilding and not rebuilding, and owners rebuild if and only if  $I_2 > I_1^{\text{crit}}$ .

*Proof.* To obtain  $I_2 = I_1^{\text{crit}}$ , combine the condition  $\pi(q_1, I_1^{\text{crit}}) = \pi(0, I_1^{\text{crit}})$  with (5) and (6), where housing quality  $q_2 = q(I_2)$  comes from (4). To show that owners rebuild if and only if  $I_2 > I_1^{\text{crit}}$ , note that  $(\partial/\partial I_2)(\pi_{I_2} - \pi_{q_1, I_2}) > 0$ . QED

This result implies that unconstrained owners face what we refer to as an “inaction zone” of  $(I_1, I_1^{\text{crit}}]$ . If infrastructure is upgraded from  $I_1$  to a level in the inaction zone, owners will not improve their house right away but only when it is exogenously destroyed. But if the infrastructure upgrade is to  $I_2 > I_1^{\text{crit}}$ , unconstrained owners will rebuild at a higher-quality  $q_2$  right away.

The investment problem for constrained owners is similar to that of unconstrained owners, except that the maximum quality they can build is  $q_L$ . As a result, in equilibrium they build  $q_L$  if their plot is empty, and otherwise they do not rebuild.

### C. Neighborhood Development

#### 1. De Novo Areas

De novo areas begin empty with infrastructure ( $I_H$ ). Constrained owners build  $q_L$ , so that the share of low-quality houses is  $1 - \theta_D$ . If  $\theta_D \geq \phi$ , then there is no deterioration in equilibrium, anticipated infrastructure is  $I_H$ , and the unconstrained owners build  $q_H$ . If  $\theta_D < \phi$ , then there is deterioration in equilibrium, anticipated infrastructure is  $I_M$ , and unconstrained owners build  $q_M = q(I_M)$ . In practice, it seems that de novo areas’ infrastructure is better than other areas’ (table 4), suggesting that at least some higher-quality infrastructure survived.

#### 2. Control Areas

Control areas begin empty and with medium-quality infrastructure ( $I_M$ ). Unconstrained owners build housing quality  $q_M$ , while constrained owners build  $q_L$ .

As discussed above, Tanzanian cities also contained areas (which are not part of our main analysis) that are similar to control areas but had a stock of low-quality housing by the time they received infrastructure  $I_M$ . In those areas, the constrained owners keep  $q_L$ , while the unconstrained owners either build  $q_M$  right away (if  $I_M > I_L^{\text{crit}}$ ) or build  $q_M$  only when their house is destroyed.

#### 3. Upgrading Areas

Upgrading areas begin with housing quality  $q_L$  and infrastructure  $I_H$ , and we consider four different cases. In the first case  $I_L^{\text{crit}} > I_H$ , so the upgrading

is minimal, all owners initially keep  $q_L$ , and infrastructure deteriorates to  $I_M$ ; in later periods, as houses are exogenously destroyed, unconstrained owners build to  $q_M$ . In the second case,  $I_H \geq I_L^{\text{crit}} > I_M$  and  $\theta_U < \phi$ , in which case everyone initially keeps  $q_L$ , infrastructure deteriorates to  $I_M$ , and unconstrained owners improve their houses to  $q_M$  when they are destroyed. In the third case,  $I_M \geq I_L^{\text{crit}}$  and  $\theta_U < \phi$ , in which case unconstrained owners build  $q_M$  right away, constrained owners keep  $q_L$ , and infrastructure deteriorates to  $I_M$ . In the final case  $I_H \geq I_L^{\text{crit}}$  and  $\theta_U \geq \phi$ , so unconstrained owners build  $q_H$ , infrastructure remains  $I_H$ , and constrained owners keep  $q_L$ . But in practice, this final case is unlikely to be relevant, because upgrading areas were targeted as poor.

#### *D. Relating the Model to the Empirical Analysis*

The model demonstrates the role of differing infrastructure investments and owner sorting in accounting for neighborhood quality. For example, consider the following scenario. De novo areas had enough unconstrained owners to ensure that their higher-quality infrastructure ( $I_H$ ) survived. In this case, the difference in the logarithm of mean housing quality between de novo and control areas is

$$\ln(\theta_D q_H + (1 - \theta_D) q_L) - \ln(\theta_C q_M + (1 - \theta_C) q_L). \quad (8)$$

This quality difference reflects both the effect of the higher infrastructure quality in de novo areas and the different composition of owners in those areas. But controlling for owner fixed effects allows us to focus on houses owned by unconstrained owners, for whom the difference in log mean housing quality between de novo and control areas is

$$\ln(q(I_H)) - \ln(q(I_M)). \quad (9)$$

In other words, under the model's assumptions, adding owner fixed effects allows us to identify the effect of de novo investments on housing quality for unconstrained owners. We acknowledge that in practice adding owner fixed effects may not solve all the potential problems, if, for example, some owners are constrained in investing in a second house (but not in the first) or have some different preferences for investing across areas. Nevertheless, the model shows that adding owner fixed effects is useful in the context of Sites and Services, where owners in different areas may have had different levels of wealth, due both to sorting and to the program's loans scheme.

The model also allows us to relate differences in infrastructure and housing quality, which we cannot measure directly, to the estimated differences in the value of housing, which are approximated by the hedonic

regressions, subject to the limitations discussed in section II. Specifically, our model predicts the following:

**Proposition 2.** For unconstrained owners who face no risk of exogenous house destruction ( $d = 0$ ),

$$\ln(I_H) - \ln(I_M) = \frac{\gamma - \alpha}{\gamma - \alpha\gamma} (\ln(\pi(q(I_H), I_H)) - \ln(\pi(q(I_M), I_M))), \quad (10)$$

and

$$\ln(q(I_H)) - \ln(q(I_M)) = \frac{1}{\gamma} (\ln(\pi(q(I_H), I_H)) - \ln(\pi(q(I_M), I_M))). \quad (11)$$

*Proof.* To derive the expression for  $\ln(I_H) - \ln(I_M)$ , use (5) and the fact that  $\pi(q_2, I_2) = \pi(0, I_2) + c(q_2)$  and then plug in  $d = 0$  to obtain  $\ln(\pi(q_2, I_2)) = \ln(q_2^\alpha I_2^{1-\alpha}) - \ln(1 - \delta)$ . Next, apply a similar calculation for  $\ln(\pi(q_1, I_1))$ , and plug in (4) to calculate  $\ln \pi(q_2, I_2) - \ln(\pi(q_1, I_1))$ . Now combine the expression for  $\ln(I_H) - \ln(I_M)$  with (4) to derive the expression for  $\ln(q(I_H)) - \ln(q(I_M))$ . QED

This result indicates that the differences across areas in log housing quality are smaller than the differences in log values. Taking the abovementioned estimate of  $\gamma$  suggests that the quality differences across neighborhoods are about  $(1/\gamma) = 0.65$  times the value differences for unconstrained owners, for low values of  $d$ , which seem empirically relevant. Our baseline estimate of the hedonic log value differences between de novo and control areas, with owner fixed effects, are around 0.5, suggesting log quality differences of around one-third.<sup>56</sup>

The model also allows us to consider differences between upgrading and control areas. As discussed in section III, upgrading areas look similar or, in some cases, worse than control areas. Table A10 suggests that poorer-quality housing in upgrading areas may in part be explained by owner fixed effects. In the context of the model, this may reflect persistence in upgrading areas of some of the initial owners (or their descendants) who were targeted by the program and may have been poorer than their counterparts in control areas ( $\theta_U < \theta_C$ ).

Finally, the similarity of housing quality in upgraded and nonupgraded slums is also consistent with the model, if we think of the nonupgraded slums as control areas with constrained owners.

### *E. Implications of the Model*

The model offers several implications for thinking about infrastructure investments for housing. First, an important theme of the paper is that

<sup>56</sup> As discussed in sec. II, the log value differences in the hedonic regressions may understate the actual value differences.

infrastructure investment may crowd in private investments. The model helps us to think about the conditions under which this takes place. In the model, infrastructure investments crowd in more private investments when its quality is sufficiently high and owners can afford to invest in housing quality. In these cases, private investment in housing quality takes place when there is a sufficient fraction of unconstrained owners, either due to their own wealth or through loans that allow them to invest. This also suggests a note of caution: if *de novo* investments were expanded widely, poor and credit-constrained residents may be unable to make full use of them, since infrastructure may deteriorate without sufficient complementary private investment.

Second, the model helps us think about the benefits of early infrastructure investments compared with *ex post* infrastructure upgrading. Upgrading areas do not always fully benefit from high levels of infrastructure investments, since in those settings infrastructure either deteriorates or leads to the scrapping of existing houses.<sup>57</sup>

Finally, turning back to our empirical findings, the model can help explain why infrastructure survived better in *de novo* areas but not in upgrading areas. The model highlights the importance of feedback from owner investments to infrastructure, which can be seen as a neighborhood externality and is sometimes overlooked when infrastructure investments are made.

## V. Concluding Remarks

This paper examines the consequences of different strategies for developing basic infrastructure for residential neighborhoods. Specifically, we study the Sites and Services projects implemented in seven Tanzanian cities during the 1970s and 1980s. These projects provided basic infrastructure, including formal plots, leaving it to residents to build their own houses. We examine the long-run development of these neighborhoods, emphasizing the comparison between *de novo* neighborhoods and other nearby areas that were greenfields when the Sites and Services program started. We also provide descriptive evidence on the development of neighborhoods whose infrastructure was upgraded.

We use high-resolution imagery and building-level survey data to study housing quality and infrastructure in the *de novo* neighborhoods and other areas in their vicinity that were also greenfields to begin with. We find that

<sup>57</sup> In reality, there are other costs of delivering infrastructure in a dense settlement that has developed organically, because it is difficult to resolve coordination failures and negative externalities once they have been put in place. In Sites and Services, e.g., the cost per square meter was similar in *de novo* and upgrading areas, even though *de novo* areas received formal plots that upgrading areas did not.



the de novo neighborhoods developed significantly higher-quality housing than other initially unbuilt areas. Our findings reflect complementary private investments that were made in response to the Sites and Services programs. We also present evidence that the initial infrastructure investments in roads and water mains were more likely to persist in de novo areas. For three cities where we have survey data, we find sizable gains in quality from de novo areas even when we control for owner fixed effects, although these fixed effects account for up to a third of the average housing quality. Our findings suggest that de novo areas increased local land values by at least \$75–\$100 per square meter, compared with total costs of no more than \$8–\$13 per square meter (all in 2017 prices).

We also report evidence that de novo neighborhoods attract more-educated residents, who can afford to pay for the higher quality on offer. But as of 2012, almost half of the adults in de novo areas still had no more than a primary school education, suggesting that some people with lower lifetime incomes also benefitted from the de novo investments. But we also note that de novo areas were unaffordable to the poorest of the urban poor, a consideration that future projects may want to take into account, perhaps by creating some smaller and more affordable plots. Such plots may also benefit the few people who may be displaced by such projects, even when they target largely empty areas.

Our paper also reports descriptive evidence on upgrading areas, comparing them with nearby control areas or, where the data permit, to slums that were not upgraded. The results suggest that upgrading areas now have either similar or worse housing quality, and the program's investments in roads and water mains did not survive well in upgrading areas. While we should be cautious in interpreting these results, they suggest that upgrading, at least as implemented in Sites and Services, was not a panacea for preexisting squatter areas. We cannot rule out that other upgrading efforts may prove more successful, but in order to provide long-lasting benefits, upgrading programs should aim to address the risk of infrastructure deterioration.

Taken together, our findings suggest that de novo investments are a policy tool worthy of consideration for growing African cities. They are considerably cheaper than building public housing and therefore more affordable for poor countries. They also offer important advantages to residents, who can invest in higher-quality housing. Our findings also suggest that it is important to ensure that the infrastructure investments do not deteriorate as a result of poor private investments. While the implementation of Sites and Services projects in Tanzania in the 1970s and 1980s was not flawless, it has taught us important lessons. We hope that these lessons can inform future planning and investment decisions in a continent that is growing in both population and income per capita but where many poor people still live in poor-quality buildings and neighborhoods.

## References

- Ali, Daniel Ayalew, Matthew Collin, Klaus Deininger, Stefan Dercon, Justin Sandefur, and Andrew Zeitlin. 2016. "Small Price Incentives Increase Women's Access to Land Titles in Tanzania." *J. Development Econ.* 123:107–22.
- Angel, Shlomo. 2012. *Planet of Cities*. Cambridge, MA: Lincoln Inst. Land Policy.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Econ. J. Appl. Econ.* 7 (1): 22–53.
- Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *J.P.E.* 115 (4): 588–638.
- Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen. 2011. "Inference with Dependent Data Using Cluster Covariance Estimators." *J. Econometrics* 165 (2): 137–51.
- Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *Q.J.E.* 127 (2): 587–644.
- Brinkhoff, Thomas. 2017. "City Population." <https://www.citypopulation.de>.
- Buckley, Robert M., and Jerry Kalarickal. 2006. *Thirty Years of World Bank Shelter Lending: What Have We Learned?* Washington, DC: World Bank.
- Castells-Quintana, David. 2017. "Malthus Living in a Slum: Urban Concentration, Infrastructure and Economic Growth." *J. Urban Econ.* 98:158–73.
- Cohen, Michael, Callisto E. Madavo, and Harold Dunkerley. 1983. "Learning by Doing: World Bank Lending for Urban Development 1972–82." World Bank, Washington, DC. <http://documents.worldbank.org/curated/en/220591468140401312/Learning-by-doing-World-Bank-lending-for-urban-development-1972-82>.
- Collin, Matthew, Justin Sandefur, and Andrew Zeitlin. 2015. "Falling off the Map: The Impact of Formalizing (Some) Informal Settlements in Tanzania." Working paper, Centre for the Study of African Economies, Univ. Oxford.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2016. "The Production Function for Housing: Evidence from France." IZA Discussion Paper no. 1037, Institute of Labor Economics, Bonn, Germany.
- Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *J. Econometrics* 92 (1): 1–45.
- Coville, Aidan, and Yu-hsuan Su. 2014. "From the Ground Up: An Impact Evaluation of the Community Infrastructure Upgrading Programme in Dar-es-Salaam." Working paper, World Bank, Washington, DC.
- Dar Ramani Huria. 2016. "Community-Based Mapping Project in Dar-es-Salaam." <http://ramanihuria.org/data/>.
- Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78 (6): 1863–903.
- Diewert, W. Erwin, Jan de Haan, and Rens Hendriks. 2015. "Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components." *Econometric Rev.* 34 (1–2): 106–26.
- DigitalGlobe. 2016. "WorldView Satellite Imagery." <http://worldview3.digitalglobe.com>.
- Duranton, Gilles, and Diego Puga. 2015. "Urban Land Use." In *Handbook of Regional and Urban Economics*, vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 467–560. Amsterdam: North-Holland.
- Duranton, Gilles, and Anthony J. Venables. 2020. "Place-Based Policies for Development." In *Handbook of Regional Science*, edited by Manfred Fisher and Peter Nijkamp. Berlin: Springer.

- Field, Erica. 2005. "Property Rights and Investment in Urban Slums." *J. European Econ. Assoc.* 3 (2–3): 279–90.
- Franklin, Simon. 2020. "Enabled to Work: The Impact of Government Housing on Slum Dwellers in South Africa." *J. Urban Econ.* 118:103265.
- Freire, Maria E., Somik Lall, and Danny Leipziger. 2014. "Africa's Urbanization: Challenges and Opportunities." Working Paper no. 7, Growth Dialogue. <http://growthdialogue.org/africas-urbanization-challenges-and-opportunities/>.
- Galiani, Sebastian, Paul J. Gertler, Raimundo Undurraga, Ryan Cooper, Sebastian Martinez, and Adam Ross. 2013. "Publisher's Note on Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums." *J. Urban Econ.* 96:166–94.
- Galiani, Sebastian, and Ernesto Schargrodsky. 2010. "Property Rights for the Poor: Effects of Land Titling." *J. Public Econ.* 94 (9): 700–729.
- Gelman, A., and G. Imbens. 2017. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *J. Bus. and Econ. Statist.* 37 (4): 447–56.
- Giglio, Stefano, Matteo Maggiori, and Johannes Stroebe. 2014. "Very Long-Run Discount Rates." *Q.J.E.* 130 (1): 1–53.
- Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2016. "Urbanization with and without Industrialization." *J. Econ. Growth* 21 (1): 35–70.
- Gollin, Douglas, Martina Kirchberger, and David Lagakos. 2017. "In Search of a Spatial Equilibrium in the Developing World." Working Paper no. 23916, NBER, Cambridge, MA.
- Harari, Mariaflavia, and Maisy Wong. 2017. "Long-Term Impacts of Slum Upgrading: Evidence from the Kampung Improvement Program in Indonesia." Working paper, Univ. Pennsylvania.
- Henderson, J. Vernon, Tanner Regan, and Anthony J. Venables. 2017. "Building the City: Urban Transition and Institutional Frictions." Discussion Paper no. 11211, CEPR, London.
- Henderson, J. Vernon, Anthony J. Venables, Tanner Regan, and Ilia Samsonov. 2016. "Building Functional Cities." *Science* 352 (6288): 946–47.
- Hornbeck, Richard, and Daniel Keniston. 2017. "Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872." *A.E.R.* 107 (6): 1365–98.
- Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Rev. Econ. Studies* 79 (3): 933–59.
- Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–94.
- Kironde, J. M. Lusugga. 1991. "Sites-and-Services in Tanzania: The Case of Sinza, Kijitonyama and Mikocheni Areas in Dar-es-Salaam." *Habitat Internat.* 15 (1–2): 27–38.
- . 1992. "Creations in Dar-es-Salaam and Extensions in Nairobi: The Defiance of Inappropriate Planning Standards." *Cities* 9 (3): 220–31.
- . 1994. "The Evolution of the Land Use Structure of Dar es Salaam 1890–1990: A Study in the Effects of Land Policy." PhD diss., Univ. Nairobi.
- . 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* 3:348–67.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Komu, Felician. 2013. "Rental Housing in Tanzania: Power and Dialects of Misidentification." *J. Building and Land Development*, special issue: 29–41.

- Laquian, Aprodicio A. 1983. "Sites, Services and Shelter: An Evaluation." *Habitat Internat.* 7 (5–6): 211–25.
- Libecap, Gary D., and Dean Lueck. 2011. "The Demarcation of Land and the Role of Coordinating Property Institutions." *J.P.E.* 119 (3): 426–67.
- Manara, Martina, and Tanner Regan. 2019. "Eliciting Demand for Title Deeds: Lab-in-the-Field Evidence from Urban Tanzania." Discussion Paper no. 19, LSE Geography and Environment, London.
- Marx, Benjamin, Thomas Stoker, and Tavneet Suri. 2013. "The Economics of Slums in the Developing World." *J. Econ. Perspectives* 27 (4): 187–210.
- . 2019. "There Is No Free House: Ethnic Patronage in a Kenyan Slum." *American Econ. J. Applied Econ.* 11, no. 4 (2019): 36–70.
- Mayo, Stephen K., and David J. Gross. 1987. "Sites and Services and Subsidies: The Economics of Low-Cost Housing in Developing Countries." *World Bank Econ. Rev.* 1 (2): 301–35.
- McMillen, Daniel P., and Christian L. Redfearn. 2010. "Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions." *J. Regional Sci.* 50 (3): 712–33.
- MMM (Marshall Macklin Monaghan). 1979. "Main Report: Five Year Development Programme." In *The Dar-es-Salaam Master Plan*. MMM, Toronto.
- NOAA (National Oceanic and Atmospheric Administration). 2012. "Version 4 DMSP-OLS Nighttime Lights Time Series." <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>.
- Owens, Kathryn. 2012. "Enabling Sustainable Markets? The Redevelopment of Dar-es-Salaam." Paper prepared for the World Bank's Sixth Research and Knowledge Symposium, Barcelona, October 8–10.
- Painter, K., and D. P. Farrington. 1997. "The Crime Reducing Effect of Improved Street Lighting: The Dudley Project." In *Situational Crime Prevention: Successful Case Studies*, 2nd ed., edited by R. V. Clarke, 209–26. Guilderland, NY: Harrow & Heston.
- Picarelli, Nathalie, Pascal Jaupart, and Ying Chen. 2017. "Cholera in Times of Floods: Weather Shocks & Health Impacts in Dar es Salaam." Working Paper no. C-40404-TZA-1, Internat. Growth Centre, London.
- Redding, Stephen J., and Daniel M. Sturm. 2016. "Estimating Neighborhood Effects: Evidence from War-Time Destruction in London." Working paper, Princeton Univ.
- Romer, Paul. 2010. "Technologies, Rules, and Progress: The Case for Charter Cities." Center for Global Development. <https://www.cgdev.org/publication/technologies-rules-and-progress-case-charter-cities#:~:text=Non%2Dresident%20fellow%20Paul%20Romer,scarce%20resources%20nor%20limited%20technologies>.
- . 2012. "Urbanization as Opportunity." OECD. <http://www.oecd.org/cfe/regional-policy/Urbanization%20as%20Opportunity%20-%20Paul%20Romer.pdf>.
- Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens III. 2010. "Housing Externalities." *J.P.E.* 118 (3): 485–535.
- Tanzania Ministry of Lands. 2012. "Rates Land Value Mikoa (Regions) 10 2012." Ministry of Lands, Housing and Human Settlements Development, Dar es Salaam.
- Tanzania National Bureau of Statistics. 2011. "Tanzania Population and Housing Census 2002." Tanzania National Bureau of Statistics, Dar es Salaam.
- . 2014. "Tanzania Population and Housing Census 2012." Tanzania National Bureau of Statistics, Dar es Salaam.
- . 2017. "2012 Census Shapefiles (machine readable data files)." Tanzania National Bureau of Statistics, Dar es Salaam.

- Theodory, T. F., and M. M. Malipula. 2012. "Supplying Domestic Water Services to Informal Settlements in Manzese, Dar es Salaam: Challenges and Way Forward." *J. Rural Planning Assoc.* 14 (2): 60–72.
- Tiba, A. D., G. Mwarabu, W. H. Sikamkono, et al. 2005. "The Implication of 20,000 Plots Project on the Emerging Form of Dar es Salaam City." MSc semester project, Dar es Salaam: Dept. Urban Planning and Management, UCLAS, UDSM.
- Turner, Matthew A., Andrew Haughwout, and Wilbert Van Der Klaauw. 2014. "Land Use Regulation and Welfare." *Econometrica* 82 (4): 1341–403.
- UN (United Nations). 2015. "World Population Prospects." Population Division, Dept. Econ. and Social Affairs, New York.
- UN-Habitat. 2012. "State of the World's Cities 2012/2013: Prosperity of Cities." <https://unhabitat.org/prosperity-of-cities-state-of-the-worlds-cities-20122013>.
- World Bank. 1974a. "Appraisal of National Sites and Services Project." Report no. 337a-TA, Urban Projects Dept., World Bank, Washington, DC.
- . 1974b. "Development Credit Agreement between United Republic of Tanzania and International Development Association (Conformed Copy)." Credit no. 495 TA, World Bank, Washington, DC.
- . 1977a. "Development Credit Agreement between United Republic of Tanzania and International Development Association (Conformed Copy)." Credit no. 732 TA, World Bank, Washington, DC.
- . 1977b. "Tanzania: The Second National Sites and Services Project." Report no. 1518a-TA, Urban Projects Dept., World Bank, Washington, DC.
- . 1984. "Completion Report: Tanzania—First National Sites and Services Project." Report no. 4941, Eastern Africa Regional Office, World Bank, Washington, DC.
- . 1987. "Tanzania: The Second National Sites and Services Project." Report no. 6828, Operations Evaluation Dept., World Bank, Washington, DC.
- . 2010. "Project Appraisal Document for a Tanzania Strategic Cities Project." Report no. 51881-TZ, Urban and Water Dept., World Bank, Washington, DC.
- . 2013. "Tanzania Strategic Cities Project Housing Survey." Produced by President's Office, Regional Administration and Local Government, World Bank, Washington, DC.