Slum Upgrading and Long-run Urban Development: Evidence from Indonesia *

Mariaflavia Harari†
University of Pennsylvania

Maisy Wong‡
University of Pennsylvania

Preliminary and incomplete
February 2018

Abstract

The United Nations estimates that a quarter of the world's urban population lives in slums. This paper sheds light on how a developing country city grows out of informality, through the lens of one of the largest slum upgrading programs in the world. The 1969-1984 Kampung Improvement Program (KIP) provided basic public goods in slums, covering 5 million people and 25% of the city of Jakarta, Indonesia. We assemble a granular database with program boundaries, historical maps, current land values, building heights, measures of land fragmentation, and a novel quality index of informal settlements based on Google Street View and field photos. Our research design compares KIP areas with historical slums that were never treated. Our findings are similar using a boundary discontinuity design. KIP areas today have 12% lower land values and buildings with 1.6 fewer floors on average, implying aggregate impacts of US$11 billion. Greater land fragmentation in KIP areas points towards the importance of land assembly costs as a barrier to formalization. These long-term costs need to be weighed against the benefits of the program. Overall, our findings suggest slum upgrading may be more cost effective for cities in early stages of urban development.

*We are grateful to Darrundono, Gilles Duranton, Marja Hoek-Smit, and Ben Olken for their advice. We thank participants at the Asian Development Review conference, the Harvard-IGC conference, the Wharton applied economics workshop, the NBERSI Urban workshop, the University of Delaware economics seminar. Kania Azrina, Xinzhu Chen, Gitta Djuwadi, Krista Iskandar, Jeremy Kirk, Melinda Martinus, Joonyup Park, Yuan Pei, Xuequan Peng, Arliska Fatma Rosi, Beatrix Siahaan, Vincent Tanutama, Ramda Yanurzha were excellent research assistants. We thank the Research Sponsors Program of the Zell/Lurie Real Estate Center, the Tanoto ASEAN Initiative, and the Global Initiatives at the Wharton School. All errors are our own.

†Wharton Real Estate. 3620 Locust Walk, 1467 SHDH, Philadelphia, PA 19104-6302. Email: harari@wharton.upenn.edu.

‡Wharton Real Estate. 3620 Locust Walk, 1464 SHDH, Philadelphia, PA 19104-6302. Email: maisy@wharton.upenn.edu.

1 Introduction

The United Nations estimates that 2.5 billion people will be added to cities by 2050 and that 95% of the urban expansion in the coming years will be in the developing world.[1] Rapid urban growth is often accompanied by the proliferation of slums, which are estimated to host a quarter of the world's urban population (UN-Habitat, 2012). Informal settlements with inadequate housing and public services are a persistent feature of urban landscapes in developing economies. While slums may provide an entry point for migrants seeking economic opportunities in cities, concerns arise about the worsening of living conditions in slums, which may limit economic mobility and agglomeration benefits (Glaeser (2011); Marx, Stoker, and Suri (2013)).

One widely implemented policy approach is to improve living conditions in slums through slum upgrading programs.[2] These programs provide local infrastructure *in situ* (on-site), including roads, and health, water, and sanitation facilities.[3] In the short run, improvements enhance neighborhood quality, raising land values (Kline and Moretti, 2014). In the long run, however, the interaction of such programs with frictions in urban land markets may make treated informal settlements more persistent than they would otherwise be, delaying the formalization of treated neighborhoods (Fox, 2014). This can distort the spatial allocation of development activity and give rise to opportunity costs of land use, that have to be weighed against a program's short-run impacts.

We shed light on how cities grow out of informality, through the lens of one of the earliest and largest scale slum upgrading programs in the world. The Kampung Improvement Program (KIP) in the city of Jakarta, Indonesia, covered 5 million beneficiaries and 25% of Jakarta.[4] We study the first three waves of KIP, implemented from 1969 to 1984 to improve roads and basic public goods. We provide novel causal estimates of the long-term impacts of KIP on the urban development of

[1] Between 2014 and 2050, 404 million people are projected to be added to Indian cities, followed by 292 million in China, and 212 million in Nigeria.

[2] For example, India and Indonesia recently announced slum upgrading programs to benefit 18 million and 10 million slum residents, respectively. These programs are the Pradhan Mantri Awas Yojana (PMAY) in India and the Kota Ku program in Indonesia (World Bank, 2016b; Government of India Ministry of Housing & Urban Poverty Alleviation, 2016). Other programs include the Favela-Barrio project in Brazil, the PRIMED project in Colombia, and programs in Bangladesh, Tanzania, Kenya, and Ghana (UN Habitat, 2011; World Bank, 2017; UN Habitat, 2017).

[3] Common alternatives to on-site slum upgrading include titling, migration restrictions, evictions, relocations into public housing, and "sites and services" programs, which provide basic public goods and services in newly created settlements when land is available.

[4] The word *kampung* refers to a traditional village in Indonesian, but it is also used to refer to informal settlements in urban areas. Unless stated otherwise, we will use *kampung*, slum, and informal settlement interchangeably.

Jakarta, a city that has started formalizing and that thus provides a unique setting to assess the long-run opportunity costs of land use.

One of the strengths of our research design is in the granularity and richness of the data. We combine unique KIP program maps that indicate the boundaries of upgraded kampung's with historical maps of kampung's before the launch of KIP. Additionally, we obtain a new administrative database of current assessed land values for 20,000 locations (sub-blocks) that span the entire city of Jakarta. Finding comprehensive and accurate data of land values is notoriously difficult for developing countries, but we validate these administrative data using transaction prices scraped from a property website. Our third novel data source is a photographic survey of a representative sample of locations in Jakarta, that combines publicly available Google Street View imagery and photos taken in the field. We divide Jakarta into 75 by 75 meter pixels and obtain photos for over 7000 pixels, for a total of over 28 thousand photos. From the photos we manually code building heights and an index of informality based on the physical aspects of the built environment. This helps us overcome one of the difficulties in this literature, namely the measurement of slums. Finally, we obtained digital maps detailing the outlines of more than a million land parcels in Jakarta, both in formal and informal areas, including information on land use. We use this database to investigate the impact of KIP on land use patterns and to quantify land fragmentation, as proxied by the number of parcels in each 75 meter pixel. We complement the above dataset with auxiliary data for topographic, distance, and historical attributes of locations.

Our primary outcomes of interest are current land values and building heights. Leveraging our granular datasets, we develop research designs that compare KIP areas to other observably identical ones, overcoming several empirical challenges.[5] As a preliminary step, we address omitted variable bias concerns by showing that KIP and non-KIP areas become relatively comparable once we restrict our comparison to KIP versus non-KIP areas nearby.

Our first identification strategy is to focus on areas that were kampung's before the implementation of the program, and compare those treated by KIP with those that were not treated, within administrative units comparable to census tracts in the United States. This addresses program selection bias and the concern that KIP targeted slums only.

Our second strategy is a boundary discontinuity design that compares observations within 200

[5] Field and Kremer (2008) discuss empirical evaluations of slum upgrading programs and difficulties that are particularly salient in urban environments, including spillovers, contamination of control areas, endogenous sorting, and general equilibrium effects.

meters on both sides of the program boundaries. The identifying assumption is that omitted variables do not change discontinuously at the boundary. We take additional steps to address concerns that the boundary discontinuity design may be too localized, and may suffer from spillovers, contamination, and displacement effects. Our results are remarkably stable using different buffer distances (200 to 500 meters). Furthermore, we do not see spatial decay patterns up to 1000 meters away from the KIP boundary, suggesting limited spatial externalities (Turner, Haughwout, and Van Der Klaauw, 2014). To complement our localized boundary analysis, we also present a more comprehensive analysis that encompasses all of Jakarta, controlling for very fine hamlet fixed effects (analogous to census block group fixed effects for the United States).

We find that KIP areas today have 12% lower land values and buildings with 1.6 fewer floors on average. Reassuringly, our estimated effects are stable in magnitude across various specifications. Given the total coverage of KIP (100 million square meters or 10,000 hectares) and current land values in historical kampung's that did not receive KIP, the 12% effect translates into an aggregate impact of US$11 billion. We also find that KIP areas are 12 percentage points less likely to have high-rises (buildings with more than three floors). Given the price premium associated with high-rises, we calculate that about 88% of the overall effect on land values is explained by the lack of tall buildings in KIP areas.

We rationalize these findings through a simple model of spatial equilibrium with frictions in the development process. We model such frictions as a fixed cost that developers face when they supply formal buildings. Formalization costs include land assembly costs, that are notoriously high in settings with weak property rights such as ours. In the model, as KIP improves amenities in a neighborhood, in the short-run this should increase local land values, as predicted by the literature on place-based policies in developed countries. If, however, KIP also increases formalization costs in treated areas, in the long run, these neighborhoods will have formalized less than control neighborhoods, resulting in relatively lower land values and building heights. In fact, there is a clear bunching pattern in the distribution of building heights with a majority of buildings in KIP areas having fewer than three floors. Formalization costs may be higher in KIP areas due to entitlement effects and a perception of more entrenched occupancy rights of residents, which increased the perceived costs of land assembly for developers. This is in line with anecdotal evidence from KIP evaluation reports.

Consistent with our proposed mechanism, we show that KIP areas today are between 14 and 20 percentage points more likely to be informal as measured by our index. Furthermore, they have a greater degree of land fragmentation, as measured by the number of parcels in each pixels, which we interpret as a proxy for land assembly costs (Brooks and Lutz, 2016). In order to assemble 5600 square meters, roughly the land area needed for a high-rise, developers in KIP areas need to acquire, on average, between 9 and 12 more parcels. A back-of-the-envelope exercise suggests that greater land fragmentation in KIP areas explains about three quarters of the effect of KIP on land values and a third of the effect on building height. We also find that KIP areas have fewer retail and office density, which we interpret as amenities associated with formal neighborhoods. Furthermore, the impacts of KIP are more pronounced in central areas, where the potential gains from formalizing are the greatest.

Our results are not consistent with alternative channels such as differences in public amenities or direct effects of KIP-provided public goods: we find similar access to schools and health centers across KIP and non-KIP areas, suggesting that the neighborhoods have converged in terms of public amenities. We also find no differential effect for areas near the original KIP investments, consistent with policy reports suggesting the initial investments have likely depreciated.

We present further tests to address additional threats to identification. The main one is negative selection of slums into KIP: the program selection rules prioritized kampung's in worse conditions, which may confound our results of negative effects of KIP. Exploiting the staggered rollout of KIP, we show that land values and heights are not differentially lower for the first wave of KIP, which targeted areas in poorer conditions, compared to the second and third waves of KIP.

Another potential confounder that could affect our boundary discontinuity exercise, is the persistent effect of being a slum. We address this by comparing outcomes on either side of the boundary of historical kampung's in non-KIP areas and showing the effects are insignificant.

Taken together, our results indicate that in the long-run KIP is associated with an opportunity cost of land use related to the delayed formalization of KIP areas. This long-run cost needs to be weighted against the program's benefits. Beyond the direct effects of KIP-provided amenities - such as sanitation - on the quality of life and human capital of the original program beneficiaries, the delayed formalization induced by KIP also benefited kampung dwellers by granting them access to jobs in central areas of Jakarta for a longer time relative to their non-KIP counterparts. In future work we will provide bounding exercises to shed light on the trade-off between short-run benefits and long-run costs of the program, acknowledging that the counterfactual for residents of kampung's is plausibly displacement to peripheral areas with poorer job access. This will shed light on the reasons for the slow pace of the formalization process in Jakarta.

Overall, our findings suggest that policy makers considering slum upgrading should account for long-run costs that may manifest themselves once a city starts formalizing. In general, the trade-off between benefits and costs will be affected by a city's stage of development: for a city that is already undergoing formalization, the opportunity costs of land use may be too large relative to the benefits. On the other hand, in a city that is comparable to Jakarta at the time of KIP,[6] slum upgrading may offer an attractive cost-benefit balance, especially if additional measures are taken in the interim to facilitate the formalization process and minimize future distortionary costs. Our estimates of the long-term impacts of slum upgrading can thus provide valuable inputs for policy makers beyond Indonesia.

This paper makes four contributions to the literature. To our knowledge, we provide the first causal estimates of the long-term impacts of a large *in situ* slum upgrading program. Second, we shed light on the formalization process and barriers to formal development, in particular land assembly costs, for which we provide a proxy based on land fragmentation. Third, we assemble a uniquely rich, granular, and comprehensive database covering policy data, land values, and urban structures in both formal and informal areas in a developing country mega-city. Fourth, we contribute to one of the main gaps in the literature on informal settlements by providing a measure of informality through an innovative approach that leverages publicly available imagery, complemented with field photos.

Our work is related to two literatures. The first is a small but growing literature on slums in developing countries (see Brueckner and Lall (2015) and Marx, Stoker, and Suri (2013) for an overview). There are relatively few papers on slum upgrading programs (see Field and Kremer (2008) for a review).[7] Michaels et al. (2017) contrast the "sites and services" approach to *in situ* slum upgrading in Tanzania. Feler and Henderson (2011) and Castells-Quintana (2017) investigate the provision of public goods and services in informal settlements, while Marx, Stoker, and Suri (2016) examine the influence of ethnic patronage in shaping private investment decisions in slums. A number of related papers investigate other policy approaches, including titling (Field, 2007; Galiani and Schargrodsky, 2010) and relocations from slums to public housing (Barnhardt,

[6]When KIP was implemented, the country was at a level of development comparable, in terms of GDP per capita, to that of Bangladesh, Kenya, Pakistan, and Tanzania today. Between 1969 and 1984, the average GDP per capita for Indonesia was around $950. Today, the countries above have GDP per capita of $970 (Bangladesh), $1133 (Kenya), $1143 (Pakistan), $842 (Tanzania). All dollar amounts are 2010 US dollars.

[7]Takeuchi, Cropper, and Bento (2008) consider the welfare impacts of a slum improvement program, and Kaufmann and Quigley (1987) and Galiani et al. (2017) study the impact of improving housing in slums.

Field, and Pande, 2017). Finally, Henderson, Venables, and Regan (2016) examine the dynamics of spatial development in a city, highlighting how frictions in the formalization process lead to misallocation of land. Our results emphasize similar barriers to formalization and misallocation patterns, leveraging policy variation from KIP.

Our research also draws upon the broader literature on urban development. We highlight development costs that can potentially distort spatial development patterns, echoing other research on land assembly costs (Brooks and Lutz, 2016), property rights, and land fragmentation patterns (Libecap and Lueck, 2013), zoning and land use regulations in the United States (Turner, Haughwout, and Van Der Klaauw, 2014) and in India (Duranton et al., 2015). Finally, a number of papers have examined the impacts of urban renewal programs on land values in US cities (Collins and Shester, 2013; Rossi-Hansberg, Sarte, and Owens, 2010), finding evidence of positive effects and positive externalities.

The rest of the paper proceeds as follows. Section 2 describes the background, Section 3 outlines the conceptual framework, Section 4 describes the data, Section 5 presents the empirical strategy, Section 6 presents our main results, Section 7 explores mechanisms, and Section 8 addresses the main threats to the identification. Section 9 concludes.

2 Background

Indonesia is the fourth most populous country in the world, with a population of 240 million, and has grown into a middle income country, with a current GDP per capita of $3800 (2010 USD). The country is currently in the midst of structural transformation.[8] The urban sector is rapidly growing at a rate of 4% per year, with slightly more than half of the population living in cities now and more than two thirds expected by 2025 (World Bank, 2016a).

2.1 Urban development and slums in Jakarta

Jakarta is a mega-city with a population of almost 11 million inhabitants, expected to grow to 16 million by 2020, and a land area of approximately 260 square miles. Beyond the urban core, the city expands in a sprawling metropolitan area home to 30 million inhabitants - the second-largest

[8] Between 1965 and 2014 the manufacturing share of GDP increased by 19 percentage points while the agricultural share fell by 35 percentage points (World Bank, 2016a).

in the world. Jakarta is the primate city not only in demographic but also in economic terms: the annual gross regional product per capita is US$14,000 as of 2015 - about 4 times the national average - and the city contributes to 1/7th of the Indonesian gross domestic product.

The urban development of Jakarta can be traced back to colonial times. Located in a coastal lowland, the city was founded in 1619 by the Dutch. The colonial administration established a system of well-defined land rights only in those parts of the city directly settled by the Dutch (*bewoude-kom* or "built-up" areas), leaving local customary land rights (*adat* law) in place elsewhere. These non-built-up areas, settled by the local population, took the traditional form of "kampung's" or urban villages.

After independence in 1945, the transition from customary land rights to private property was never fully completed and many areas of Jakarta are still under adat law. The latter encompasses a wide spectrum of rights (such as the right to build or the right to use) that typically entail a limited degree of tenure security and that are often difficult to codify and enforce (Leaf, 1993). During the decades post-independence, the city faced heavy in-migration of low-income households. Concerned with overcrowding, policy makers sought to alleviate congestion in the city's kampung's through slum upgrading and by declaring Jakarta a "closed city" in 1970. In the subsequent decades, between 1980 and 2002, almost one-quarter of all land within Jakarta city limits was converted from non-urban uses for agriculture or wetlands and water (World Bank, 2011). As the city grew and land became more and more scarce, in the 1990s some of the kampung's started being redeveloped into formal neighborhoods.

The process of formalization begins with the selection of a suitable location, based on market potential, proximity to roads, and presumed level of compensation to be paid by current residents (World Bank, 1995). The next phase is the land assembly, perceived as the main barrier to formalization from the developers' standpoint (Rakodi and Firman, 2009). This involves identifying existing claimants, verifying their land tenure status (sometimes unclear to the residents themselves) and negotiating the purchase of individual parcels, often through the mediation of a middle-man (Leaf, 1993).[9] Since property rights are not well-defined, this phase tends to be lengthy and complex, involving disputes over land ownership and hold-up issues. The process also tends to be politically charged, as it raises equity concerns related to the compensation and displacement of current residents. Absent a well-developed housing finance system, low-income households are

[9] To mitigate speculation, parcels are purchased privately until about half of the land is assembled, after which the project is made public and negotiations take place with the local village heads (World Bank, 1995)

priced out of the formal housing sector, and kampung's constitute their main housing option in the city.

With a population growing at a rate of 2.5% a year, Jakarta currently faces a massive housing backlog, which is at the center of the current political debate. Recent government estimates indicate that more than 70,000 units will need to be built per year in order to face the current unmet demand. Although no comprehensive census of slum areas exists, estimates suggest that a quarter of the city's population lives in kampung's today (McCarthy, 2003).

The Jakarta government's approach towards the proliferation of informal settlements has evolved through the years. Beyond closing the city in the 1970s, past policies include slum clearance and slum upgrading. More recently, efforts have centered around public housing and subsidizing the development of low income housing, mostly at the periphery of the city.[10] Related projects include improving infrastructure and public transit systems, with targeted subsidies for the poor, in order to alleviate the city's chronic traffic congestion. There are also plans to add a subway system and other large-scale residential development projects.

2.2 The Kampung Improvement Program

KIP is one of the earliest and largest slum upgrading programs ever implemented in the world. The program in Jakarta covered 11,000 hectares (25% of the city) and 5 million beneficiaries, with a total outlay between $450 million and $550 million (2016 USD). KIP was later expanded to other cities, eventually covering 50,000 hectares and 15 million beneficiaries in Indonesia (World Bank, 1995).

History of KIP. The history of slum upgrading in Jakarta can be traced back to the 1920's, when the Dutch upgraded villages (*kampoeng verbeteering*) surrounding their communities, to protect themselves from the spread of diseases, floods, and fires. Following in-migration into Jakarta after the Dutch left, it was estimated that kampung's constituted 60% of Jakarta's total area and 75% of its population by 1960's. Policy makers were concerned about floods, health hazards, over-crowding, and political riots. At that time, Indonesia was resource-constrained, being one of the poorest countries in the world (poorer than India, Bangladesh, and Nigeria). Therefore, slum

[10]Efforts to promote redevelopment of central slums while rehousing current residents on site have been attempted but appear unpromising, as developers typically do not want to mix affordable housing with luxury developments and the government lacks the political capital to enforce these types of arrangements.

upgrading appeared as an affordable policy tool to provide immediate assistance to a large number of kampung residents at a low cost (Darrundono, 2012).

KIP was implemented in multiple five-year plans (*Pelita*). This study considers the first three waves of KIP in Jakarta, which focused on physical upgrades: *Pelita I* (1969-1974) , II (1974-1979) and III (1979-1984). In 1986 there was a negative budget shock as oil prices plummeted. Beginning in the 1990's, KIP was expanded to other cities and the government also embarked on a new wave of slum upgrading that broadened the scope beyond physical upgrades, to include community building and microcredit programs (Kessides, 1997). However, this expansion was disrupted by the 1997 Asian financial crisis. Recently, the government has expanded slum upgrading efforts again with the KotaKu program, slated to cover 154 cities and 9.7 million beneficiaries.

KIP is generally held up by practitioners and policy makers as an example of a successful slum upgrading program. A 1995 evaluation report by the World Bank concluded that targeted urban improvement programs "can have a very positive impact on low-income areas", in the form of improved quality of life for kampung residents, access to clean water, sanitation, reduced flooding. There was also some evidence that non-KIP kampung's had caught up. Moreover, the demographic composition appeared to be stable, despite concerns about displacement of poor households by rich ones.

Program Details. The primary objective of KIP was to improve neighborhood conditions in kampung's. The government aimed to benefit as many low-income households as possible in the shortest period of time. Given the limited budget, KIP focused on providing basic and relatively inexpensive upgrades on site, distributed over a large coverage area. Planners assumed a useful life of 15 years for the physical upgrades. It is therefore plausible that the initial KIP investments, implemented more than four decades ago, have significantly depreciated by now.

The program was rolled out prioritizing kampung's according to need. Specifically, the selection criteria were neighborhood conditions (e.g. sanitation facilities, flood damage, road quality), age of the kampung, population density, and income. Given the time constraint and limited information, planners prioritized physical components that were easy to observe. Moreover, selected kampung's had to be distributed evenly across the five municipalities of Jakarta (see Darrundono (1997) and Darrundono (2012) for more details about the program).

A typical kampung upgrade required around two years. The process began with the KIP Technical Unit selecting a list of eligible kampung's based on the selection criteria discussed above. The list then had to be approved by a steering committee. Finally, the upgrading plan was contracted out to local contractors.

The physical upgrades provided as part of KIP consisted primarily of three components. First, the government wanted to improve access in kampung's by widening and paving vehicular roads, bridges, and footpaths. Widening roads was of particular importance as policy makers wanted to ensure that ambulances and fire fighters could access slum locations in the event of emergencies.

The second component focused on sanitation and water management. This included the provision of public water supply and sanitation facilities. As many areas in Jakarta are prone to floods, the government also installed drainage canals and waste disposal facilities, to minimize the clogging of waterways. Third, KIP included the construction of community buildings such as primary schools and neighborhood health clinics.

3 Conceptual framework

In this Section we outline a simple framework to characterize how a neighborhood improvement policy will affect land values and building heights, in a context with formal and informal settlements. In particular, we adapt the standard spatial equilibrium model (Rosen, 1979; Roback, 1982) to introduce a friction in land markets in the form of a fixed cost of formalization. Our model illustrates how a neighborhood improvement program can increase land values in treated neighborhoods in the short run, but can lead to lower land values relative to control neighborhoods in the long run when there are formalization costs. This contrasts with the standard prediction in the literature on place-based policies in developed countries (Kline and Moretti, 2014; Glaeser and Gottlieb, 2008).

Below, we first describe housing supply and the formalization process, followed by housing demand and equilibrium in the housing market. Next, we explain how a slum upgrading program will affect land values and building heights.[11] We focus on the contrast between kampung's and formal neighborhoods and relegate other details - that are standard in the Rosen-Roback framework - to the Appendix.

[11] We focus on housing markets to deliver intuition, but our conclusions hold in a full general equilibrium model that incorporates local labor markets as well.

3.1 Housing market and formalization process

We focus on characterizing the formalization process to show how kampung's (K) and formal (F) neighborhoods form endogenously in equilibrium. The two types of neighborhoods will differ by land values (p), building height (h), exogenous amenities (θ), as well as endogenous amenities (β) that arise in equilibrium if a neighborhood formalizes. Developers face a fixed cost if they choose to supply buildings with height above $\bar{h}$, which we denote as "formal". Developers will formalize if prices in formal neighborhoods are high enough to overcome the fixed costs. We capture higher prices in formal neighborhoods by allowing them to have additional endogenous amenities β associated with formalization. In equilibrium, kampung's will have lower land values, shorter buildings, and fewer amenities (given by θ). By contrast, formal neighborhoods will have higher land values, taller buildings, and more amenities (θ and β).

Formally, developers choose how many floors of housing space h to build in a given location.[12] The cost of building at height h is

$$C(h) \quad = \quad h^{\delta} + cI_{[h>\bar{h}]}, \, \delta > 1 \text{ and } c > 0. \tag{1}$$

The key friction in our model is an additional fixed cost for building tall buildings which gives rise to the discontinuous cost function above. To fix ideas, assume $\bar{h}$ is 2 floors. For buildings of 1 or 2 floors, the cost is simply h^{δ}. For buildings over 2 floors, there is an additional fixed cost c. This non-linearity in the costs of building high-rises stems from engineering costs (e.g. steel frames have to be used for taller buildings), compliance to regulations (e.g. minimum lot sizes), and land assembly. While we acknowledge that formal and informal neighborhoods differ in many respects, for the purposes of our model, we define neighborhoods with heights below $\bar{h}$ as "kampung" (K) and neighborhoods with taller buildings as "formal" (F), respectively.

Denoting the price of housing per square meter in a neighborhood as p, the developers' profit maximization problem for a given neighborhood reads:

$$\max_{h} \left\{ ph - h^{\delta} - cI_{[h>\bar{h}]} \right\} \tag{2}$$

Solving the developers' problem delivers a discontinuous supply curve with a flat portion (cor-

[12] Assume all locations have the same amount of land, normalized to 1, so that housing space coincides with building height.

responding to the housing supply of kampung's) and an upward-sloping portion (corresponding to formal neighborhoods):

$$h^S = \begin{cases} \bar{h} \text{ if } p \leq \bar{P}(c) \text{ (kampung)} \\ \left(\frac{p}{\delta}\right)^{\frac{1}{\delta-1}} \text{ if } p > \bar{P}(c) \text{(formal)} \end{cases} \tag{3}$$

where the threshold $\bar{P}$ is increasing in c. Intuitively, developers choose to formalize, i.e. develop at a height greater than $\bar{h}$, when the price of housing p is high enough relative to the fixed cost c, otherwise they choose the corner solution $h = \bar{h}$. In other words, when formalization costs are high, there is a narrower range of prices for which developers are willing to formalize.

Next, let us characterize housing demand. Consumers choose how much housing to consume and which neighborhood to be in. Each neighborhood is characterized by the presence of two types of amenities. There is a bundle of amenities θ -such as local public goods or favorable topography - that are exogenous to the model. There is a second bundle of amenities from formalization - such as malls or office complexes - that are present only in neighborhoods that formalize and are endogenous to the model's equilibrium. The latter are captured by parameter β such that

$$\beta \begin{cases} = 1 \text{ if } h \leq \bar{h} \\ > 1 \text{ if } h > \bar{h}. \end{cases} \tag{4}$$

The utility of consumers depends positively on income (which we assume is exogenous and normalized to 1) minus housing costs and on amenities θ and β in each location. Conditional on the choice of neighborhood, their utility maximization problem reads:

$$\max_h \theta\beta \left(1 - ph\right)^{1-\alpha} h^{\alpha} \tag{5}$$

From the consumers' optimization problem one can derive the aggregate demand for housing space, which depends positively on population in a given location, denoted as N, and negatively on housing price p.

Finally, we impose spatial equilibrium, so consumers are indifferent across neighborhoods and have no incentives to deviate. In each neighborhood, the housing market clears with housing supply equal to housing demand. The indifference condition implies that the equilibrium price of housing in each neighborhood is ultimately pinned down by the amenities:

$$logp = \frac{1}{\alpha}(log\theta + log\beta) + const. \tag{6}$$

Figure 1a depicts two housing market equilibria, in the kampung and formal neighborhoods. Using the optimality condition for developers, that for consumers, and the spatial equilibrium condition, we characterize the equilibrium prices (horizontal axis) and building heights (vertical axis) in a given neighborhood, as a function of amenities(θ and β) and the fixed cost (c). The housing supply curve (h^S) has two segments: it is flat in informal neighborhoods and upward sloping above $\bar{h}$). Given housing supply, consumers' neighborhood choices and housing demand h^D pin down equilibria in the kampung and formal neighborhoods:

Equilibrium in kampung's. When amenities in a neighborhood are low relative to formalization costs, the housing market equilibrium corresponds to the intersection of h_K^D and h^S. The neighborhood is a kampung, characterized by low heights - constrained at $\bar{h}$ - and low prices:

$$\log h_K^S = \log \bar{h}. \tag{7}$$

$$\log p_K = \frac{1}{\alpha}\log\theta + const. \tag{8}$$

Because of low amenities, housing prices are too low to overcome the costs of formalization, and these neighborhoods stay informal. This equilibrium conceptualizes what most of Jakarta was like before KIP.

Equilibrium in formal neighborhoods. When amenities in a neighborhood are large relative to formalization costs, the equilibrium corresponds to the intersection of h_F^D and h^S. The neighborhood is formal, with taller buildings and higher prices, reflecting not only differences in θ but also endogenous amenities from formalization β:

$$\log h_F = \frac{1}{\alpha(\delta-1)}(\log\theta + \log\beta) + const. > \log\bar{h}. \tag{9}$$

$$\log p_F = \frac{1}{\alpha}(\log\theta + \log\beta) + const. \tag{10}$$

Through the lens of this model, a city starts formalizing once housing demand pushes prices high enough to overcome formalization costs, with the high-θ neighborhoods formalizing first. As a neighborhood formalizes, its price and building height jump discontinuously to new levels that reflect the newly acquired amenities from formalization β (this echoes the model in Henderson, Venables, and Regan (2016)).

3.2 Comparing KIP and non-KIP neighborhoods

Let us now consider two ex ante identical neighborhoods and examine the impacts of KIP being introduced in one of them. We will use our framework to characterize the static equilibrium shortly after the introduction of KIP (which we label "short run") and today (which we refer to as the "long run"). The short-run equilibrium is one in which amenities and demand are low everywhere to start with, so that both neighborhoods begin in the informal equilibrium, and KIP increases amenities in one of the neighborhoods. The long-run equilibrium is one in which the original KIP-provided amenities have converged in the two neighborhoods and the city as a whole has experienced a boost in housing demand, that captures the steep increase in urbanization rates that Jakarta has experienced in recent years.

1. In the short run, as a result of the improved public amenities (θ), the KIP neighborhood will have higher prices.

KIP improves amenities θ in the treated neighborhood, which induces an outward shift in the demand curve. Improved amenities will attract population into the neighborhood; this, in turn, will bid land prices up. Thus, in the short-run, KIP neighborhoods will have higher prices than non-KIP ones. While we do not have systematic historical price data to test this prediction, the 1995 KIP evaluation report provides suggestive evidence that surveyed KIP kampung's indeed had between 20 and 115% higher prices than non-KIP kampung's as of the early 1990s (World Bank, 1995).[13]

2. In the long run, if the initial KIP investments have depreciated and formalization costs are greater in KIP, the KIP neighborhood may have lower prices and lower heights.

If KIP not only increases θ but also increases formalization costs, through an increase in c KIP will also induce an increase in the price threshold $\bar{P}$ past which developers choose to formalize.

[13] Beyond the forces at play in our model, land values may have increased in KIP areas in the short run because of crowd-in of private investments by residents. Increased perception of tenure security or complementarities between public and private investments may have induced them to upgrade their dwellings. For instance, World Bank (1995) indicates that some household upgraded their floors once they realized that the neighborhood was less prone to flooding after KIP.

This corresponds to an outward extension of the flat portion of the supply curve, as depicted in Figure 1b. In the long run, the housing demand curve has also shifted outward, reflecting in-migration from outside Jakarta. In the example depicted in Figure 1b, the long-run equilibrium in the KIP neighborhood falls in this flat region and the neighborhood remains informal with heights at level $\bar{h}$.

Figure 1c illustrates higher prices and building heights in (formal) non-KIP neighborhoods relative to KIP neighborhoods. Assume that amenities θ have equalized between the KIP and non-KIP neighborhood: over time, KIP-provided amenities depreciated in the KIP neighborhood while they converged in the non-KIP neighborhood.[14] As the non-KIP neighborhood catches up in terms of amenities and faces the same Jakarta-wide boost in housing demand, it experiences the same outward shift in the demand curve that KIP neighborhoods faced (dashed line in Figure1c). But because formalization costs in the non-KIP neighborhood have not increased, this new demand curve may intersect the supply curve in its upward sloping portion, in which the equilibrium price is high enough to overcome formalization costs. In this case, the non-KIP neighborhood will formalize. Through the term β, this further increases demand for housing space in that neighborhood, resulting in a further outward shift. In this example, the new equilibrium is at the intersection of the two solid lines and corresponds to a higher price and higher building height than in the KIP neighborhood.

Higher prices and heights in the non-KIP neighborhood thus arise in the long run because demand is now high enough to justify formalization in those neighborhoods, but not in KIP areas that have higher formalization costs.

Formalization costs may be higher in KIP areas for several reasons. First, land assembly costs may be higher because KIP residents are more reluctant to give up their land for redevelopment. This could be related to entitlement effects and perceptions of more entrenched occupancy rights, resulting from being beneficiaries of a government program (Fox, 2014). For example, World Bank (1995) documents stronger perceptions of tenure security in KIP neighborhoods. While KIP beneficiaries are more likely to report that they have a right of ownership to their land relative to respondents in non-KIP places, the report also notes that the statistics are "known to be overestimated. Many people are unaware of their status regarding land and said they owned it though admitted they had no land certificate or document to prove it" (Table 13, p111). Furthermore, Garr (1989) provides anecdotal evidence of entitlement effects arising in one of the KIP-improved

[14]We provide empirical evidence in support of this in Section 5.

kampung's in the city of Surabaya, observing that the resident "was [not] apprehensive of...the government efforts to oust him from this location for either legal or environmental reasons..." due to a "KIP-induced...secure perception of a future unclouded by the potential of government enforcement efforts". While in the short run improved tenure security may have encouraged residents to make investments in the neighborhood, which would contribute to higher prices in KIP areas, this could also make developers more reluctant to attempt land assembly in KIP neighborhoods. In the long run, when prices would justify the redevelopment of kampung's, it is plausible that they would not target these neighborhoods but rather non-KIP ones. Describing the early kampung formalization efforts of the 1990s, World Bank (1995) reports that developers were indeed considering KIP status when selecting kampung's for potential redevelopment. Beyond entitlement effects, KIP residents may also have a higher reservation selling price because of higher incomes or higher mobility costs. Finally, physical redevelopment costs may be higher because KIP structures need to be dismantled.

This model makes a number of simplifying assumptions. First, as in the standard version of the Rosen-Roback model, consumers are perfectly mobile and homogeneous. In a model with mobility costs or heterogeneous preferences for certain locations or amenities, we would no longer have a perfect equalization of utility across neighborhoods and spatial equilibrium would hold only for the marginal consumer. However, the qualitative impacts of amenities on prices and heights would be the same. If we allowed for heterogeneous consumers differing in income and with non-homothetic preferences over housing, this will tend to further increase prices and housing space per capita in the non-KIP neighborhood once the latter formalizes, as it would attract higher-income households.

Second, the model does not allow for congestion or agglomeration. A richer model may introduce congestion in KIP amenities by making the utility derived from those amenities decreasing in population. This effect will tend to attenuate the response of population and prices to changes in amenities, but, again, will not change the sign of the impacts.

Our model also rules out spatial externalities to surrounding neighborhoods. In the presence of spillovers, the difference in price between the informal KIP neighborhood and the formalized non-KIP one will be attenuated. In the empirics, we leverage our granular data to explore the extent of spatial externalities.

We test the model's predictions on prices and heights in Section 6. The predictions on rates of formalization, formalization costs, amenities θ and β are tested in Section 7.

4 Data

This Section discusses our primary data sources: KIP policy maps, historical settlements maps, the assessed land values database, our photographic survey, and land use maps. In the Appendix, we discuss the sources of other data, including distance and topography controls, administrative boundaries, market transactions, and current amenities.

4.1 Policy maps

Our source for KIP coverage is a 2011 publication by the Jakarta Department of Housing (DPGP, 2011), consisting of more than 200 physical maps of Jakarta, with a detailed indication of KIP boundaries as well as assets provided as part of the program (such as roads or sanitation facilities). The goal of this publication was to make a detailed inventory of KIP investments 30 years from the original program. In addition to KIP policy boundaries, these maps also detail the individual assets provided as part of KIP: infrastructure (paved road segments), sanitation facilities (garbage collection bins, public taps, public toilets, deep water wells, drainage canals) and community buildings (markets, health centers, and schools). Extensive ground surveying was performed by the Jakarta Department of Housing mapping team to ensure accuracy. We were able to access the raw Autocad files that form the basis of these maps, which allows us to achieve a considerable level of precision in georeferencing and tracing the items of our interest - primarily, boundaries of KIP areas - taking full advantage of the 1:5000 resolution.

Figure 2 displays KIP treated areas as unshaded polygons. The large scale and systematic coverage of KIP across Jakarta is apparent from this map. This is indeed one of the strengths of our empirical setting: if KIP were a small-scale program, there may be greater external validity concerns, as a scaling up of the program may not produce the same results we detect. Moreover, the comprehensive coverage of KIP across the city mitigates concerns that our estimates are merely driven by a displacement of economic activity from KIP areas.

As part of our empirical strategy, we also perform a boundary discontinuity exercise around KIP boundaries. For this purpose, we manually select a subset of 163 "clean" boundary segments that provide an uncontaminated discontinuity sample, following Turner, Haughwout, and Van Der Klaauw (2014). Details of this selection procedure are provided in the Appendix. Figure A1 displays the boundaries used in our discontinuity analyses.

4.2 Historical settlements

We identify areas that were kampung's before the implementation of the program through two maps, that we georeferenced and digitized. The first is a 1959 U.S. Army map of Jakarta (U.S. Army Map Service, 1959) with a scale 1:50,000. Our second source is a 1937 map of the then-Batavia, with a scale 1:23,500, issued by the Dutch Department of Urban Development and Housing. In both maps, kampung's are clearly demarcated as distinct from "built-up" areas originally settled by the Dutch. We consider historical kampung's areas that are marked as "kampung" in either the 1959 or the 1937 map. These areas correspond to the shaded region in Figure 2.

4.3 Assessed land values

We observe assessed land values (*nilai tanah*) in Jakarta from a digital map created by the National Land Agency for property tax purposes. In developing countries, it is notoriousy challenging to obtain reliable data of property values with comprehensive coverage. In the absence of directly observed market transaction data, researchers often turn to tax record data. However, the concern is that transaction values reported to the tax office do not reflect true transaction prices because of incentives to evade sales transactions taxes (10% of the transaction price in Indonesia).

In 2015, in an effort to bring more transparency to land markets and to improve property tax collection, the Jakarta tax office transitioned to a computer-assisted mass appraisal assessment approach (similar to valuation models used in developed countries). The newly developed valuation model utilizes data from brokers, online websites, administrative and notary offices, including market transaction prices. Then further adjustments are made, based on property and neighborhood attributes, data quality, as well as the timing of the transaction. The land office also implements field visits to verify property and location attributes. The estimated property value is then decomposed into a building component (estimated using replacement costs) and a land component. The property tax due for each building is then calculated as the assessed land value multiplied by the area of the building. The result of this assessment process is a database with assessed land values in Rupiah per square meter at the sub-block level, where a sub-block is the smallest unit in zoning and city planning maps. The dataset covers 20,000 sub-blocks, shown in Figure 3. Each point in the map is a sub-block, with the solid gray boundaries representing village boundaries and the lighter grey boundaries corresponding to hamlet boundaries. A few of the strengths of our database are apparent from this map, such as its comprehensive coverage of all of Jakarta and its high resolution.

To gauge the extent to which the assessed values reflect market prices, we also scraped and manually geo-referenced a sub-sample of transactions prices from Indonesia's largest property website (more details are in the Appendix). Figure 4 presents a scatterplot indicating that assessed values are positively correlated with market transactions. The correlation coefficient is 0.56. Each point represents values averaged at the hamlet level.

4.4 Photo survey, heights, and the informality index

We draw information on current building heights and on whether a given location is a kampung by observing photographs of urban views. We collect over 28 thousand photos from a representative sample of 7104 locations through an innovative approach that combines publicly available geocoded imagery from Google Street View and photos collected on the field. Observing these photos, our enumerators then coded the maximum building height in each location the quality of the location based on the photos.

For our analysis of urban views as well as land use patters (discussed below), our units of observation are 75 meter by 75 meter pixels that we obtain by superimposing a grid over the territory of Jakarta. Each pixel has a size comparable to the land area required for an average high-rise development project in Jakarta, based on reports from the Jakarta City Planning Agency.

The first step in our data collection is the definition of the sampling frame. The ability to extract information from a representative sample is one of the advantages of collecting outcomes such as building height through our approach, as opposed to relying on existing sources - such as administrative permit records - that suffer from potential under-reporting of buildings (for tax evasion purposes). We start from our grid of 75 meter pixels and sample from the two subsamples that we consider for our baseline specifications (detailed in Section 5 below): the historical kampung's sample, comprising areas that were slums before the implementation of KIP, as per our 1959 map, and the boundary discontinuity sample, comprising areas within a 500 meter distance from the KIP boundaries used in the analyses. We select a random sample of 5000 pixels from the historical kampung's sample and 2500 pixels from the discontinuity sample. We stratify by terciles of distance from the National Monument within each subsample, to ensure we have a broad spatial distribution. The proportion of pixels in the first, second and third distance terciles are respectively 50%, 40% and 10%. Within each distance stratum, we draw half of the observations from KIP and half from non-KIP areas.[15] Because there is some overlap between the two subsamples, in the end we have 7104 observations in total, of which 5280 are in the historical kampung's sample and 1824 are in the 500 meter discontinuity sample. Figure A2 shows our sampled pixels.

Once we have identified our survey pixels, the next step is to obtain 4 photos for each pixel, corresponding to the north, south, east, and west angles from the pixel's center. We rely partly on Google Street View, feeding in the coordinates of the pixels' centroids, and partly on a photo field survey. 5401 of our pixels are covered by Google Street View, with imagery collected mostly in 2015-2017 (about 10% were collected in 2013). Importantly, the coverage of Google Street View is selective: Google cars are typically too large to go through the narrow streets of kampung's and cannot access gated commercial developments. For the 1703 pixels not covered by Google StreetView, we obtain photos from enumerators sent to the field.[16] The majority of photos from the field indeed correspond to areas with narrow streets or private developments. In total, we have approximately 28,000 photos.

The third step is the classification of photos. For each pixel, we derive two indicators. The first is the height, in floors, of the tallest building in the view, provided it lies within 37.5 meters of the pixel centroid. Pixels with no buildings, corresponding to large roads, parks, or empty lots, were assigned a height of 0 and a "no building" dummy. The average height in our sample is 4 floors, with a median of 2 floors. The distribution of heights is displayed in Figure 5 and shows significant bunching in correspondence of 2 floors, consistent with the predictions of the model.

The second indicator we derive from the photos is an informality index ranging from 0 to 4, where 0 corresponds to "very formal" and 4 to "very informal". A well-known gap in the literature on informal settlements is the inherent difficulty in defining and identifying slums. For the purposes of defining our indicator, the assessment is based on observable characteristics of the neighborhood and of the buildings visible in the photos.[17] These characteristics include, at the neighborhood level, the width and paving of roads, the presence of trash, the presence of

[15] The proportion of KIP and non-KIP pixels in the original samples is comparable - about 45% KIP and 55% non-KIP.

[16] Enumerators were provided with latitude and longitude coordinates of the locations they had to survey. They were asked to collect GPS coordinates of the photos they took, so that the accuracy of the location could be verified. In a few cases the enumerators could not reach the exact location due to buildings, walls, or roads blocking the access. In these cases, we used the closest available photos from Google Street View, which always falls within 150 meters of the pixel centroid. Our results are robust to dropping those locations.

[17] Approaches to detect slum areas from aerial, as opposed to ground imagery are reviewed in Kuffer, Pfeffer, and Sliuzas (2016). Recently, street-view imagery has been employed to detect urban change in the U.S. context by Naik et al. (2017).

messy electrical wiring, the density of structures and their regularity (e.g., whether building heights are even). At the building level, we consider the irregularity of structures (e.g., the presence of setbacks), the durability of materials, and the size of windows.

Examples of photos that we classified as belonging to each of the five categories can be found in Figure 6. We calibrate the index so as to capture the varying degrees of informality in a city that is on the brink of becoming formalized. The index takes value 0 for areas that are completely formalized and comparable to a developed country city ("very formal"); 1 for neighborhoods that appear formal but retain some of the traditional features of kampung's, such as the narrow roads ("formal"); 2 for kampung's that are overall in good conditions (e.g. they have paved roads and concrete buildings; "slightly informal"); 3 for kampung's that are in worse conditions ("informal") and 4 for areas that are "very informal".

We trained two research assistants from Jakarta who independently performed the assessment. We then consider an average of the index as coded by the two enumerators; our results are robust to different aggregation approaches. The median value of the resulting index in our full photo sample is 2.

4.5 Land use patterns

Our source for land use patters is a series of detailed digital cadastral maps created by the Jakarta Department of Housing in 2011.[18] Such maps include the outlines of land parcels, footprints of individual buildings, as well as roads, and waterways. Information is also provided on whether each structure is commercial or residential. It is worth emphasizing that these maps also cover informal areas, providing an exceptionally detailed mapping of structures, as it can be appreciated in Figure 7.

From these maps, we are able to derive two sets of outcomes. We first construct two measures of formal commercial density at the 75 by 75 meter pixel level by computing the land share of each pixel corresponding to retail and office buildings respectively.

Next, we consider the spatial configuration of land parcels, specifically land fragmentation. Our interest in fragmentation lies in the fact that it plays an important role in the land assembly process. For concreteness, consider a developer starting a high-rise development project that requires an amount of contiguous land roughly as large as one of our 75 meter by 75 meter pixels. In order

[18]These maps are made available through the website of the Jakarta Regional Disaster Management Agency.

to do this, a number of adjacent land parcels will have to be purchased from current owners. In the presence of many small parcels, private market frictions to the land assembly process are more pronounced as there are more claimants and a greater likelihood of strategic delay and holdout (Brooks and Lutz, 2016). Informal areas are typically characterized by a more fragmented pattern, with a large number of small and irregular parcels.

We consider the number of parcels in each pixel as a simple metric of land fragmentation that proxies for the costs of land assembly. For robustness, in the Appendix we also consider alternative metrics, such as average parcel size within a pixel, which is negatively correlated with fragmentation, and the K land fragmentation index (Januszewski, 1968),[19] ranging from 0 (in the limit case of an infinite number of parcels) to 1 (for a single parcel).

Table 1 includes summary statistics for the full sample. Panel A reports statistics for the assessed values database of close to 20,000 sub-blocks for which we have assessed values data. The average value is 12.45 million Rupiahs per square meter (around 93 USD per square foot). Panel B reports statistics for land use patterns with close to 90,000 pixels. Panel C reports summary statistics from over 7,000 pixels covered by our photo survey sample.

5 Empirical framework

We estimate variations of the following estimating equation:

$$Y_{ij} = \alpha + \beta 1(KIP_{ij}) + \gamma X_{ij} + \delta_j + \varepsilon_{ij} \tag{11}$$

where Y_{ij} is the outcome for unit i in location j. The key regressor, $1(KIP)$, is a dummy equal to 1 if i is located in a KIP treated area, according to our policy maps. Additionally, X_{ij} represents a vector of pre-determined controls, δ_j is a village (or hamlet) fixed effect, [20] and ε_{ij} is an idiosyncratic error

[19]For N parcels each of which has area $a_i, i = 1..N$, the K index is defined as $K = \frac{\sqrt{\sum_{i=1}^{n} a_i}}{\sum_{i=1}^{n} \sqrt{a_i}}$.

[20]The primary administrative units in Indonesia are districts (*kabupaten*), sub-districts (*kecamatan*), villages (*kelurahan*), hamlets (*Rukun Warga*, RW), and block groups (*Rukun Tetangga*, RT). Our empirical analysis will focus mostly on villages or hamlets. As a reference, villages have an average area of 2.5 square kilometers, and are comparable to U.S. census tracts. Hamlets have an average area of 0.24 square kilometers and are comparable to U.S. census block groups.

term. Our outcomes include assessed land values (where the unit of analysis is a sub-block in the assessed values database) and land use patterns (where the unit of analysis is a pixel).

The parameter of interest is β, which captures the long-term impacts of KIP, comparing treated areas relative to non-treated areas. One major threat is that β is biased downwards because KIP targeted low-quality informal settlements. To circumvent the concern that β is identified from comparing slums to formal settlements, our first strategy is to restrict our sample to historical kampung's that existed before KIP, and compare those that were part of KIP with those that were never treated (see Section 4.2). We label this the historical kampung's sample.

Our second strategy is to implement a boundary discontinuity design comparing observations located on either side of KIP treated areas boundaries. The identification assumption is that unobserved determinants do not change differentially around the boundaries, conditional on the controls. The strength of this strategy is that we can restrict the sample to narrow distance bands (within 200 meters), which holds constant many potential unobserved determinants of land values. We label this the discontinuity sample. We include boundary fixed effects and quadratic distance controls to KIP boundaries; the results are robust to other types of distance controls (e.g. logs).

For the purposes of defining the discontinuity sample, we manually inspected all policy boundaries shown in Figure 2 and only considered boundary segments such that the control group is not contaminated by treated areas nearby, as discussed in Section 4.1. Figure A1 in the Appendix illustrates the boundaries we include in our boundary discontinuity analyses.

KIP neighborhood boundaries are pre-determined because these largely depend on hamlet boundaries defined before KIP; in fact, many of these neighborhood units were introduced by the Japanese during World War II.[21] Nevertheless, a concern may arise that KIP boundaries coincide with current administrative boundaries, so that effects estimated from a boundary discontinuity design are driven by differences across administrative units. Importantly, we verified that the boundaries in our sample do not merely follow administrative boundaries and our conclusions are similar if we control for village fixed effects, as we show below.

Additionally, we perform a number of tests to address several concerns with the boundary discontinuity design. First, the results may be too local and not representative of the rest of Jakarta. In particular, the differences across the boundaries may reflect a mere reshuffling of development activity across the boundaries. We address this by also presenting a full sample analysis that covers

[21] During the occupation in World War II, the Japanese army introduced the system of *Rukun Warga* (RW hamlets, smaller than villages) and *Rukun Tetangga* (RT blocks, smaller than hamlets) for security purposes.

all of Jakarta to assess how representative the boundary discontinuity estimates are. Second, there could be spillover and contamination effects across the boundaries. To investigate this, we present estimates of KIP effects at different distance bands from the boundary, up to 1000 meters. To the extent that access to KIP-related amenities declines with distance, the localized spillover effects should also exhibit spatial decay patterns (Turner, Haughwout, and Van Der Klaauw, 2014). We do not find evidence of any decay, suggesting spillovers are limited.

In all of our specifications, we control for location-specific natural advantage through a series of variables capturing access and geography. Specifically, we control for distance, in logs, from the National Monument in Merdeka Square (which we interpret as the historical city center), from the headquarters of the Indonesia Stock Exchange, and from historical main roads. An important dimension of natural advantage in Jakarta is flood proneness, which played a significant role in the city's settlement patterns. We account for it in multiple ways: we control for distance, in logs, from the coast, from waterways and from any permanent or semi-permanent water body; moreover, we include a binary indicator for whether a hamlet is flood-prone, according to flood information reported in Openstreetmap. Finally, we account for local topography by including slope and elevation controls. All of our data sources are documented in Appendix B. Our results are unaffected if we drop the controls that are not pre-determined (distance to the stock exchange building, flood proneness in a hamlet, distance to the nearest waterways).

As a preliminary step, Table 2 presents regressions to assess how comparable KIP and non-KIP areas are in the sub-block level dataset (columns 1 and 2) and in the pixel level dataset (columns 3 and 4). Each pair of cells report the coefficient and p-value, in square brackets, from a regression of location attributes X_{ij} on the treatment dummy and geographic fixed effects, with standard errors clustered at the village level. We report estimates from the historical kampung's sample with village fixed effects (odd columns) and for the 200 meter boundary discontinuity sample with KIP boundary fixed effects (even columns). The results are similar if we include hamlet fixed effects, or if we cluster standard errors at the hamlet level.

Overall, the distance and topography attributes are relatively comparable, after conditioning on the fixed effects. A few coefficients are significant statistically but small in magnitude. Pixels in KIP areas are less steep (columns 3 and 4), but the magnitudes of the coefficients (0.24 and 0.21 degrees) are negligible relative to the mean (2 degrees). KIP pixels appear closer to the stock exchange (column 4), but this difference is negligible in magnitude (0.5 %) and would, if anything, bias against finding a negative effect of KIP on land values.

Next, KIP areas also appear to be further from waterways (columns 2 to 4). The coefficients range from 14% to 27%. Locations that are farther from waterways tend to have higher land values, biasing against finding lower assessed land values in KIP areas. These locations could have higher land values because they tend to be less flood-prone. Finally, our binary indicator for being in a flood-prone hamlet is significant for the sub-block level boundary discontinuity sample (column 2). Within 200 meter distance bands, observations on the KIP side of the boundary are 10 percentage points more likely to be in a flood-prone hamlet. While this might appear a large effect, it likely is driven by the construction of this variable which may generate spurious differences, accentuated by narrow distance bands.[22]

Importantly, our main results are nearly unchanged if we drop these four variables, indicating that these differences are not driving the differences in land values and building heights. Taken together, these results are reassuring that treated and control areas are reasonably comparable, once we control for fine-grained geographic fixed effects. In Appendix Table A1 we show the same exercise for the full sample; results are similar.

6 Main Results

Our model suggests that, if KIP increases formalization costs in treated areas, in the long run KIP neighborhoods should be characterized by lower land values and lower building heights than control areas. In this Section, we take these predictions to the data. We consider the long-run impact of KIP on land values and heights in Sections 6.1 and 6.2 respectively. In Section 7 we examine more closely the delayed formalization mechanism proposed in the model as well as heterogeneous effects.

6.1 Effect of KIP on land values

Table 3 presents our main results for the effect of KIP on assessed land values (columns 1 to 3). The dependent variable is log of price per square meter. The unit of analysis is the sub-block. We

[22] The main issue is that this variable is not continuous across space: given that KIP boundaries are sometimes near hamlet boundaries, a small change in underlying flood proneness as the KIP boundary is crossed may translate into a discontinuous jump from 0 to 1 in our binary indicator. In fact, this same coefficient is only 2 percentage points and is insignificant in the historical kampung's sample, where we consider variation within villages, and is negative and insignificant in the pixel level sample.

report regressions from the historical kampung's sample (column 1), the 200 meter discontinuity sample (column 2), and the full sample (column 3). Standard errors are clustered at the village level, by KIP boundary, or at the hamlet level, respectively.

Column 1 presents our baseline estimate showing KIP areas have 12% lower land values compared to historical kampung's within the same village that were not part of KIP. Restricting the sample to historical kampung's addresses the concern that KIP was a program implemented only in slums and that our finding of lower land values could be driven by pre-treatment differences between kampung's and formal areas. This specification includes 196 village fixed effects with 88 villages that have within-village variation in KIP status.

In the boundary discontinuity analysis in column 2, we estimate a similar difference (13%) comparing observations within 200 meters of KIP boundaries. The identifying assumption for this strategy is that unobserved determinants of land values do not change discontinuously at the boundaries, conditional on observables. We additionally control for the distance to the KIP boundary and its square, as well as 124 boundary fixed effects.

In Table A2, we show that the estimates are similar across different buffer distances ranging from 150 meters (the optimal bandwidth as per Calonico, Cattaneo, and Titiunik (2014)) to 500 meters. Once again, we find similar effect sizes with KIP areas having 14 to 16 percent lower land values. One concern that arises with the boundary discontinuity sample is contamination or spillovers between treatment and control areas. The stability of the estimates across the buffer distances suggests limited evidence of spatial externalities due to KIP. Below, we present further analyses to show limited evidence of spatial externalities.

Another potential concern with the boundary discontinuity exercise is that we may be picking up differences across administrative units, to the extent that KIP boundaries coincide with administrative boundaries. In Table A3 we mitigate this concern by showing that the effects of KIP at the boundary are similar when considering variation within the same administrative unit, as we control for village fixed effects with a 500 meter distance band.

To address concerns on the representativeness of the historical kampung and the boundary discontinuity sample, in column 3 we also present a specification that utilizes the full sample, finding a similar effect (-11%). This sample includes 2,060 hamlet (*Rukun Warga, RW*) fixed effects, with 304 hamlets that have within-hamlet variation in treatment status.

Taken together, the results in Table 3 indicate that, in the long run, KIP areas have 11 to 13 percent lower assessed land values. Reassuringly, the magnitudes are similar across the three specifications. The 12 percent effect in column 1 is large, translating into an aggregate effect of $11 billion US dollars.[23]

In the remaining empirical analyses, the historical kampung and the 200 meter boundary discontinuity specifications of columns 1 and 2 will be our baseline ones. In instances in which a larger sample is required for heterogeneity analyses, we will present specifications using the full sample.

6.2 Effect of KIP on heights

The second prediction of the model relates to building heights. In the long-run, KIP neighborhoods will have lower heights, bunching around the $\bar{h}$ threshold corresponding to the maximum height attainable with the informal building technology. We test this by drawing on building heights data from our photo survey. Results are reported in columns 4 through 6 of Table 3. The unit of analysis is a pixel and the dependent variable is the number of floors of the tallest building in view in a pixel. Column 4 reports the historical kampung's specification, column 5 reports effects for the 200 meter boundary analysis, and column 6 includes the full photo survey sample. We include the same set of controls as in our land values analysis, but also add sampling strata fixed effects as well as a dummy for pixels with public spaces and roads.

Once again, we find relatively stable effects across the specifications, with the average building height being 1.3 to 1.6 floors lower in KIP areas compared to observably identical non-KIP areas. This is a large effect, relative to a mean of 5 floors and a standard deviation of 9 for the control group in the historical sample.

Strikingly, Figure 5 plots the distributions of building heights for the treated (shaded bars) and control (unshaded bars) groups, restricting the sample to historical kampung's only. At the left tail of the distribution, we observe bunching around 2 floors for KIP areas, consistent with the model. At the right tail, non-KIP areas have twice as many buildings with five to ten floors (73 buildings in KIP versus 138 buildings in non-KIP) and similarly, for buildings with more than ten floors (116 buildings in KIP versus 262 buildings in non-KIP).

Our calculations suggest that the effect sizes for building heights corroborate the magnitude

[23] The average assessed value for historical kampung's is 12 million Rupiahs per square meter (around US$89 per square foot), translating into an effect size of 1.4 million Rupiahs per square meter, or US$11 per square foot (at an exchange rate of 13,371 Rupiahs to US dollars). We obtain the aggregate effect by multiplying by the total area under KIP, 100 million square meters (10,000 hectares).

of the aggregate land values impact ($11 bilion). We first estimate a hedonic regression using land values in non-KIP historical kampung's only. The key coefficients are a dummy for buildings with four to ten floors, and a dummy for buildings with more than ten floors (the omitted group is buildings with three or fewer floors). We include the standard controls for the heights specification, including village fixed effects. The estimated price premiums are non-linear, with a price increase of 22 log points (25%) for buildings with four to ten floors, and 42 log points (52%) for buildings with more than ten floors. We apply these price coefficients to translate the height effects into aggregate value impacts, concluding that the loss in tall buildings explains 88% of the aggregate impact of land values.[24] The results are similar if we only used formal areas (defined using our informality index) to estimate the price premium. This is reassuring as we obtained this using only land values in non-KIP or formal areas, addressing the concern that measurement error or mis-reporting biases in the assessed values data may be differential in KIP versus non-KIP areas.

Table A4 in the Appendix repeats the analysis for different height outcomes. We focus on the historical kampung's sample only but the results are similar for the boundary sample. In column 1 we consider a binary indicator for whether the building has strictly more than 3 floors. Consistent with the bunching pattern in Figure 5, we find that KIP areas are 12 percentage points less likely to have buildings with more than 3 floors, compared to a mean of 0.24 for the control group in the historical kampung's sample. In column 2 the dependent variable is the log of height; we obtain an effect of -19 log points (-21 percent, implying an effect size of 1.1 floors). In columns 5 to 8 we address the concern that some of the survey photos may be outside of the pixel due to GPS imprecision in the field. Results are similar if we only consider photos taken within the 75 by 75 meter pixel (columns 5 and 6) and if we replace field photos taken outside of the pixel with photos from Google Street View that are closer to the intended location (columns 7 and 8).

7 Mechanisms

In our model, differences in land values and heights between KIP and non-KIP neighborhoods are driven by delayed formalization in KIP neighborhoods, which in turn is associated with greater

[24]KIP areas have 146 fewer buildings with more than ten floors. Combined with the 52% price premium in price per square meter (relative to a price of 13.4 millions of Rupiahs per square meter), and the total footprint of each building (which we assume is 75 by 75 meters), the total heights effect (expressed in land values) is US$1.3 billion. The effect for buildings from four to ten floors is US$900 million. Therefore, the aggregate impact from the heights analysis is US$2.1 billion. The aggregate land value impact for the estimation sample is $2.4 billion.

formalization costs in KIP areas. In this Section, we empirically examine this channel by providing evidence on the degree of formalization in KIP and non-KIP areas, and by examining land fragmentation as a proxy for land assembly costs. We then consider other channels, including differences in current and historical amenities (β and θ in the model) and spatial spillovers associated with KIP investments. Finally, we present heterogeneous effects indicating larger treatment effects in central areas relative to peripheral areas, consistent with the model, which predicts the formalization channel is stronger when the potential gains from formalization are larger.

7.1 Informality and land fragmentation

We begin by showing that KIP areas are more likely to be informal kampung's, as measured by our photo survey. Figure 8 plots the distribution of our informality index for treated and control pixels, within the historical kampung's sample. As discussed in Section 4.4, 0 indicates very formal areas and 4 indicates very informal neighborhoods. A clear pattern emerges analogous to the histogram for building heights: control areas (unshaded bars) are more likely to be formal whilst KIP areas are more likely to have a higher degree of informality, as per our index.

In Table 4, we present evidence that KIP areas are 15 percentage points more likely to be informal kampung's today, relative to non-KIP areas that were historical kampung's (column 1). The dependent variable is 1 if the informality index (averaged across both research assistants) is strictly above one. The mean for non-KIP historical kampung's is 40%. The effect is similar in columns 2 and 3: 20 percentage point (p.p.) for the boundary analysis and 14 p.p. for the full sample. Table A5 in the Appendix shows that our conclusions remain the same using different approaches, including defining the kampung dummy using other threshold values for the informality index besides one (columns 1 through 4), including the continuous informality index as the dependent variable (columns 5 and 6), pooling the index values for the two research assistants (and adding a fixed effect for one of them), as opposed to averaging between the two scores (columns 7 and 8).

We next turn to formalization costs, which are at the heart of our proposed mechanism. We focus on land assembly costs and consider the number of land parcels within a 75 meter by 75 meter pixel as a proxy. This simple metric of land fragmentation captures the idea that, in order to assemble a large enough plot for a formal building, a developer needs to first purchase adjacent land parcels from current claimants. This process is costlier, the more parcels need to be assembled, as a larger number of claimants translates into a greater likelihood of hold up issues and strategic delay (Brooks and Lutz, 2016). In Table 4, columns 4 through 6, we show that KIP areas have a greater degree of land fragmentation. KIP areas have on average 9 to 13 more parcels per pixel, relative to a mean of 19 parcels. Besides our standard set of distance and geography controls, we also control for the total log length of roads in the pixel, as the presence of road intersections may mechanically increase observed fragmentation.

In order to appreciate to what extent the greater land fragmentation in KIP areas contributes to the lower land values and heights, we can relate the results above to the impact of fragmentation on land values and heights. In Table 5 we restrict the sample to historical kampung's within KIP. Columns 1 to 3 show that one additional parcel per pixel is associated with a 1% decline in land values in KIP areas, 0.06 fewer floors, and 1 p.p. increase in the likelihood of being a kampung today. Benchmarked to these figures, KIP's impact of raising the parcel count by 9 translates into a 9 p.p. effect on prices (75% of the overall 12% price effect), -0.54 floors (33% of the overall -1.64 effect on floors), and a 9 p.p. increase in the likelihood of being informal (60% of the overall 15 p.p. effect on the kampung indicator).

Table A6 in the Appendix shows the conclusions are similar if we use other proxies of land fragmentation, such as average parcel area and the K fragmentation index (discussed in Section 4.5). The average parcel area in KIP is smaller by approximately 560 square meters (about 20% of the area of a pixel) to 960 square meters. Impacts on the K index are best understood in standardized terms: relative to control pixels, KIP pixels are more fragmented by roughly 2/5ths to 2/3rds of a standard deviation of the K index. As a reference, the average value of the K index is 0.5 in the control sample.

7.2 Amenities

Next, we turn to amenities, which have their model counterpart in θ and β. Non-KIP neighborhoods are predicted to have better endogenous amenities from formalization (β in the model), which we test in Table 6. Furthermore, our model postulates that exogenous amenities θ have converged or depreciated between KIP and non-KIP areas. In the next tables we provide evidence that public amenities have converged (Table 6) and that the original KIP investments may have fully depreciated (Table 7). This rules out that differences in land values between KIP and non-KIP areas are driven by differences in public amenities.

Current amenities. Table 6 demonstrates that KIP areas today have less formal amenities (β) and similar access to public amenities (θ), relative to non-KIP areas. We restrict the sample to historical kampung's only, but the conclusions are similar using the full sample. Columns 1 and 2 focus on formal commercial developments, such as formal retail and office developments. The dependent variables measure the share of each pixel that is associated with retail activity or office developments, respectively, according to the administrative database on land use patterns (described in Section 4.5). Our estimates indicate KIP areas have 2 p.p. lower retail density and 4 p.p. lower office development density.

By contrast, we find small differences in accessibility to public amenities (columns 3 through 6). The information is drawn from OpenStreetMap, a crowd-sourced database which supplements our administrative source (see the Appendix for details). In particular, the outcomes are logs of the distance (from the middle of the pixel) to the nearest school, hospital, police station, and bus stop. These can be viewed as being part of the vector θ in the model. There are no effects for schools and police stations. The 7 percent effect for hospitals translates to 70 meters (small relative to a mean of 1000 meters), and the 31 percent for bus stops is equivalent to 258 meters (small relative to a mean of 835). This corroborates our discussion above that non-KIP areas have converged with KIP areas in terms of access to public goods.

Initial KIP investments. In addition to contemporaneous amenity and development levels, Table 7 shows heterogeneous effects of historical KIP investments. Specifically, we examine the three primary KIP policy components - roads, sanitation, and buildings (health centers and schools). We observe the location and type of KIP investments from the policy maps (discussed in Section sec:policyData). For each assessed land value observation, we quantify the amount of KIP investments located within a 500 meter buffer, for observations in KIP and non-KIP areas. This allows for the possibility that residents in non-KIP areas were also able to access KIP investments. To quantify road investments, we calculate the length of roads built by KIP within 500 meters of each observation. To quantify the prevalence of sanitation facilities and public buildings, we count the number of facilities within 500 meters. Additionally, we de-meaned these three measures of KIP investments so that the coefficient on the treatment indicator corresponds to the average treatment effect (evaluated at the average of these three investment measures). Column 1 reports the results for the historical sample and column 2 presents results for the full sample.

Interestingly, we do not find differential treatment effects by the type of investment on current land values. The lack of heterogeneity in effects by policy investments is consistent with the notion that differences in initial public investment levels may have equalized across KIP and non-KIP areas by now. As discussed in Section 2, most of the initial investments have plausibly depreciated over three to four decades.

7.3 Potential gains from formalizing

Next, we show that the effects of KIP are larger in areas where the potential gains from formalization are greater, by examining effects in the city center versus the periphery. To overcome the fixed cost of formalization, the model suggests that formalization occurs in neighborhoods where prices are above the threshold $\bar{P}$. If a neighborhood is in an equilibrium characterized by very low aggregate demand for housing (such as the periphery of Jakarta), formalization is unlikely and the neighborhood remains in the kampung equilibrium depicted in Figure 1a. In these areas, we expect to see smaller differences between KIP and non-KIP areas, because it is less likely that non-KIP historical kampung's have formalized.

In Table 8 we show that the effects on land values and heights are larger in the city center than the periphery. We define the center of the city as the National Monument, that is mentioned in many planning documents as a major landmark (that pre-dated KIP) and remains a hub for the city today. We extend our most saturated specification with the full sample (column 3 of Table 3), dividing Jakarta into terciles according to the distance to the National Monument (central, middle, periphery) and interacting these tercile indicators with the treatment indicator. The omitted group is the periphery. We report heterogeneous effects for land values (column 1) and building heights (column 2).

Column 1 demonstrates that the average land value in central KIP areas is 14% lower. The coefficient on the interaction term is large and statistically significant (-0.14) but the direct effect is insignificant. The interaction term is also large and significant for the middle tercile (-0.13), but the effect at the periphery is small (-0.01) and insignificant. Column 2 shows a similar pattern for building heights. Since the program was implemented several decades ago, KIP areas are concentrated in central areas (67%), with 27% in the middle, and 6% in the periphery. We can repeat our aggregation exercise to calculate the implied total opportunity cost of land use, accounting for the heterogeneity in KIP impacts and coverage across different parts of the city. Our calculations suggest the opportunity cost of land use is \$9.3 billion in the center, \$2 billion in the middle, and \$35 million in the periphery.

Overall, this pattern of heterogeneous effects is consistent with the interpretation of delayed formalization in KIP areas. Central areas, that have greater market potential, are the ones associated with a larger opportunity cost of land use and will tend to be formalized earlier, leading to a larger observed gap in land values.

7.4 Spatial externalities

Finally, we study potential spatial externalities by examining whether the effect on land values exhibits any spatial decay, as we move away from the KIP boundaries (Turner, Haughwout, and Van Der Klaauw, 2014). Localized externalities provide an economic motivation for slum upgrading policies. Residents of nearby non-slum areas are often concerned over negative spillovers stemming from unsanitary living conditions and depressed public and private investment in slums.

We estimate spatial decay patterns by exploring heterogeneous effects by distance bands. We repeat the same specifications as Table 3, but the omitted group is now the *treatment* group and the key regressors include binary indicators for areas outside of KIP and within 100 meter distance bands. We start with 0 to 100 meters, and include up to 10 distance bands (up to 900 to 1000 meters). We drop observations beyond 1000 meters.[25]

Figure 9 plots the effects by distance bands and the 95% confidence intervals, for the historical kampung's specification. Consistent with the stability of our coefficients in Table A2, we find no significant spatial decay patterns. The full set of regression results is reported in Table A7, in which all columns confirm the lack of spatial decay. In columns 1 and 2, the 95% confidence intervals are quite stable and overlap with each other. In column 3, with hamlet fixed effects, the coefficients beyond the 500 meters bands are not significant because there is little within-hamlet variation beyond 500 meters.[26] The conclusions are similar using the historical kampung's sample in columns 4 and 5.

8 Threats to identification

This Section discusses the main threats to identification. We first discuss program selection bias followed by potential confounding due to the persistence of slums. We also address concerns about biases in assessed land values.

[25]The average village has an area of 2.5 square kilometer, implying an equivalent area radius of 892 meters.
[26]The average hamlet has an area of 243,156 square meters, implying an equivalent area radius of 278 meters.

Program selection. Table 9 examines the extent to which program selection biases could explain our effects. Here, we augment our analysis by decomposing the main treatment effect found in Table 3 into three program indicators corresponding to the three KIP *Pelita* waves (five-year plans). We present results for land values (columns 1 and 2) and building heights (columns 3 and 4) using the historical kampung's sample and the full sample.

In theory, the differences across the three waves reflects differences in selection effects (the program selection rule prioritized the worst places first), but also differences in direct treatment effects (KIP investments are older in earlier waves). However, as we showed in Table 7, the direct effect from KIP investments are small by now, plausibly due to depreciation. Therefore, across the three waves, there should be less heterogeneity in treatment effects associated with the timing of the investments. By contrast, the selection rule would suggest that the first Pelita wave targeted the worst places. If the lower land values and lower building heights reflect selection bias only, then, we should find the largest effect in the first wave, which is not what we find.

Reassuringly, there is no monotonic pattern in effect sizes, with larger effects (in absolute magnitude) for the Pelita I wave and progressively smaller effects for the later waves. For historical kampung's in column 1, the effects on land values are -13% for the first wave, -9% for the second wave, and -14% for the third wave, with overlapping confidence intervals. Similarly, column 3 shows that the effects on building heights are -0.38 for the first wave, -1.75 for the second wave, and -1.04 for the third wave. The conclusions are similar using the full sample (columns 2 and 4).

Persistence of slums. To further probe the extent to which the lower land values at the boundary reflect the fact that KIP areas were slums, we present a falsification test using historical slum boundaries (Table 10). Specifically, column 1 restricts the sample to locations within 200 meters of historical slum boundaries. Here, we include 45 historical slum boundaries that are not contaminated, control for distance to the boundary (and its square), and cluster standard errors by boundary segments. In column 2 we consider observations within a 500 meter distance band. If slums have a persistent effect on land values, we should find a negative and significant effect when we compare areas that were historical kampung's against areas that were not. Instead, we find an insignificant effect for both the 500 meter and the 200 meter samples. This is also consistent with our baseline specification with the historical sample only, which effectively differences out any confounding due to the persistence of slums by comparing within historical kampung's only.

Bias from land values assessment. Next, we consider the possibility that the negative land values are driven by differences in the quality of the assessed land values in our dataset. First, to the extent that the measurement error is smooth across KIP boundaries, it is unlikely that the boundary discontinuity estimates are driven by measurement error.

Nevertheless, one concern is that KIP areas are more likely to be informal today, and property data for informal settlements are less likely to be reported. Table A8 in the Appendix investigates whether KIP areas are less likely to be represented in the assessed values dataset. The unit of analysis is a 75 by 75 meter pixel and the dependent variable is whether we observe an assessed value for the pixel. In contrast to concerns that KIP areas are less likely to be included the assessed values database, we actually find a positive coefficient on the treatment dummy. Column 1 includes the full sample with hamlet fixed effects and column 2 restricts the sample to historical kampung's only, with village fixed effects.

Furthermore, we show in Table A9 in the Appendix that the negative impact on land values survives if we drop areas that are more likely to be informal. This is a difficult exercise because there is no direct measure of informality in the full sample. As a way around this, we assume that areas with greater land fragmentation (as defined in Section 4.5) are more likely to be informal. We show that our results survive even when we drop pixels in the top quartile of our proxies of land fragmentation. This reduces the concern that our results are driven by measurement error that is correlated with informality and the treatment indicator.

9 Conclusion

As cities in developing countries prepare to accommodate an unprecedented wave of urban expansion, policy makers are debating different policy approaches to facilitate the massive urbanization flows. In this paper, we provide novel causal evidence of the long-term impacts of the world's largest scale slum upgrading program, the Kampung Improvement Program, in the city of Jakarta. Our setting is unique in that we have granular data on a city that has started formalizing, over 40 years from the implementation of the program.

Across empirical exercises and robustness checks, we consistently find that KIP areas have 12% lower land values. KIP areas also display less commercial activity and greater land fragmentation. These findings suggest that the program slowed down the formalization process, resulting in an unintended long-run distortionary cost. Aggregating over KIP treated areas, this opportunity cost of land use amounts to US$11 billion.

In future work, we will complement our analysis with bounding exercises aimed at informing policy, shedding light on the trade-off between the program's benefits and its long-run costs. In doing so, we acknowledge that the counterfactual for current residents of kampung's, should their neighborhood be formalized, is plausibly displacement to peripheral areas with poorer job access. We hope that our exercise will shed light on the reasons for the slow pace of the formalization process in Jakarta.

As governments of many countries today are considering slum upgrading, our findings can help inform their policy decisions in a number of ways. One general lesson is that slum upgrading entails long-run distortionary costs, and that such costs should be weighed against a program's benefits. As these distortionary costs will manifest themselves once a city starts formalizing, the balance between costs and benefits will partly depend on the timing of the implementation of the program and the stage or development that a city is in. For a city comparable to Jakarta in 1969, slum upgrading may be an attractive policy option, since these long-run costs will materialize far in the future. On the other hand, for a city that has already started formalizing, the opportunity cost of land use may already be too high for slum upgrading to have a positive net present value. Another policy implication is that slum upgrading may be more successful if complemented by other measures, aimed at facilitating formalization and the development of well-functioning land markets. In future work, it will be interesting to elaborate on these trade-offs further and develop a stylized model to present scenarios in which slum upgrading can be more or less attractive than other policy counterfactuals.

References

Barnhardt, Sharon, Erica Field, and Rohini Pande. 2017. "Moving to Opportunity or Isolation? Network Effects of Randomized Housing Lottery in Urban India." *American Economic Journal: Applied Economics* 9 (1):1–32.

Brooks, Leah and Bryon Lutz. 2016. "From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly." *American Economic Journal: Economic Policy* 8 (3):69–105.

Brueckner, Jan K. and Somik V. Lall. 2015. "Cities in Developing Countries: Fueled by Rural-Urban Migration, Lacking in Tenure Security, and Short of Affordable Housing." In *Handbook of Regional and Urban Economics Volume 5*, edited by J. Vernon Henderson Gilles Duranton and William C. Strange, chap. 21. Elsevier Science.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6):2295–2326.

Castells-Quintana, David. 2017. "Malthus living in a slum: Urban concentration, infrastructure and economic growth." *Journal of Urban Economics* 98:158–173.

Collins, William J. and Katharine L. Shester. 2013. "Slum Clearance and Urban Renewal in the United States." *American Economic Journal: Applied Economics* 5 (1):239 – 73.

Darrundono. 1997. "Kampung Improvement Programme, Jakarta, Indonesia." Aga Khan Award for Archtecture brief.

Darrundono. 2012. "Proyek Muhammad Husni Thamrin."

DPGP. 2011. "Updating aset MHT wilayah administrasi dan kepulauan seribu." Dinas Perumahan Dan Gedung Pemerintah Daerah (DPGP), Provinsi Daerah Khusus Ibukota Jakarta.

Duranton, Gilles, Ejaz Ghani, Arti Grover Goswami, and William Kerr. 2015. "The Misallocation of Land and Other Factors of Production in India." Policy Research Working Paper Series 7221, World Bank.

Feler, Leo and J. Vernon Henderson. 2011. "Exclusionary Policies in Urban Development: Urban-Servicing Migrant Households in Brazilian Cities." *Journal of Urban Economics* 69 (3):253–272.

Field, Erica. 2007. "Entitled to Work: Urban Property Rights and Labor Supply in Peru." *The Quarterly Journal of Economics* 122 (4):1561–1602.

Field, Erica and Michael Kremer. 2008. "Impact Evaluation for Slum Upgrading Interventions." .

Fox, Sean. 2014. "The Political Economy of Slums: Theory and Evidence from Sub-Saharan Africa." *World Development* 54:191–203.

Galiani, Sebastian, Paul Gertler, Ryan Cooper, Sebastian Martinez, Adam Ross, and Raimundo Undurraga. 2017. "Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums." *Journal of Urban Economics* 98 (1):187–213.

Galiani, Sebastian and Ernesto Schargrodsky. 2010. "Property rights for the poor; Effects of land titling." *Journal of Public Economics* 94 (9):700–29.

Garr, Daniel. 1989. "Indonesia's Kampung improvement program: Policy issues and local impacts for secondary cities." *Journal of Planning education and Research* 9 (1):79–83.

Glaeser, Edward. 2011. *Triumph of the City: How Our Greatest Invention Makes us Richer, Smater, Greener, Healthier and Happier*. Penguin Press, New York, NY.

Glaeser, Edward L. and Joshua Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity* 2 (1):155–239.

Government of India Ministry of Housing & Urban Poverty Alleviation. 2016. "Pradhan Mantri Awas Yojana: Housing for All (Urban) Scheme Guidelines."

Henderson, J. Vernon, Anthony Venables, and Tanner Regan. 2016. "Building the City: Sunk Capital, Sequencing, and Institutional Frictions." SRC Discussion Paper 196.

Januszewski, Jozef. 1968. "Index of Land Consolidation as a Criterion of the Degree of Concentration." *Geographia Polonica* 14:291–296.

Kaufmann, Daniel and John M Quigley. 1987. "The Consumption Benefits of Investment in Infrastructure: The Evaluation of Sites-and-services Programs in Underdeveloped Countries." *Journal of Development Economics* 25 (2):263–284.

Kessides, Christine. 1997. "World Bank Experience with the Provision of Infrastructure Services for the Urban Poor: Preliminary Identification and Review of Best Practices."

Kline, Patrick and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6:629–662.

Kuffer, Monika, Karin Pfeffer, and Richard Sliuzas. 2016. "Slums from space-15 years of slum mapping using remote sensing." *Remote Sensing* 8 (6).

Leaf, Michael Leon. 1993. "Land regulation and housing development in Jakarta, Indonesia: From the "big village" to the "modern city"." .

Libecap, Gary D. and Dean Lueck. 2013. "The Demarcation of Land and the Role of Coordinating Property Institutions." *Journal of Political Economy* 119 (3):426 –57.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri. 2013. "The Economics of Slums in the Developing World." *Journal of Economic Perspectives* 27 (4):187–210.

———. 2016. "There Is No Free House: Ethnic Patronage in a Kenyan Slum." Working paper.

McCarthy, Paul. 2003. "The case of Jakarta, Indonesia." *UN Habitat Case Study* .

Michaels, Guy, Dzhamilya Nigmatulina, Ferdinand Rauch, Tanner Regan, Neeraj Baruah, and Amanda Dahlstrand-Rudin. 2017. "Planning Ahead for Better Neighborhoods: Long Run Evidence from Tanzania." .

Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and Cesar A. Hidalgo. 2017. "Computer Vision Uncovers Predictors of Physical Urban Change." *Proceedings of the National Academy of Sciences* 114 (29):7571–7576.

NASA and METI. 2011. "ASTER Global Digital Elevation Model, Version 2." USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

Pekel, Jean-Francois, Andrew Cottam, Noel Gorelick, and Alan Belward. 2016. "High-resolution mapping of global surface water and its long-term changes." *Nature* 540:418–422.

Rakodi, Carole and Tommy Firman. 2009. "Planning for an Extended Metropolitan Region in Asia: Jakarta, Indonesia."

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6):1257–1278.

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter Mieszkowski and Mahlon Straszheim. Johns Hopkins University Press.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens. 2010. "Housing Externalities." *Journal of Political Economy* 118 (3):485–535.

Takeuchi, Akie, Maureen Cropper, and Antonio Bento. 2008. "Measuring the Welfare Effects of Slum Improvement Programs: The Case of Mumbai." *Journal of Urban Economics* 64 (1):65–84.

Turner, Matthew A, Andrew Haughwout, and Wilbert Van Der Klaauw. 2014. "Land Use Regulation and Welfare." *Econometrica* 82 (4):1341–1403.

UN Habitat. 2011. "Building Urban Safety Through Slum Upgrading."

UN-Habitat. 2012. "State of the World's Cities 2012/2013."

UN Habitat. 2017. "Participatory Slum Upgrading Program."

U.S. Army Map Service. 1959. "Djakarta - 1:50,000." Edition 1-AMS, Series T725.

World Bank. 1995. "Enhancing the Quality of Life in Urban Indonesia: The Legacy of Kampung Improvement Program."

———. 2011. "Jakarta - Urban challenges in a changing climate." Tech. rep., The World Bank, Washington, D.C.

———. 2016a. "Indonesia's Urban Story."

———. 2016b. "International Bank For Reconstruction and Development Project Apprasial Document on a Proposed Loan in the Amount of US$216.5 Million to the Republic of Indonesia for a National Slum Upgrading Project."

———. 2017. "Tanzania - Urban Water Supply Project."

Tables

Table 1: Summary statistics

Variable name	N	Mean	SD
Panel A: Sub-block level data			
Assessed land values (thousand Rupiahs per sqm)	19862	12386.49	14686.83
Distance to Monument (km)	19862	11.13	4.83
Distance to main road (km)	19862	6.91	4.09
Distance to stock exchange (km)	19862	10.49	4.30
1(Flood-prone hamlet)	19862	0.38	0.48
Elevation (m)	19862	22.90	15.91
Slope (degrees)	19862	4.97	3.35
Distance to coast (km)	19862	11.84	7.48
Distance to nearest waterways (km)	19862	0.48	0.59
Distance to surface water occurrence (km)	19862	2.39	1.44
Panel B: Pixel level data			
Retail density	89463	0.02	0.10
Business density	89463	0.04	0.16
Parcel counts	89463	15.80	16.18
Average parcel area (sqm)	89463	1757.04	2182.16
K-Index	89463	0.56	0.31
Panel C: Photo survey data			
Building height (nr floors)	7104	4.21	7.50
Informality index	7104	1.11	1.12
Informality index > 0	7104	0.47	0.50

Notes: Panel A reports summary statistics for land values and pre-determined attributes for the full sample of assessed values database (19,862 sub-blocks with assessed land values). Panel B reports summary statistics for 89,463 (75 meter by 75 meter) pixels, used to capture land use patterns and land fragmentation measures. Panel C reports summary statistics for 7104 pixels, including building height, informality index and binary "kampung" indicator.

Table 2: Comparing KIP and non-KIP areas

Unit of analysis:	Sub-block level		Pixel level	
Sample:	Historical kampung	BDD 200m	Historical kampung	BDD 200m
	(1)	(2)	(3)	(4)
Log distance to Monument	-0.01	0.0007	0.004	0.004
	[0.25]	[0.90]	[0.53]	[0.25]
Log distance to main road	-0.01	-0.003	0.02	0.02
	[0.72]	[0.86]	[0.49]	[0.21]
Log distance to stock exchange	-0.01	-0.002	-0.01	-0.005*
	[0.24]	[0.57]	[0.41]	[0.10]
1(Flood-prone hamlet)	0.02	0.10*	-0.03	-0.001
	[0.55]	[0.05]	[0.17]	[0.96]
Elevation (m)	0.005	0.15	-0.36	0.07
	[0.99]	[0.75]	[0.11]	[0.78]
Slope (degrees)	0.15	-0.15	-0.24**	-0.21*
	[0.60]	[0.51]	[0.03]	[0.07]
Log distance to coast	-0.002	-0.003	-0.002	0.003
	[0.92]	[0.64]	[0.76]	[0.47]
Log distance to nearest waterways	0.12	0.27***	0.14***	0.23***
	[0.13]	[0.004]	[0.003]	[0.00003]
Log distance to surface water occurrence	0.002	-0.0008	-0.01	0.02
	[0.96]	[0.95]	[0.65]	[0.23]
Geography FE	Village	KIP Boundary	Village	KIP Boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports results from a series of OLS regressions with a pre-determined attribute as the dependent variable and the treatment indicator as the key regressor. For each variable, the top row reports the coefficient, and the bottom row reports the p-value in brackets. The unit of analysis for columns 1 and 2 is a sub-block in the assessed land values database, and the unit of analysis for columns 3 and 4 is a 75m×75m pixel. Columns 1 and 3 report results for the historical kampung's sample, and columns 3 and 4 report results for the boundary discontinuity sample, within a 200 meter distance band. We have 3 distance controls (log distance to the National Monument, main road, and stock exchange) and 6 topography controls. Standard errors are clustered by village in columns 1 and 3, and clustered by KIP boundary in columns 2 and 4.

Table 3: Effect of KIP on land values and building heights

Dependent variable:	**Log land values**			**Building heights**		
Sample:	Historical kampung	BDD 200m	Full sample	Historical kampung	BDD 200m	Full sample
	(1)	(2)	(3)	(4)	(5)	(6)
KIP	-0.12***	-0.13**	-0.11***	-1.64***	-1.40**	-1.25**
	(0.05)	(0.06)	(0.03)	(0.38)	(0.64)	(0.58)
N	3147	1345	19862	5280	1452	7104
R-Squared	0.73	0.81	0.85	0.32	0.52	0.43
Distance	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	N	Y	N	N	Y	N
Geography FE	Village	KIP Boundary	Hamlet	Village	KIP Boundary	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table reports results from an OLS regression with the log of assessed land values (column 1 through 3) and building heights (columns 4 through 6) as the dependent variables. The key regressor is the treatment indicator. For columns 1 through 3, the unit of analysis is a sub-block in the assessed land values database, and the estimation sample for column 3 consists of 19,862 sub-blocks. Column 1 restricts the estimation to the historical kampung's sample with 196 village fixed effects. Column 2 reports the boundary discontinuity design specification with observations within 200 meters from a KIP boundary, controlling for distance to the KIP boundary and with 124 KIP boundary fixed effects. Column 3 includes the full sample with 2060 hamlet fixed effects. For columns 4 to 6, the unit of analysis is a a 75m$\times$75m pixel, and the estimation sample for column 4 consists of 7,104 pixels. Columns 4 through 6 replicate the sample restrictions of columns 1 through 3, but for building heights (measured in number of floors) as the dependent variable, controlling for tercile distance dummies and an indicator for no building pixels. In addition, all columns include 9 attributes (defined in Table 2) as controls. Standard errors are clustered by village in columns 1, 3 ,4, and 6, and by KIP boundaries in columns 2 and 5.

Table 4: Effect of KIP on informality and land fragmentation

Dependent variable:	**1(Kampung)**			**Parcel count**		
Sample:	Historical kampung	BDD 200m	Full sample	Historical kampung	BDD 200m	Full sample
	(1)	(2)	(3)	(4)	(5)	(6)
KIP	0.15***	0.20***	0.14***	8.65***	12.54***	10.13***
	(0.02)	(0.05)	(0.03)	(1.07)	(1.13)	(0.55)
N	5280	1452	7104	11015	5412	89463
R-Squared	0.24	0.31	0.43	0.51	0.40	0.52
Distance	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	N	Y	N	N	Y	N
Geography FE	Village	KIP boundary	Hamlet	Village	KIP boundary	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table reports OLS regressions of the effect of KIP on informality land fragmentation. The unit of analysis is a 75m×75m pixel. Columns 1 through 3 examine the binary "kampung" indicator (as defined in Section 4.4). The specifications of columns 1 through 3 are analogous to columns 4 through 6 in Table 3. Columns 4 through 6 replicate the sample restrictions of columns 1 through 3, but for parcel count as the dependent variable. In addition, all columns include 9 attributes (defined in Table 2) as controls. Standard errors are clustered by village in columns 1, 3 ,4, and 6, and by KIP boundaries in columns 2 and 5.

Table 5: Land values, building heights and informality index on fragmentation within KIP

Dependent variable:	**Log land values**	**Building heights**	**1(Kampung)**
	(1)	(2)	(3)
Parcel count	-0.01***	-0.06***	0.01***
	(0.00)	(0.02)	(0.00)
N	1207	2699	2699
R-squared	0.80	0.37	0.27
Distance	Y	Y	Y
Topography	Y	Y	Y
Geography FE	Village	Village	Village

* 0.10 ** 0.05 *** 0.01

Notes: This table reports OLS regressions of log land values, building heights and a binary kampung indicator on parcel count, to assess to what extent land fragmentation explains the effect of KIP on those variables. The unit of analysis is a sub-block in column 1, and a 75m×75m pixel in columns 2 and 3. The key regressor is parcel count. The estimation sample is restricted to KIP areas. Standard errors are clustered by village.

Table 6: Effect of KIP on current amenities

Dependent variable:	Density		Log distance to			
	Retail	Business	School	Hospital	Police	Bus Stop
	(1)	(2)	(3)	(4)	(5)	(6)
KIP	-0.02***	-0.04**	0.04	0.07*	0.01	0.31***
	(0.00)	(0.02)	(0.06)	(0.04)	(0.04)	(0.06)
N	11015	11015	11015	11015	11015	11015
R-Squared	0.15	0.26	0.24	0.58	0.70	0.55
Distance	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y
Geography FE	Village	Village	Village	Village	Village	Village

* 0.10 ** 0.05 *** 0.01

Notes: This table reports specifications analogous to those in column 1 of Table 3, but the dependent variables are formal development density and access to amenities. The sample is the historical kampung's one. The dependent variables are share of retail (column 1) and business development within a pixel (column 2), log of distance to the nearest school, hospital, police station, and bus stop (columns 3 through 6). Standard errors are clustered by village.

Table 7: Heterogeneous treatment effects by KIP component

Dependent variable:	**Log land values**	
Sample:	Historical kampung	Full sample
	(1)	(2)
KIP	-0.08	-0.11***
	(0.05)	(0.04)
Length of Paved Roads (in km)	-0.015	0.001
	(0.012)	(0.013)
Number of Sanitation Facilities	0.001	0.002
	(0.007)	(0.008)
Number of Public Buildings	0.007	0.007
	(0.031)	(0.018)
KIP X Length of Paved Roads	-0.008	-0.001
	(0.013)	(0.010)
KIP X Number of Sanitation Facilities	0.005	-0.005
	(0.007)	(0.008)
KIP X Number of Public Buildings	-0.011	0.019
	(0.030)	(0.017)
N	3147	19862
R-Squared	0.73	0.85
Distance	Y	Y
Topography	Y	Y
Geography FE	Village	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table extends the historical kampung's sample specification in column 1 of Table 3 to report heterogeneous treatment effects by the three policy components (roads, sanitation, buildings) in column 1. Column 2 reports the same specification for the full sample. The unit of analysis is a sub-block in the assessed land values database. The intensity of KIP policy investments is measured by length of paved roads, number of sanitation facilities, and number of public buildings within a 500 meter buffer around each observation. The KIP intensity variables have been demeaned so that the coefficient on the treatment indicator reflects the effects when evaluated at average intensity levels. The omitted category is non-KIP areas. Standard errors are clustered by village.

Table 8: Heterogeneous treatment effects by distance from the city center

Dependent variable:	**Log land values**	**Building heights**
	(1)	(2)
Medium	0.005	0.52
	(0.05)	(1.34)
Central	-0.10	0.76
	(0.08)	(1.66)
KIP X Central	-0.14***	-1.42**
	(0.05)	(0.69)
KIP X Medium	-0.13**	-0.45*
	(0.06)	(0.26)
KIP X Peripheral	-0.01	-0.44
	(0.06)	(0.57)
N	19862	7104
R-Squared	0.85	0.43
Distance	Y	Y
Topography	Y	Y
Geography FE	Hamlet	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table reports regressions analogous to the full sample specification in column 3 of Table 3, augmented with heterogeneous treatment effects. We create dummy variables defined by the terciles of distance from the city center (defined as the National Monument). The omitted group is the periphery. The dependent variables are land values (column 1) and building heights (column 2). The unit of analysis is a sub-block in the assessed land values database in column 1, and a 75m$\times$75m pixel in columns 2 and 3. Standard errors are clustered by village.

Table 9: Effect of individual KIP waves on land values and building heights

Dependent variable:	**Log land values**		**Building heights**	
	Historical kampung	Full sample	Historical kampung	Full sample
	(1)	(2)	(3)	(4)
KIP I (1969-1974)	-0.13	0.01	-0.38	-0.34
	(0.09)	(0.07)	(0.47)	(0.60)
KIP II (1974-1979)	-0.09	-0.14**	-1.75***	-1.43**
	(0.06)	(0.05)	(0.52)	(0.63)
KIP III (1979-1984)	-0.14*	-0.09**	-1.04	-0.64
	(0.07)	(0.04)	(0.83)	(1.17)
N	3147	19862	5280	7104
R-Squared	0.73	0.85	0.32	0.43
Distance	Y	Y	Y	Y
Topography	Y	Y	Y	Y
Geography FE	Village	Hamlet	Village	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table reports the historical kampung's sample and full sample specifications of Table 3, disaggregating the treatment dummy by KIP wave. Column 1, 2, 3, and 4 respectively report specifications analogous to those in column 1, 2, 4, and 5 of Table 3, with the key regressors being three dummies corresponding to the three different KIP *pelita* (five-year plan) waves. Standard errors are clustered by village.

Table 10: Effect of historical kampung's on land values: boundary discontinuity

Dependent variable:	Log land values	
	BDD 200m	BDD 500m
	(1)	(2)
Historical Kampung	-0.001	-0.04
	(0.04)	(0.06)
N	1790	2628
R-Squared	0.49	0.49
Distance	Y	Y
Topography	Y	Y
Distance to historical kampung boundary	Y	Y
Geography FE	Historical Kampung Boundary	Historical Kampung Boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports results from an OLS regression with the log of assessed land values as the dependent variable. The key regressor is the historical kampung indicator. The unit of analysis is a sub-block in the assessed land values database. Column 1 reports results with historical kampung boundary fixed effects, 9 distance and topography attributes, tercile distance controls, and a control for distance to a historical kampung boundary for a sample of observations within 500 meters of a historical kampung boundary. Column 2 restricts the sample to sub-blocks within 200 meters of a historical kampung boundary. Standard errors are clustered by historical kampung boundary.

Figures

Figure 1a: Kampung and formal equilibrium

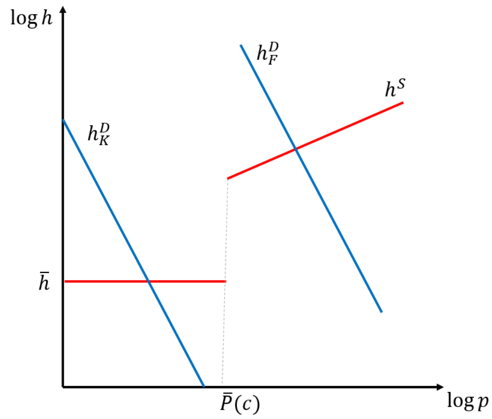


Figure 1b: Long-run, KIP neighborhood

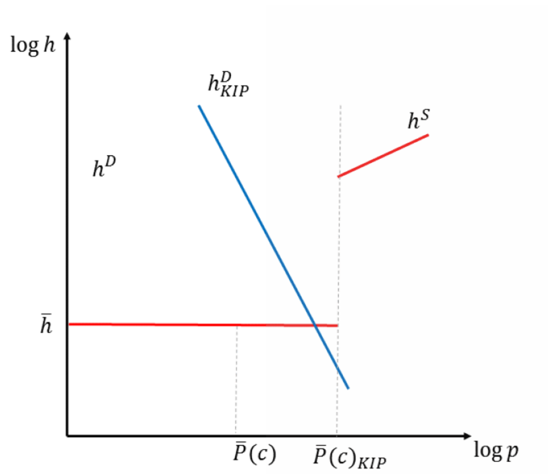


Figure 1c: Long-run, non-KIP neighborhood

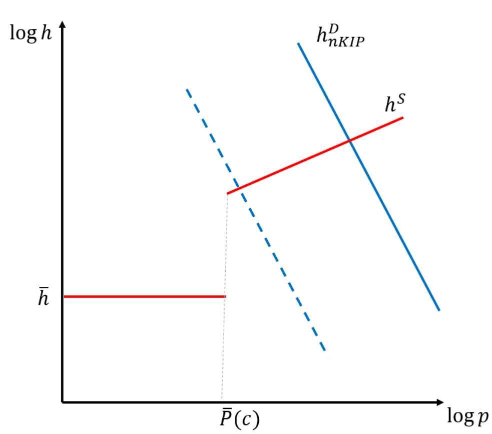
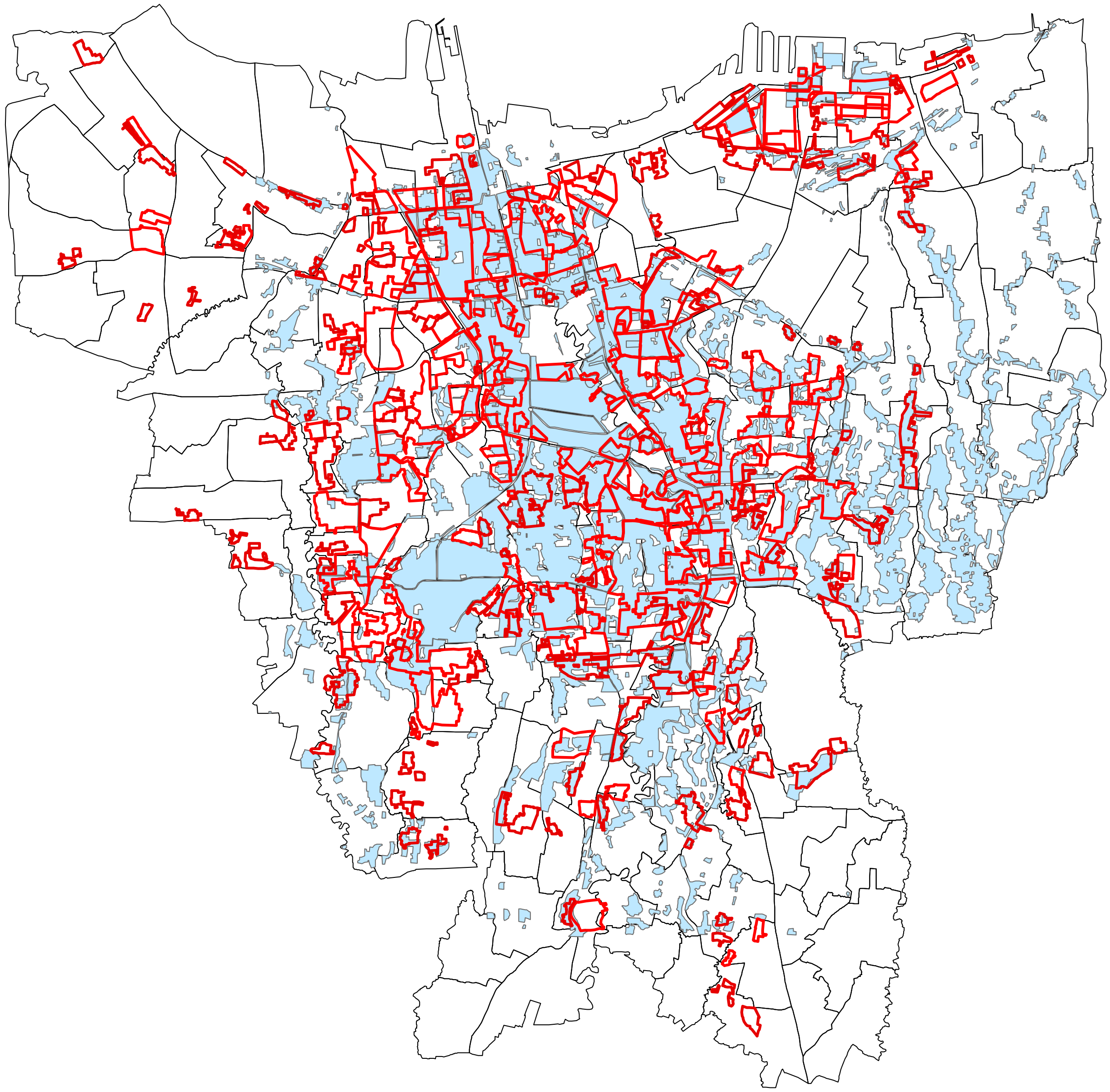
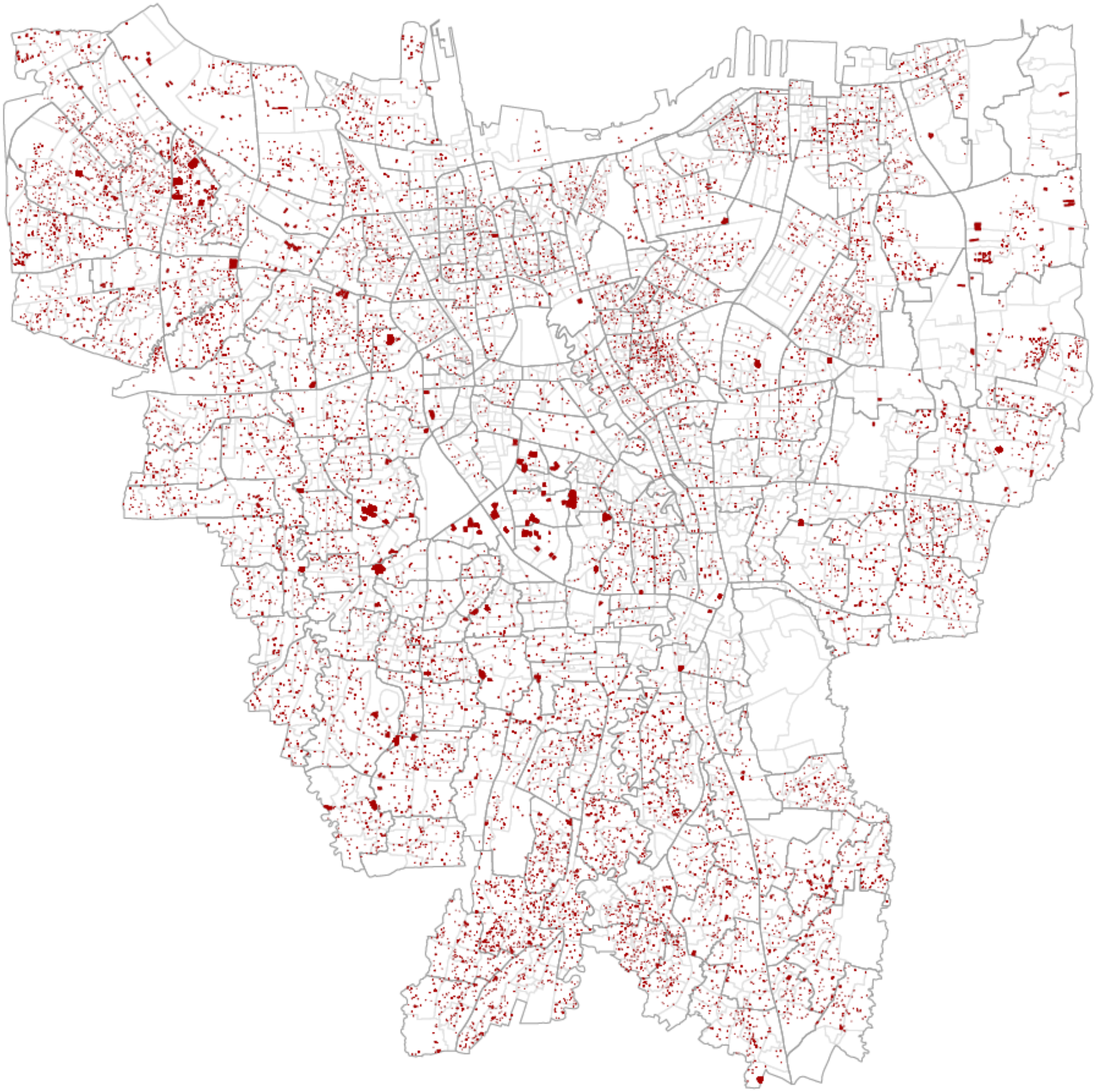


Figure 2: KIP boundaries and historical kampung's



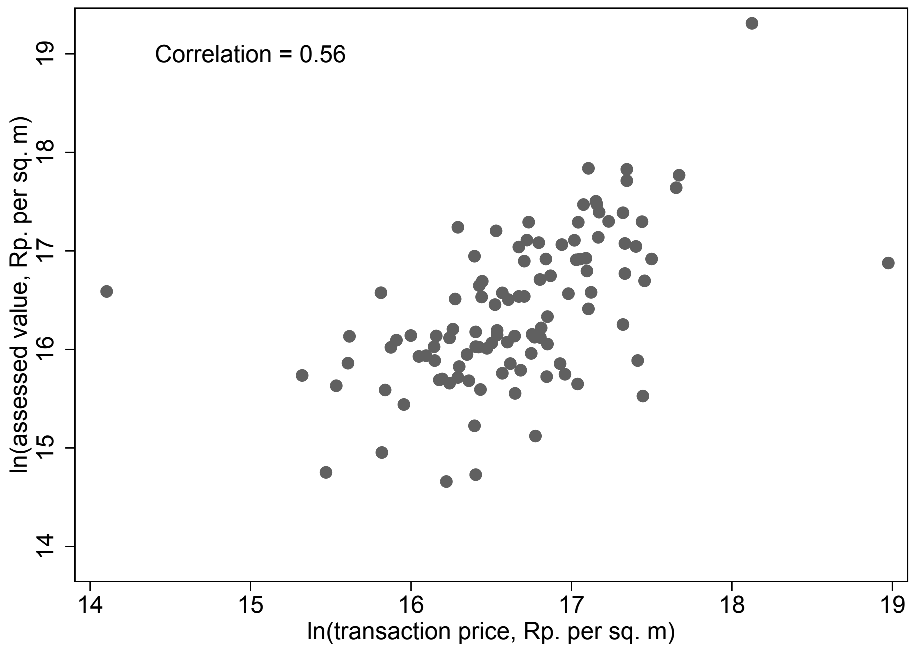
Notes: Map showing KIP boundaries (thick border) and historical kampung's that existed before KIP (shaded region).

Figure 3: Map of the assessed land values database



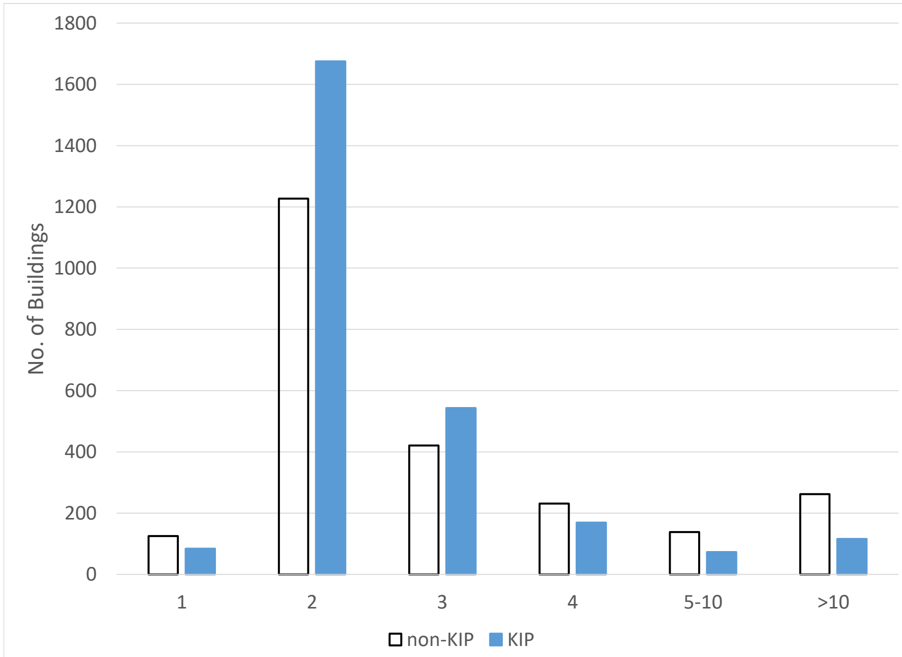
Notes: Map showing the coverage of the assessed land values database throughout Jakarta. Each shaded polygon corresponds to a sub-block. Thick boundaries correspond to villages. Light boundaries correspond to hamlets.

Figure 4: Scatterplot of assessed land values and transaction prices



Notes: This figure correlates the log of assessed land values and the log of transaction prices from www.brickz.id. Each point represents values averaged at the hamlet level.

Figure 5: Building heights histogram



Notes: This histogram shows the distribution of building heights for KIP and non-KIP pixels. The horizontal axis represents number of floors.

Figure 6: Informality index photos

0

1

2

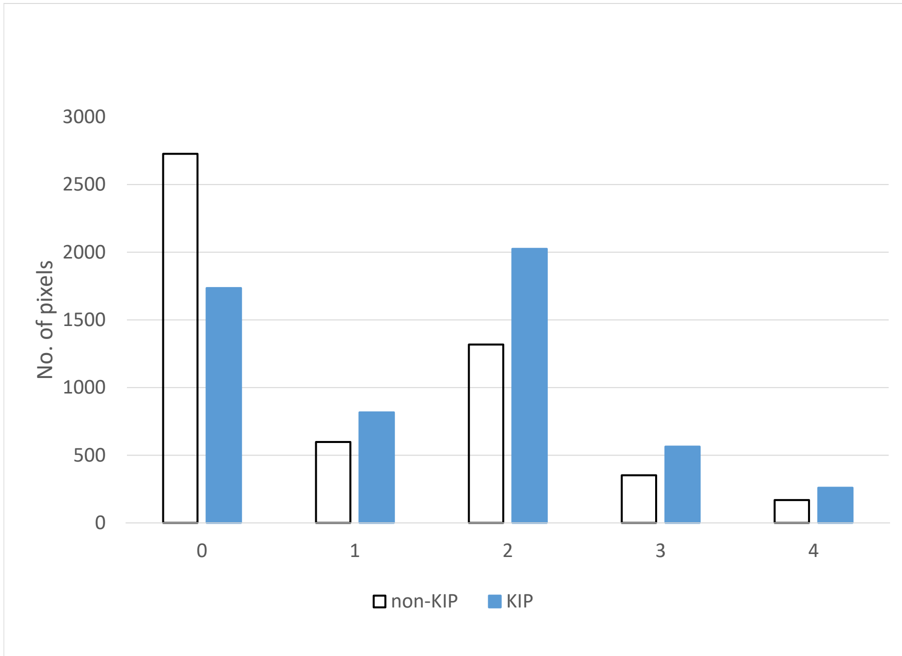
3

4

Figure 7: Example of cadastral map

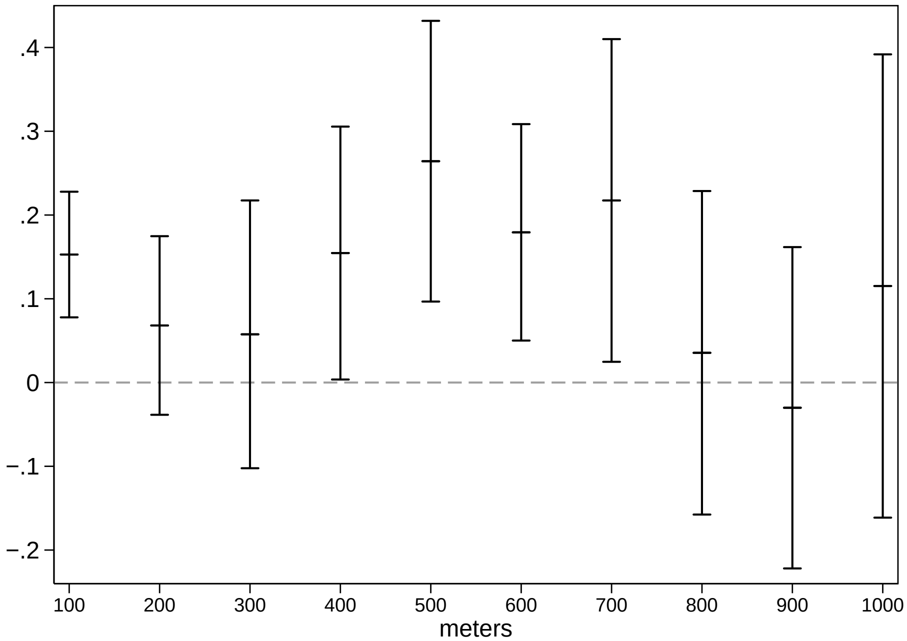
Notes: Map showing building footprints for one sub-district. The solid red boundaries indicate KIP treated areas. Smaller gray polygons represent building outlines.

Figure 8: Informality index histogram



Notes: This histogram shows the distribution of the informality index. Index values range from 0, corresponding to "very formal", to 4, corresponding to "very informal".

Figure 9: Effect of KIP on land values: Spatial decay



Notes: Each point corresponds to a coefficient and 95 percent confidence interval for the distance bins regressors in the specification reported in column 5 of Table A7.

Appendix Tables

Table A1: Comparing KIP and non-KIP areas, full sample

Sample:	**Full sample**	
	Sub-block level	Pixel level
Log distance to Monument	0.001	-0.001
	[0.69]	[0.43]
Log distance to main road	-0.01	0.01
	[0.79]	[0.45]
Log distance to stock exchange	-0.001	0.004
	[0.84]	[0.35]
1(Flood-prone hamlet)	..	..
	[..]	[..]
Elevation (m)	0.40	-0.04
	[0.37]	[0.80]
Slope (degrees)	-0.34	-0.20***
	[0.28]	[0.003]
Log distance to coast	-0.0006	0.001
	[0.83]	[0.39]
Log distance to nearest waterways	0.31***	0.24***
	[0.00]	[0.00]
Log distance to surface water occurrence	0.01	0.02*
	[0.48]	[0.07]
Geography FE	Hamlet	Hamlet

* 0.10 ** 0.05 *** 0.01

Notes: This table replicates Table 2, but for the full sample. The unit of analysis is a sub-block in the assessed land values database for column 1 and a 75m×75m pixel for column 2. Standard errors are clustered by village.

Table A2: Boundary analysis, robustness to different distance bands

Dependent variable:	**Log land values**			**Building heights**		
	BDD	BDD	BDD	BDD	BDD	BDD
	150m	300m	500m	150m	300m	500m
	(1)	(2)	(3)	(4)	(5)	(6)
KIP	-0.14**	-0.16***	-0.15***	-1.35*	-1.16***	-0.95***
	(0.06)	(0.05)	(0.05)	(0.70)	(0.39)	(0.33)
N	985	2044	2825	968	2601	3586
R-Squared	0.81	0.81	0.81	0.52	0.43	0.40
Distance	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	Y	Y	Y	Y	Y	Y
Geography FE	KIP Boundary	KIP Boundary	KIP Boundary	KIP Boundary	KIP Boundary	KIP Boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports robustness checks of the boundary discontinuity design in Table 3. Columns 1 to 3 report specifications analogous to those in column 3 of Table 3, and columns 4 to 6 repeat the analysis for building heights as a dependent variable. Standard errors are clustered by KIP boundary.

Table A3: Boundary analysis with village fixed effects, within 500 meters

Sample:	**BDD 500m**			
Dependent variable:	**Log land values**		**Building heights**	
	(1)	(2)	(3)	(4)
KIP	-0.15***	-0.12***	-0.95***	-1.54***
	(0.05)	(0.06)	(0.33)	(0.55)
N	2825	2825	3586	3586
R-Squared	0.81	0.86	0.40	0.43
Distance	Y	Y	Y	Y
Topography	Y	Y	Y	Y
Distance to KIP boundary	Y	Y	Y	Y
KIP Boundary FE	Y	Y	Y	Y
Village FE	N	Y	N	Y

* 0.10 ** 0.05 *** 0.01

Notes: This table replicates the boundary discontinuity specifications of Table 3 but for a distance band of 500 meters. The even columns include village fixed effects. Standard errors are clustered by KIP boundary.

Table A4: Robustness checks for building heights

Dependent variable:	**1(Height>3)**		**Log(height)**		**Height, exclude if outside pixel**		**Height, replace from StreetView**	
Sample:	Historical kampung	BDD 200m	Historical kampung	BDD 200m	Historical kampung	BDD 200m	Historical kampung	BDD 200m
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KIP	-0.12***	-0.09***	-0.19***	-0.13**	-1.75***	-1.43**	-1.69***	-1.43**
	(0.02)	(0.03)	(0.04)	(0.06)	(0.41)	(0.63)	(0.39)	(0.63)
N	5280	1452	5064	1415	5000	1368	5280	1452
R-Squared	0.29	0.37	0.37	0.52	0.34	0.55	0.34	0.54
Distance	Y	Y	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	N	Y	N	Y	N	Y	N	Y
Geography FE	Village	KIP boundary	Village	KIP boundary	Village	KIP boundary	Village	KIP boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports robustness checks of regressions of building heights on treatment. The odd columns report specifications analogous to those in column 4 of Table 3, and even columns report specifications analogous to those in column 5 of Table 3. In columns 1 and 2 the dependent variable is a dummy equal to one if the building height is greater than three floors. In column 3 and 4 we consider the log of building heights as a dependent variable. In columns 5 and 6 we drop observations outside the pixel. Column 7 and 8 replace those observations with heights observed from the closest possible Street View photo. Standard errors are clustered by village in odd columns and by KIP boundaries in even columns.

Table A5: Robustness checks for informality index

Dependent variable:	**1(Informality index > 0)**		**1(Informality index > 2)**		**Informality index**		**Informality index, pooling**	
Sample:	Historical kampung	BDD 200m	Historical kampung	BDD 200m	Historical kampung	BDD 200m	Historical kampung	BDD 200m
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KIP	0.12***	0.15***	0.05***	0.05	0.29***	0.39***	0.29***	0.39***
	(0.02)	(0.04)	(0.02)	(0.04)	(0.05)	(0.12)	(0.05)	(0.11)
N	5280	1452	5280	1452	5280	1452	10560	2904
R-Squared	0.28	0.33	0.13	0.22	0.25	0.34	0.23	0.30
Distance	Y	Y	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	N	Y	N	Y	N	Y	N	Y
Geography FE	Village	KIP boundary	Village	KIP boundary	Village	KIP boundary	Village	KIP boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports robustness checks of regressions of the kampung indicator on treatment. The odd columns report specifications analogous to those in column 4 of Table 3, and even columns report specifications analogous to those in column 5 of Table 3. The dependent variable for columns 1 and 2 is a dummy indicator to that is 1 if the informality index is greater than zero. The dependent variable for columns 3 and 4 is a dummy indicator to that is 1 if the informality index is greater than two. In columns 5 and 6 the dependent variable is the average of the continuous informality index across the two research assistants (RAs). Columns 7 and 8 report a pooled specification with the informality index as coded by both RAs, adding an RA fixed effect. Standard errors are clustered by village in odd columns and by KIP boundaries in even columns.

Table A6: Effect of KIP on land fragmentation, robustness to different fragmentation measures

Dependent variable:	**Parcel count**		**Avg. parcel area**		**K-Index**	
Sample:	Historical kampung	BDD 200m	Historical kampung	BDD 200m	Historical kampung	BDD 200m
	(1)	(2)	(3)	(4)	(5)	(6)
KIP	8.65***	12.54***	-557.21***	-966.25***	-0.12***	-0.19***
	(1.07)	(1.13)	(103.77)	(107.38)	(0.02)	(0.02)
N	11015	5412	11015	5412	11015	5412
R-Squared	0.51	0.40	0.37	0.35	0.43	0.40
Distance	Y	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y	Y
Distance to KIP boundary	N	Y	N	Y	N	Y
Geography FE	Village	KIP boundary	Village	KIP boundary	Village	KIP boundary

* 0.10 ** 0.05 *** 0.01

Notes: This table reports OLS regressions of the effect of KIP on land fragmentation. The unit of analysis is a 75m×75m pixel. Columns 1, 3, and 5 respectively examine the effect of KIP on the count of buildings in a pixel, the average area of buildings in a pixel, the K index of a pixel (as defined in Section 4 in the paper). The sample is restricted to historical slums. Columns 2, 4, and 6 report the boundary discontinuity design specification with observations within 200 meters from a KIP boundary, controlling for distance to the boundary, for the same outcome variables. All columns control for distance and topography attributes and for the length of roads in each pixel. Standard errors are clustered by village in odd columns, by KIP boundaries in even columns.

Table A7: Effect of KIP on land values: Spatial decay

Dependent variable:	**Log land values**				
Sample:	Full sample			Historical kampung	
Distance to KIP boundary (m)	(1)	(2)	(3)	(4)	(5)
[1, 100]	0.18***	0.19***	0.10***	0.19***	0.15***
	(0.07)	(0.04)	(0.04)	(0.05)	(0.05)
[101, 200]	0.15*	0.25***	0.14***	0.16**	0.07
	(0.09)	(0.05)	(0.05)	(0.08)	(0.06)
[201, 300]	0.28***	0.27***	0.13***	-0.03	0.06
	(0.06)	(0.05)	(0.05)	(0.10)	(0.10)
[301, 400]	0.34***	0.27***	0.15***	0.13	0.15*
	(0.07)	(0.05)	(0.06)	(0.10)	(0.09)
[401, 500]	0.32***	0.26***	0.14**	0.30**	0.26**
	(0.08)	(0.05)	(0.06)	(0.15)	(0.10)
[501, 600]	0.39***	0.24***	0.08	0.15	0.18**
	(0.08)	(0.06)	(0.07)	(0.11)	(0.08)
[601, 700]	0.34***	0.21***	0.04	0.27*	0.22*
	(0.08)	(0.06)	(0.07)	(0.16)	(0.12)
[701, 800]	0.38***	0.21***	0.05	0.30**	0.04
	(0.08)	(0.06)	(0.08)	(0.15)	(0.12)
[801, 900]	0.31***	0.21***	0.07	0.10	-0.03
	(0.09)	(0.06)	(0.08)	(0.12)	(0.12)
[901, 1000]	0.42***	0.22***	0.04	0.42***	0.12
	(0.10)	(0.07)	(0.07)	(0.15)	(0.17)
N	14185	14185	14185	2886	2886
R-squared	0.53	0.75	0.86	0.55	0.75
Distance	Y	Y	Y	Y	Y
Topography	Y	Y	Y	Y	Y
Geography FE	District	Village	Hamlet	District	Village

* 0.10 ** 0.05 *** 0.01

Notes: This table replicates the assessed values analysis of Table 3, but the omitted group is the treatment group and the key regressors are dummies corresponding to different distance bands; each dummy equals 1 if the observation is not treated and within a given 100 meter distance band from KIP boundaries. Observations beyond 1000 meters are excluded. Standard errors are clustered by village.

Table A8: Selection into assessed values database

Dependent variable	**1(Has assessed values)**	
Sample	Full sample	Historical kampung
	(1)	(2)
KIP	0.03***	0.04***
	(0.00)	(0.01)
N	89463	11015
R-Squared	0.09	0.08
Distance	Y	Y
Topography	Y	Y
Geography FE	Hamlet	Village

* 0.10 ** 0.05 *** 0.01

Notes: This table reports pixel-level regressions for the full sample (column 1) and the historical kampung's sample (column 2), with the dependent variable being an indicator equal to 1 if the pixel includes an observation from the assessed land values database. The key regressor is the treatment indicator. Standard errors are clustered by village.

Table A9: Effect of KIP on assessed land values, dropping observations with high land fragmentation

Dependent variable	**Log assessed value**		
Sample	Historical kampung		
	(1)	(2)	(3)
KIP	-0.11*	-0.11*	-0.11**
	(0.06)	(0.06)	(0.05)
N	2187	2179	2180
R-Squared	0.75	0.74	0.74
Distance	Y	Y	Y
Topography	Y	Y	Y
Geography FE	Village	Village	Village
Land fragmentation	Count	Area	Kindex

* 0.10 ** 0.05 *** 0.01

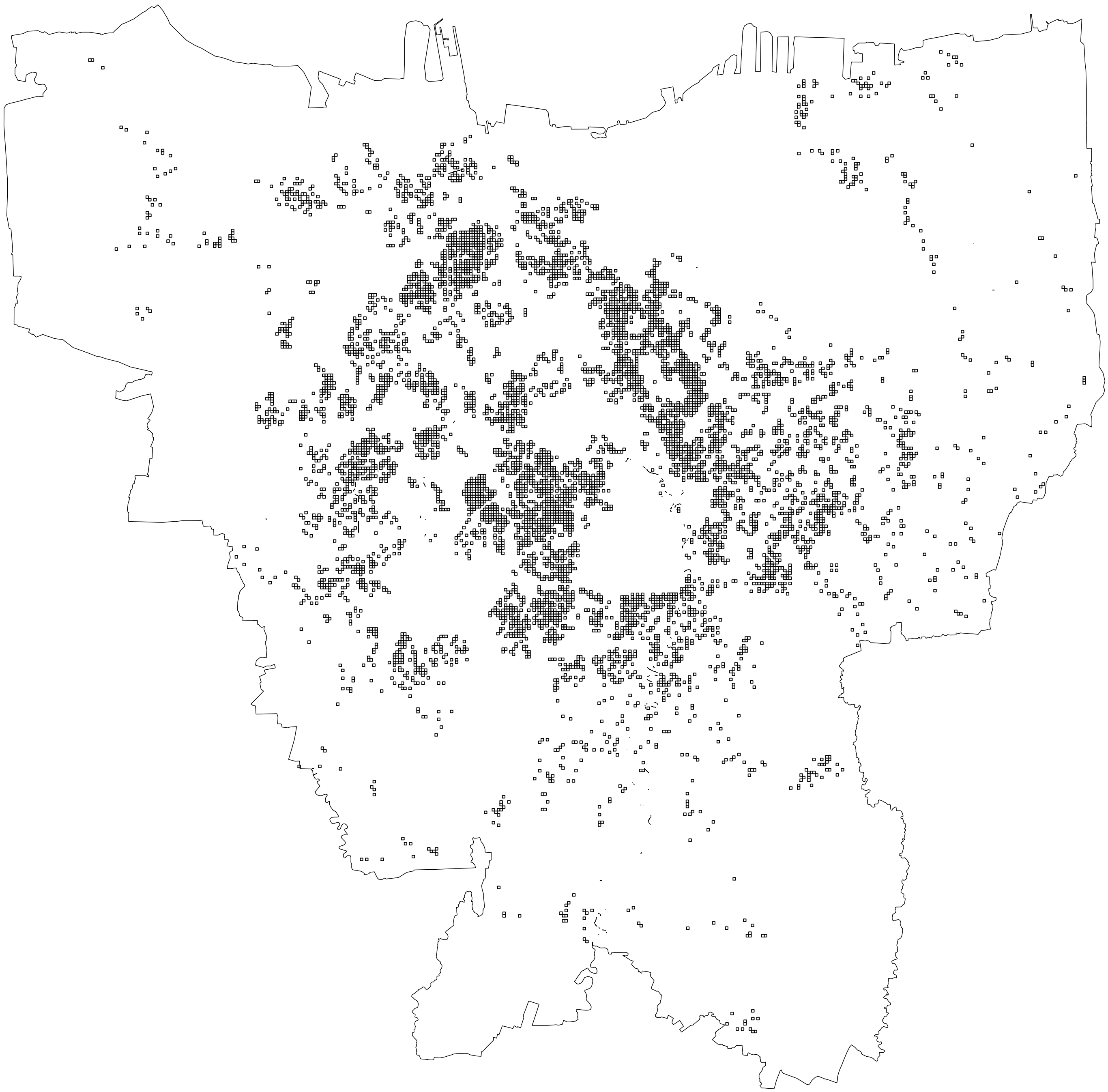
Notes: This table replicates the specification in column 1 of Table 3 but drops observations with land fragmentation in the top quartile, as measured by parcel count (column 1), average area of a parcel (column 2), and the K-index (column 3). Standard errors are clustered by village.

Appendix Figures

Figure A1: KIP boundary segments

Notes: Map showing KIP boundary segments selected for the boundary discontinuity design.

Figure A2: Photographic survey locations



Notes: Map showing the 7104 pixels sampled for the photographic survey.

Appendix

A Model

The building blocks of the model are housing supply, housing demand and spatial equilibrium.

Housing supply

Developers produce housing space H in a given location using land l and building height h, where $H = l \cdot h$. Construction costs are convex in height, but in order to build above a certain height threshold $\bar{h}$ there is an additional location-specific fixed cost c that developers need to incur. The cost of building at height h over a land surface l is

$$C(h,l) \quad = \quad \left(h^{\delta} + cI_{[h>\bar{h}]}\right) l \text{ with } \delta > 1 \text{ and } c > 0. \tag{A.1}$$

Buildings with a height below and above $\bar{h}$ can be thought of as "kampung" and "formal" respectively.

Assuming all locations have the same fixed amount of land normalized to 1 and denoting the location-specific price of housing per square meter as p, the developers' profit maximization problem for a given location reads:

$$\max_{h} \pi(h) \ , \ \pi(h) := \left\{ph - h^{\delta} - cI_{[h>\bar{h}]}\right\}. \tag{A.2}$$

Define $h^{*}(p) := \arg\max_{h} ph - h^{\delta} = \left(\frac{p}{\delta}\right)^{\frac{1}{\delta-1}}$. We assume that $\bar{h}$ is small enough, in particular that $\bar{h} < h^{*}(p)$. Otherwise, developers will choose $\bar{h}$ and the neighborhood will always be a kampung regardless of the extent of formalization costs c. The fact that Jakarta has started formalizing suggests that we are no longer in this region.

Developers choose to formalize, i.e. supply at a height $h^{*}(p)$ above $\bar{h}$, if

$$p\bar{h} - \bar{h}^{\delta} < ph^{*} - (h^{*})^{\delta} - c. \tag{A.3}$$

Otherwise, they choose the corner solution $\bar{h}$. This can be shown graphically in Figure A.1, in which we depict two examples of profit functions $\pi(h)$ for given p. For a fixed cost of c, the profit function is represented by the blue solid line and the profit-maximizing h is $h^{*}(p)$. For a fixed cost $c' > c$, the profit function is represented by the union of the solid blue line for for $h < \bar{h}$ and the solid green line for $h > \bar{h}$. In this case the profit-maximizing h is $\bar{h}$. In the data, we observe bunching at a height of 2 floors in kampung's, which is consistent with our prediction of a corner solution.

Equation A.3 simplifies to

$$\bar{h}p - \bar{h}^{\delta} + c \leq Ap^{B} \tag{A.4}$$

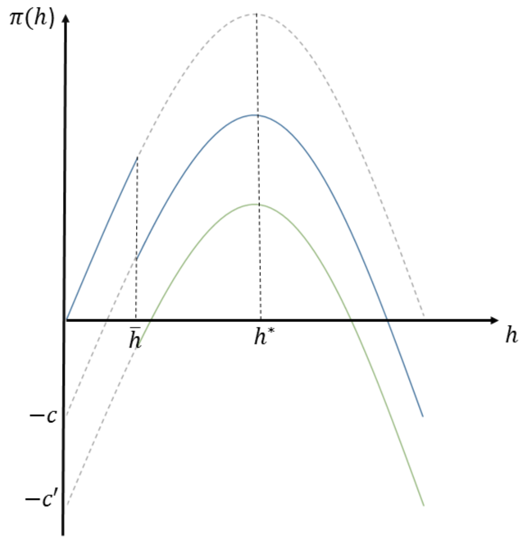
where $A := \frac{\delta-1}{\delta}\frac{1}{\delta^{\frac{1}{\delta-1}}} > 0$ and $B := \frac{\delta}{\delta-1} > 1$. Provided that $h < (c)^{\frac{1}{\delta}}$, the condition in equation A.4 is satisfied for p greater than a given threshold $\bar{P}(c)$ that is increasing in c. This can be show graphically in Figure A.2, where the shaded region on the p axis represents the range of prices for which developers do not formalize.

Housing supply in a given location is thus given by

$$h^{S} = \begin{cases} \bar{h} \text{ if } p \leq \bar{P}(c) \\ \left(\frac{p}{\delta}\right)^{\frac{1}{\delta-1}} \text{ if } p > \bar{P}(c) \end{cases} \tag{A.5}$$

where the threshold $\bar{P}$ is increasing in c. Intuitively, when formalization costs are high, there is a narrower range of prices for which developers are willing to formalize.

Figure A.1: Profit-maximizing h



Note that, because the land area in each location is normalized to 1, h^S is also equal to the total supply of housing space.

Housing demand

Assume there is a homogeneous population of consumers who move to the city as long as they can attain positive utility there. Once in the city, they optimally choose in which neighborhood to locate, facing no mobility costs. Each location is characterized by a bundle of exogenous amenities θ and by an endogenous amenity β such that

$$\beta \begin{cases} = 1 \text{ if } h \leq \bar{h} \\ > 1 \text{ if } h > \bar{h}. \end{cases} \tag{A.6}$$

The utility of consumers depends positively on amenities θ and β in each location and on the consumption of a numeraire good C and housing space H. Assume they all face the same exogenous income normalized to 1. Conditional on the choice of neighborhood, their utility maximization problem reads:

$$\max_{C,H} \theta\beta C^{1-\alpha} H^{\alpha} s.t.\, C = 1 - pH \tag{A.7}$$

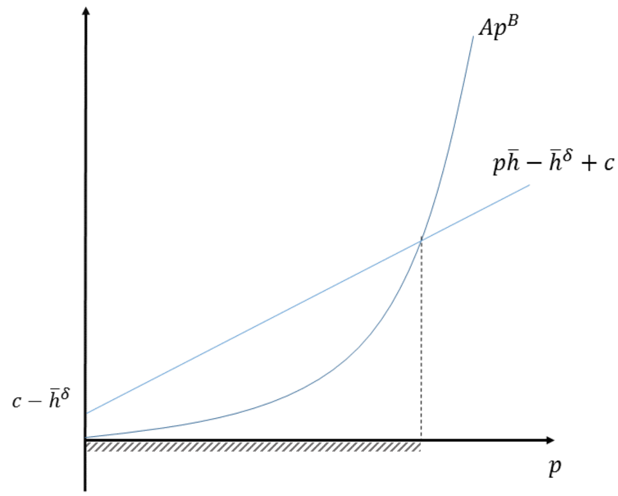
where $\alpha \in (0,1)$ is identical across consumers and locations. Each household optimally consumes

$$H^* = \frac{\alpha}{p}. \tag{A.8}$$

Denoting population in a given location as N, aggregate demand for housing space is thus equal to to

$$h^D = \frac{N\alpha}{p}. \tag{A.9}$$

Figure A.2: Range of prices in the kampung



Equilibrium

To close the model, we invoke spatial equilibrium. Housing markets are cleared such that consumers are indifferent across locations, otherwise they would move. Their equilibrium utility must be constant in all locations:

$$\theta\beta\left(1 - pH^*\right)^{1-\alpha} H^{*\alpha} = \overline{\upsilon} \tag{A.10}$$

which using A.8 can be rewritten as

$$\log p = \frac{1}{\alpha}(\log\theta + \log\beta) + \Phi \tag{A.11}$$

where $\Phi := \log\left[(1-\alpha)^{\frac{\alpha}{1-\alpha}}\alpha\right]$.

From A.11 it follows that a neighborhood will formalize when its amenities are large enough relative to the fixed cost c, i.e. when

$$\log\theta + \log\beta > \alpha\bar{P}(c) + \alpha\Phi. \tag{A.12}$$

Using the optimality condition for developers (A.5), that for households (A.8), and spatial equilibrium (A.10), the model can be solved for the three endogenous variables housing prices p, population N, and building height h as a function of the model's parameters, in particular the location-specific ones such as amenities θ and β and the fixed cost c.

The first step is to solve for the housing market equilibrium as a function of N, equating aggregate housing supply (A.5) and housing demand (A.9). The solution is depicted graphically in Figure 1a. When the aggregate demand in a neighborhood is low, the housing market equilibrium is one in which the neighborhood is a kampung with buildings at a height of $\bar{h}$. In Figure 1a this equilibrium corresponds to the intersection between h_K^D and h^S, where the red line represents supply of housing space and the blue line represents demand. When the aggregate demand for space is high enough, it will intersect the housing supply curve in its upward sloping portion and the neighborhood will be formalized, with building heights greater than $\bar{h}$. In Figure 1a this corresponds to the intersection of h_F^D and h^S. [27]

[27]Given the discontinuity in the supply curve, there could be intermediate levels of demand for which the demand curve does

Denote prices and quantities in kampung's and formal neighborhoods with the subscripts K and F. Analytically, the housing market equilibrium in the kampung case is defined by the following:

$$\begin{cases} \log h_K^D = -\log p + \log\alpha + \log N \\ \log h_K^S = \log\bar{h} \end{cases}$$

which implies

$$\log p_K = -\log\bar{h} + \log\alpha + \log N. \tag{A.13}$$

The equilibrium in the formal case is defined by

$$\begin{cases} \log h_F^D = -\log p + \log\alpha + \log N \\ \log h_F^S = \frac{1}{\delta-1}\log p - \frac{1}{\delta-1}\log\delta \end{cases}$$

which implies:

$$\log p_F = \frac{\delta-1}{\delta}\log\alpha + \frac{1}{\delta}\log\delta + \frac{\delta-1}{\delta}\log N \tag{A.14}$$

$$\log h_F = \frac{1}{\delta}\log\alpha - \frac{1}{\delta}\log\delta + \frac{1}{\delta}\log N. \tag{A.15}$$

In order to pin down N, we equate the equilibrium price resulting from the housing market equilibrium (A.13 and A.14 respectively) with the price dictated by spatial equilibrium (A.11). Denoting the equilibrium population of kampung's and formal neighborhoods as N_K and N_F, we obtain:

$$\log N_K = \frac{1}{\alpha}\log\theta + \Phi - \log\alpha + \log\bar{h} \tag{A.16}$$

and

$$\log N_F = \frac{\delta}{(\delta-1)\alpha}(\log\theta + \log\beta) - \frac{1}{\delta-1}\log\delta - \log\alpha + \frac{\delta}{\delta-1}\Phi. \tag{A.17}$$

We can then solve for the equilibrium h in the formal neighborhood by plugging equation A.17 into A.15 and we obtain:

$$\log h_F = \frac{1}{\alpha(\delta-1)}(\log\theta + \log\beta) - \frac{1}{(\delta-1)}\log\delta + \frac{\Phi}{\delta-1}. \tag{A.18}$$

To sum up, the equilibria are as follows:

$$\text{Kampung:} \begin{cases} \log p_K = \frac{1}{\alpha}\log\theta + \Phi \\ \log h_K^S = \log\bar{h} \end{cases}$$

$$\text{Formal:} \begin{cases} \log p_F = \frac{1}{\alpha}(\log\theta + \log\beta) + \Phi \\ \log h_F = \frac{1}{\alpha(\delta-1)}(\log\theta + \log\beta) - \frac{1}{(\delta-1)}\log\delta + \frac{\Phi}{\delta-1}. \end{cases}$$

It can be easily verified that $p_F > p_H$ and $h_F > h_H$ (the latter follows from $\bar{h} < h^*$).

not intersect the supply curve. In Figure 1a, this corresponds to demand curves that do not intersect any segment of the solid red line, but intersect the dashed vertical line in correspondence of $p = \bar{P}(c)$. In such cases, there would be excess demand for housing. In our model, this case is ruled out in our because we assume that consumers move to the city only if they can attain positive utility there, which they would not be able to do if they cannot consume any housing there. Therefore, those demand curves do not arise in equilibrium.

B Data

Market transactions

We compare assessed property values using real estate transaction prices scraped from the Brickz Indonesia website (www.brickz.id). Brickz has been collecting data of property sales since January 2015; sales are reported for properties advertised in the Rumah123 website (www.rumah123.com), an online property portal advertising sales and rentals. We scraped all the data available for sales of apartments and houses in Jakarta as of October 2016. For each entry, Brickz reports number of rooms, square footage, sale price and a street address, with varying precision. By a combination of Google API and manual search, we were able to geocode about 3800 entries at the street and street number level.

Program boundaries

Our source for KIP program boundaries is the Jakarta Department of Housing (DPGP, 2011) publication discussed in Section 4.1. We employ this data source both to create our main explanatory variable - a binary indicator for whether a sub-block or a pixel falls within a KIP treated area - and for our boundary discontinuity exercise. Selecting boundary segments for the latter exercise involves an additional data processing step. In a standard boundary discontinuity setup, there is a single boundary and observations lying on either side of the boundary are univocally classified as treated or not. Our setting is more complicated since we have over 500 polygons corresponding to KIP treated areas; these polygons typically have irregular shapes and are sometimes located in close proximity to one another. As a result of this layout, units that are classified as "control" - i.e. outside KIP - relative to one particular boundary segment may be "treated" - i.e. within KIP - relative to another boundary segment. We thus exclude contaminated units by manually selecting a subset of "clean" boundary segments, following Turner, Haughwout, and Van Der Klaauw (2014). These boundary segments are displayed in Figure A1.

Geography

Our elevation and slope controls are computed based on the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (NASA and METI, 2011), with a resolution of 30 meters.

We control for local hydrology through two variables. The first is a binary indicator for whether a hamlet is classified as "flood-prone" in Openstreetmap, a collaborative mapping project which provides crowd-sourced maps of the world and is continuously updated. Our second measure is the distance, in logs, from the nearest permanent or semi-permanent water body. We detect water bodies drawing upon the European Commission Joint Research Centre's Global Surface Water Dataset (Pekel et al., 2016), a global 30 meter resolution raster map reporting the occurrence of water bodies from March 1984 to October 2015. We consider pixels corresponding to water for at least 50% of the sample period.

Other variables

In order to control for access, we include a number of distance controls, all expressed in logs: distance from the headquarters of the Indonesia Stock Exchange, from the National Monument in Merdeka (which we identify as the pre-determined city center), from the coast, from historical main roads (following a 1959 map) and from waterways. Roads and waterways are drawn from Openstreetmap.

Finally, boundaries of sub-city administrative divisions ("kelurahan" or villages and "rukun warga" or hamlets), as well as the location of schools, hospitals, police stations and bus stops are also drawn from Openstreetmap.