

In da house!*

Stefano Polloni[†]
Brown University

This Version: May 3, 2018

Abstract

We estimate economic spillovers from subsidized housing projects on local neighborhoods using housing transaction records and high-resolution census data in South Africa. With a differences-in-differences design, we find that projects lead to a X% decline in nearby housing prices. We attribute this decline to externalities from informal settlements constructed within housing projects.

*We are grateful to Matthew Turner and Jesse Shapiro for their feedback, advice and support.

[†]Department of Economics, Brown University, Box B, Providence, RI 02912 E-mail: stefano.polloni@brown.edu

1. Introduction

In developing countries, 30% of urban populations live in slums where households often suffer high rates of crime, low access to infrastructure, insecure property rights, and unsanitary conditions (United Nations [2015]). These negative externalities often combine to create poverty traps, preventing long-term economic development (Marx et al. [2013]). Governments have responded by replacing slums with new homes and moving slum dwellers to public housing projects. While these policies are often motivated by immediate health and economic benefits for recipients, economic theory emphasizes how high-quality public housing provides an incentive for neighbors to invest more in their own homes and communities, reducing negative externalities and steering communities away from poverty traps.

In this paper, we analyze the impacts of public housing on the development of surrounding neighborhoods. We study a large-scale housing program in South Africa, which has allocated over 4.3 million dwellings and houses over 13% of the total population (Department of Human Settlements [2012,2015]; GHS 2009-2013). Already one of the largest housing programs in the developing world, this program continues to respond to large backlogs in housing demand with an even mix of upgrading slum areas with new houses as well as constructing large public housing developments. This program is intended to not only serve as “a key strategy for poverty alleviation” for direct beneficiaries, but also generate community-wide benefits, “leveraging growth in the economy,” “combating crime, promoting social cohesion and improving quality of life for the poor,” and “utilizing housing as an instrument for the development of sustainable human settlements, in support of spatial restructuring” (Department of Human Settlements [2004]).

We combine administrative data on [XX] housing projects with data on property transactions, demographics, and building construction to measure the local impacts of these projects. We estimate a persistent [XX]% decline in home prices within [XX] meters of a project. We then use heterogeneity in the extent to which projects eradicate existing slums as a natural experiment to quantify negative externalities from slum areas. Our estimates imply that increasing slum density

by [XX]% leads to a [XX]% decline in local housing prices. We find some evidence that this effect scales non-linearly with the size of the housing projects, which has theoretical implications for the extent to which slum areas represent poverty traps.

To identify these effects, we use a differences-in-differences strategy leveraging both the exact timing of housing project construction as well as precise geographic proximity of surrounding areas. Like Diamond and McQuade [2016]’s similar analysis of public housing in the US, we leverage the substantial uncertainty in project timing due to difficulties coordinating many stakeholders and sources of funding (Tissington [2011]). Similar to the context in the US, we find little evidence that local housing markets anticipate these projects. We also build on Diamond and McQuade [2016] by using planned but unconstructed project areas as counterfactual areas, which help to control for demographic or housing market trends that might be correlated with the choice of project locations.

Our negative spillover estimates stand in contrast to a large literature in development that has focused on measuring housing impacts on direct recipients. Relying on small-scale experimental designs, previous studies have linked public housing to improvements in employment outcomes (Franklin [2016]), self-reported wellbeing (Galiani et al. [2017]; Devoto et al. [2012]), and child health outcomes (Cattaneo et al. [2009]). Data limitations both in finding a large enough sample of housing projects and in identifying outcomes at a precise spatial scale have prevented previous studies identifying spillover effects. Taken alongside these previous studies, our findings suggest that policymakers may benefit from weighing direct benefits to recipients against potential negative effects to the local housing neighborhoods in designing future housing policy.

To understand the welfare implications of this policy, we also develop a model similar to...

1.1. Where are houses built?

Between 2000 and 2010, subsidized housing efforts in South Africa have primarily focused on constructing and allocating single-story, two-room (30 to 40 square meter) dwellings to households in groups of 50 to 500 per project. These housing projects are evenly divided between two categories

(Department of Human Settlements [2012,2015]):

1. **Greenfield developments** involve the construction of housing projects primarily on undeveloped state-owned land although in some cases, municipalities will work with private developers to purchase inexpensive, undeveloped private land for these projects. Finding undeveloped land often requires policymakers to locate these projects far from city centers and economic opportunities.
2. **In-situ upgradings** replace existing informal settlements with housing developments.¹ Since informal settlements are often located closer to city centers, the resulting housing projects may provide better employment opportunities (Tissington [2011]).

Facing substantial housing demand, the Department of Human Settlements has continued to issue grants to provincial governments to keep the rate of yearly housing allocations roughly constant (Department of Human Settlements [2012,2015]). While the location and types of projects are determined by provincial and municipal governments, construction is subcontracted to private developers who also act as project managers assisting in the allocation of houses to beneficiaries (Durojaye et al. [2013]).

Since housing projects require coordination between many stakeholders, these projects often face unanticipated delays and cancellations due to labor and land procurement issues, difficulties gaining support from local government agencies, environmental impact assessments, and inadequate bulk infrastructure provision (Department of Human Settlements [2012,2015]). In one example, political disagreements with local stakeholders led to the abandonment of a large project in Gauteng (Dlmini [2017]).

1.2. Who are the beneficiaries?

The National Department of Human Settlements issues guidelines for eligibility and maintains an official waiting list for eligible households for greenfield developments. Eligibility requires

¹While in some cases these programs refer simply to the provision of land titles and municipal services (water, electricity, etc.), this paper focuses on cases where informal settlements are replaced by fully-serviced, single-story houses.

citizenship, no previous property ownership, being married or having financial dependents, and having a monthly household income below R3,500 (Durojaye et al. [2013]).² The share of households reporting at least one member on the waiting list has remained stable at over 13% from 2009 to 2013.³ Before construction, each project is assigned beneficiaries in a first-come, first-served basis according to the waiting list in their province or municipality. For in-situ upgrading projects, previous inhabitants of informal settlements receive renovated houses while any remaining houses are allocated according to the housing waiting list.

In practice, these guidelines are only loosely followed. Recent reports point to cases of corruption in the allocation of houses while in some instances, housing projects are organized with the assistance of local community groups who ultimately select the beneficiaries (Durojaye et al. [2013]; Mathoho [2010]). Research suggests that Beneficiaries are often selected over the course of project construction and sometimes even after construction has finished (Durojaye et al. [2013]).

Beneficiaries are expected to pay a small one-time payment in order to receive title for their houses. Guidelines also prevent beneficiaries from reselling their houses (and discourage renting) within their first 7 years of ownership. Despite these guidelines, only 82% of project houses are reported being still occupied by their original beneficiaries within five years of construction.⁴ Anecdotal evidence suggests that project managers are aware of this active secondary market but have difficulty policing these transactions (Matsena [2018]).

2. A Model of Residential Choice and Public Housing

We develop a simple model of residential choice in order to both guide the reduced form empirical analysis as well as provide a framework evaluating the welfare effects of public housing policy. The key insight of the model is that while public housing projects provide high-quality houses and infrastructure, they also cross-subsidize informal housing markets, which can exacerbate conges-

²The Gauteng Province has implemented their own waiting list since 2008 in order to exert greater control over the allocation process.

³This figure is calculated from the General Household Surveys from 2009 to 2013

⁴This figure is calculated from the General Household Surveys from 2009 to 2013

tion externalities caused by these informal markets.

Basic Setup

A city is comprised of N residential neighborhoods indexed $n \in \{1, \dots, N\}$. Each neighborhood has a formal and informal housing sector, indexed $s \in \{F, I\}$. Each neighborhood is distinguished by its local housing supply and later by negative congestion externalities from informal housing. All jobs are in the central business district and pay a fixed wage rate, and commuting to the CBD is costly.

Housing Demand

Measure one of freely-mobile agents inelastically demand one unit of housing. An agent i has idiosyncratic tastes for neighborhoods, represented by the vector $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{2N}$ of neighborhood-sector specific valuations. The distribution of preferences in the population is given by the continuous and well-behaved function $f(\boldsymbol{\varepsilon})$ such that $\int f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = 1$. The indirect utility of individual i living in sector s of neighborhood n is given by:

$$\begin{aligned} u_{ins} &= A_{ns} - R_{ns} + \varepsilon_{ins} \\ &= v_{ns} + \varepsilon_{ins} \end{aligned}$$

where A_{ns} may be interpreted as the net-of-commuting and amenity-adjusted wage received in n when living in sector s , R_{ns} is the rental rate of housing, and ε_{ins} is individual i 's idiosyncratic taste for neighborhood-sector pair (n, s) . In this framework, neighborhood-sector tastes, ε_{ins} , are assumed to be uncorrelated across space and sector. This assumption rules out cases where households may prefer living closer to certain areas or in certain types of housing.

For notational convenience, we denote the array of quantities A_{ns} , R_{ns} and ε_{ins} in their matrix form as \mathbf{A} , \mathbf{R} and $\boldsymbol{\varepsilon}_i$. Individual i takes exogenous quantities $\{\mathbf{A}, \boldsymbol{\varepsilon}_i\}$ and endogenous rents \mathbf{R} as given, and locates in pair (n, s) yielding the highest indirect utility. Aggregate housing demand H_{ns}

in (n, s) is given by:

$$\begin{aligned} H_{ns} &= \int I\left(u_{ins} = \max_{n's'} \{u_{in's'} \mid \mathbf{A}, \mathbf{R}\}\right) f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\ &= D_{ns}(\mathbf{A}, \mathbf{R}) \end{aligned}$$

Housing Supply

Neighborhoods are endowed with L_n units of land. A share θ_{nF} of the total land stock is available for formal residential development, while the rest, $\theta_{nI} = 1 - \theta_{nF}$, is vacant or public land suited for informal housing. We assume that θ_{nF} is determined exogenously. In each sector, price-taking landlords supply housing by combining land and materials M with CRS technology $q(L, M)$. Materials are supplied at fixed cost c on a large national market. Because land is fixed in every market, the marginal cost of housing is increasing. The government may also choose to subsidize housing in market (n, s) at a rate of δ_{ns} per unit. The inverse supply curve in (n, s) is:

$$R_{ns} = S(H_{ns}, \theta_{ns} L_n) - \delta_{ns}$$

where $\partial S / \partial H > 0$ and $\partial S / \partial L < 0$. For notational convenience, we denote the array of quantities H_{nh} , L_n and θ_n in their vector form as \mathbf{H} , \mathbf{L} and $\boldsymbol{\theta}$, respectively.

Equilibrium

Given density $f(\boldsymbol{\epsilon})$ and the exogenous quantities $\{\mathbf{A}, \mathbf{L}, \boldsymbol{\theta}\}$, an equilibrium in this city consists of rents \mathbf{R}^* and housing quantities \mathbf{H}^* such that all housing markets clear, that is, $\forall(n, h)$:

$$\begin{aligned} H_{ns}^* &= D_{ns}(\mathbf{A}, \mathbf{R}^*) \\ \text{and } R_{ns}^* &= S(H_{ns}^*, \theta_{ns} L_n) - \delta_{ns}, \end{aligned} \tag{1}$$

and all agents live somewhere:

$$\sum_n \sum_s H_{ns}^* = 1$$

Welfare Implications of Housing Subsidies

The sum of individuals' utility is given by:

$$U = \int \max_{n's'} \{u_{in's'} \mid \mathbf{A}, \mathbf{R}^*\} f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}$$

and total landlords' profit is:

$$\begin{aligned} \Pi &= \sum_n \sum_s \left(\int_0^{H_{ns}^*} [R_{ns}^* - (S(x, \theta_{ns} L_n) - \delta_{ns})] dx \right) \\ &= \sum_n \sum_s \left(R_{ns}^* H_{ns}^* + \delta_{ns} H_{ns}^* - \int_0^{H_{ns}^*} S(x, \theta_{ns} L_n) dx \right) \end{aligned}$$

Total social welfare in this economy is therefore $W = U + \Pi$. We are interested in the welfare implications of the government subsidizing formal housing in a subset $\mathcal{N} \subset \{1, \dots, N\}$ of neighborhoods. Let $\delta_{ns} = \delta$ if $n \in \mathcal{N}$ and $s = F$, and $\delta_{ns} = 0$ otherwise. To characterize the marginal social benefit from the subsidy, $\partial W / \partial \delta$, we first note a result shown in Busso et al. [2013]:

$$\begin{aligned} \frac{\partial U}{\partial v_{ns}} &= \int \frac{\partial}{\partial v_{ns}} \max_{n's'} \{u_{in's'} \mid \mathbf{A}, \mathbf{R}\} f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\ &= \int I(u_{ins} = \max_{n's'} \{u_{in's'} \mid \mathbf{A}, \mathbf{R}\}) f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\ &= H_{ns} \end{aligned}$$

Using the above, we write the derivative of total utility with respect to subsidy δ as:

$$\frac{\partial U}{\partial \delta} = \sum_n \sum_s H_{ns}^* \left(-\frac{\partial R_{ns}^*}{\partial \delta} \right)$$

The derivative of total profits with respect to δ is:

$$\frac{\partial \Pi}{\partial \delta} = \sum_{n \in \mathcal{N}} H_{nF}^* + \sum_n \sum_s \left[H_{ns}^* \frac{\partial R_{ns}^*}{\partial \delta} + R_{ns}^* \frac{\partial H_{ns}^*}{\partial \delta} + \delta_{ns} \frac{\partial H_{ns}^*}{\partial \delta} - \frac{\partial H_{ns}^*}{\partial \delta} (R_{ns}^* + \delta_{ns}) \right]$$

Summing and simplifying:

$$\frac{\partial W}{\partial \delta} = \frac{\partial U}{\partial \delta} + \frac{\partial \Pi}{\partial \delta} = \sum_{n \in \mathcal{N}} H_{nF}^* \quad (2)$$

The total cost of the subsidy is given by $TC = \sum_n \sum_s H_{ns}^* \delta_{ns}$ and its marginal cost is thus:

$$\frac{\partial TC}{\partial \delta} = \sum_{n \in \mathcal{N}} \left(H_{nF}^* + \delta \frac{\partial H_{nF}^*}{\partial \delta} \right)$$

The extra term in this expression relative to (2) represents the marginal deadweight loss from an increase in δ . Unsurprisingly, we find that subsidies in an economy with perfect housing markets lead to economic inefficiencies. Furthermore, the magnitude of these inefficiencies depends critically on the population responses in subsidized markets.

Housing subsidies with Slum Externalities

Densely populated slums pose fire hazards, increase health risks and overburden existing public infrastructure, e.g. water and sewage networks. We consider an external utility cost from informal housing of the form:

$$A_{ns} = \bar{A}_{ns} + a \left(\frac{H_{nI}}{L_n} \right)$$

where $a'(\cdot) < 0$. With this specification, the private decision of locating in (n, I) negatively

impacts all residents in n because of congestion effects. The utility response to δ now depends on how both rents and amenities change in equilibrium:

$$\frac{\partial U}{\partial \delta} = \sum_n \sum_s H_{ns}^* \left(a' \left(\frac{H_{nl}^*}{L_n} \right) \frac{\partial H_{nl}^*}{L_n \partial \delta} - \frac{\partial R_{ns}^*}{\partial \delta} \right)$$

Therefore:

$$\frac{\partial W}{\partial \delta} = \sum_{n \in \mathcal{N}} H_{nF}^* + \sum_n a' \left(\frac{H_{nl}^*}{L_n} \right) \frac{\partial H_{nl}^*}{\partial \delta} \frac{(H_{nF}^* + H_{nl}^*)}{L_n}$$

Combining the equilibrium conditions in (1) and using the implicit function theorem, it is possible to show formally that $\partial H_{nl}^* / \partial \delta < 0 \forall n$. The extra term in the above expression – relative to equation (2) – represents the marginal welfare *gain* from reduced slum density. The subsidy δ makes formal housing in \mathcal{N} more attractive relative to informal housing. Marginal residents moving to \mathcal{N} make remaining residents better-off because of reduced congestion. The marginal deadweight loss (MDWL) of the subsidy is now :

$$\begin{aligned} MDWL &= \frac{\partial TC}{\partial \delta} - \frac{\partial W}{\partial \delta} \\ &= \sum_{n \in \mathcal{N}} \delta \frac{\partial H_{nF}^*}{\partial \delta} - \sum_n a' \left(\frac{H_{nl}^*}{L_n} \right) \frac{\partial H_{nl}^*}{\partial \delta} \frac{(H_{nF}^* + H_{nl}^*)}{L_n} \end{aligned}$$

Efficiency considerations in this setting depend on the population responses in subsidized markets, but also on responses in informal markets and on the shape of slum externality $a(\cdot)$. We note that $\delta = 0$ implies $MDWL < 0$, that is, some level of subsidy δ is welfare improving when compared to no subsidy. This is expected since the social benefits exceed the private benefits of moving from informal to formal housing.

Subsidy Spillovers

South Africa's housing program involves low density housing projects as well as large improvements to surrounding infrastructure such as electricity, water, and sanitation to service these new houses. Both have the additional impact of lowering construction costs for informal housing by providing space in and around housing projects to set up new informal dwellings as well as improving access to bulk services. Since the policy effectively lowers construction costs for informal housing, we model this by assuming subsidies in the formal sector spillover to the informal sector at no additional cost for the government. Formally, we let $\delta_{ns} = \delta$ if $n \in \mathcal{N}$ and $s = F$, $\delta_{ns} = \alpha\delta$ if $n \in \mathcal{N}$ and $s = I$, and $\delta_{ns} = 0$ otherwise, where $\alpha \in \mathbb{R}^+$.

Construction cost spillovers for the informal sector alter the welfare implications in two ways. First, suppliers of informal housing benefit from an increase in profit due to the indirect subsidies. Second, for $n \in \mathcal{N}$, the sign of informal housing response $\frac{\partial H_{nl}^*}{\partial \delta}$ is now ambiguous and depends on the magnitude of α .⁵ We use the notation $\frac{\partial H_{nl}^*}{\partial \delta}(\alpha)$ to represent this dependence. When α is large, the indirect subsidies in (n, I) dominate and net-migration is positive, i.e. $\frac{\partial H_{nl}^*}{\partial \delta} > 0$, making incumbent residents in (n, I) and (n, F) worse-off. With similar derivations as above, we obtain:

$$\frac{\partial W}{\partial \delta} = \sum_{n \in \mathcal{N}} (H_{nF}^* + \alpha H_{nI}^*) + \sum_n a' \left(\frac{H_{nI}^*}{L_n} \right) \frac{\partial H_{nI}^*}{\partial \delta}(\alpha) \left(\frac{H_{nF}^* + H_{nI}^*}{L_n} \right)$$

and therefore:

$$MDWL = \sum_{n \in \mathcal{N}} \delta \frac{\partial H_{nF}^*}{\partial \delta} - \sum_{n \in \mathcal{N}} \alpha H_{nI}^* - \sum_n a' \left(\frac{H_{nI}^*}{L_n} \right) \frac{\partial H_{nI}^*}{\partial \delta}(\alpha) \left(\frac{H_{nF}^* + H_{nI}^*}{L_n} \right)$$

⁵ This can also be shown formally by combining the equilibrium conditions in (1) and using the implicit function theorem. For $n \in \mathcal{N}^C$, $\frac{\partial H_{nl}^*}{\partial \delta}$ remains unambiguously negative.

Empirical Predictions

First, in terms of direct impacts, this model predicts that the housing subsidy program increases local formal and informal housing, improves public services, and attracts informal housing growth in nearby areas.

Second, informal settlement growth nearby public housing projects exacerbates congestion externalities, depressing housing prices in nearby formal and informal housing markets relative to housing markets further away. While this model attributes decreases in nearby home prices to congestion externalities, other mechanisms that may generate a similar decline in home prices that are not modeled here.

One mechanism includes housing externalities for formal housing where changes in the quality of neighbor's houses affects an individual's own decision to invest in their house. If subsidy houses have low quality relative to their neighborhoods, then local homeowners may simply invest less in improving their houses. This theory implies that project houses have lower quality than preexisting formal houses, which can be tested in the data. Another mechanism involves allowing for heterogeneity in tastes for neighboring households. If subsidy programs target recipients unlike preexisting households (ie. lower income), then neighboring home prices may decline simply because recipients are undesirable neighbors. We can similarly test for evidence of changing demographics in project neighborhoods to address this alternative theory.

3. Data

Understanding the spillover impacts of public housing requires (1) outcome measures at high spatial and geographic resolutions as well as (2) a precise measure of the location, timing, and size of housing projects.

3.1. Measuring Spillover Outcomes

To measure formal housing market impacts, we use over 500,000 housing transactions from the South African National Deeds Office covering the universe of transactions for suburbs in the bottom 20% of the housing market between 2003 and 2011 in the Gauteng Province (including the Johannesburg metro area).⁶ These data include the price, exact location, plot size, buyer name, and seller name for each transaction. To isolate spillover effects, we focus on transactions occurring within 1.2 kilometers of a housing project bringing the final sample to around 100,000 transactions. We exclude the top 1% of prices as well as prices below 1,000 Rand, which are likely composed of measurement error or the exchange of property title between family members.

Since government deeds records are unlikely to capture informal housing markets, we also include a building census of all structures in the Gauteng Province in 2001 and 2011. Using a combination of high-resolution satellite imagery and local field teams, these data record the precise location of each structure, identifying structures within over 30 categories including formal and informal residential dwellings. Out of 3,817,840 structures, this building census includes 1,628,073 formal structures and 1,560,345 informal structures. These data serve as both outcome measures of informal housing development as well as additional measures of public housing construction.

For demographic and economic outcomes, we turn to the census of population for 2001 and 2011 as well as the yearly General Household Survey for 2005 through 2014. Both surveys include information about dwelling type, employment, income, and demographics for each household. The General Household Survey includes additional details on participation in housing programs over time for a random sample of around 35,000 households in the Gauteng Province. To spatially link these households in both samples to their corresponding housing projects, we conduct the analysis at the census block level, which is the smallest geography available with 17,840 blocks in the Gauteng Province and 4,383 within 1.2 kilometers of a housing project.

⁶The bottom 20% suburbs were selected relative to prices in 2003 and followed every year from 2003 to 2011. These data were provided by the Affordable Land and Housing Data Centre, which tracks affordable housing markets.

Figure 1: Top-Five Sellers

Seller Name	Observations
City Of Johannesburg Metropolitan Municipality	29,087
City Of Johannesburg	27,672
City Of Tshwane Metropolitan Municipality	24,780
Ekurhuleni Metropolitan Municipality	21,758
Gauteng Provincial Housing Advisory Board	13,058
Total Observations	549,704

3.2. Defining Housing Projects

Our primary measure of housing projects matches a set of housing program characteristics using the transaction records in the deeds data. This procedure results in spatially distinct clusters that form the definition of housing projects for the empirical analysis (Tissington [2011]). Due to data limitations, we are not able to differentiate between greenfield and in-situ upgrading projects in the analysis.

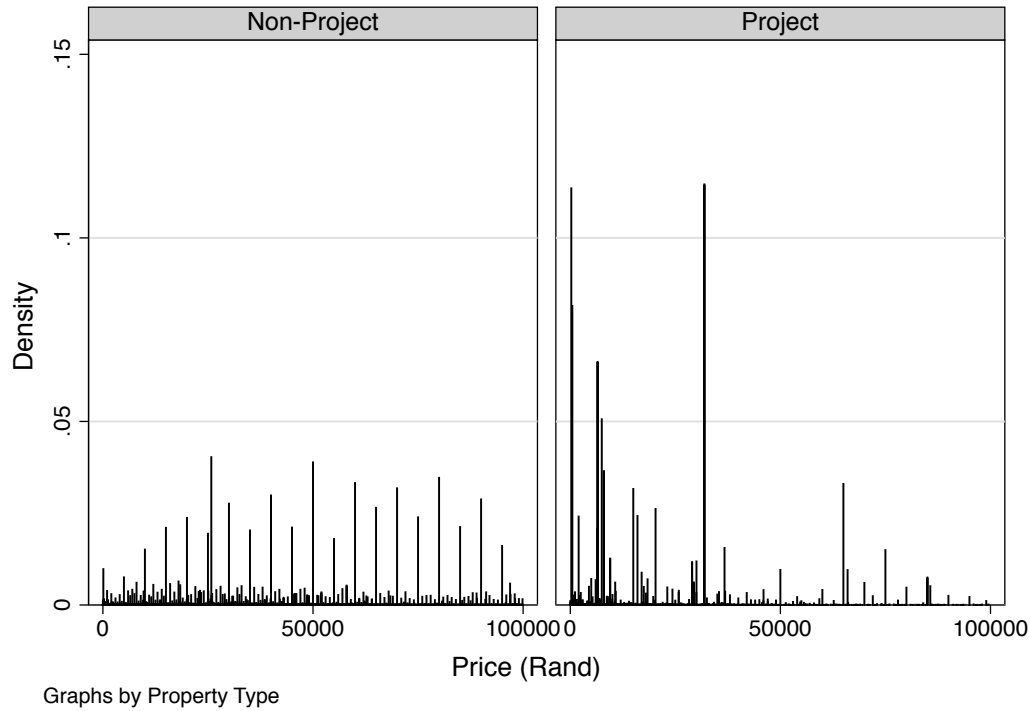
3.2.1. Seller Identity:

We focus on transactions from government entities or housing authorities from seller name records. To account for large developers or banks being listed as sellers for housing projects, we also include seller names that appear most frequently. We exclude transactions flagged as large buildings used for commercial purposes ($<2\%$ of transactions). Figure 1 shows the top five sellers, which account for a large share of the total transactions and all represent government housing programs in the region.

3.2.2. Subsidy Value:

We exclude transactions with purchase prices that are more than R50,000 above the yearly subsidy value ($<4\%$ of remaining transactions). Figure 2 provides a histogram of deed prices under R100,000 for transactions that meet our project definition and transactions that do not. Since project houses are distributed for free by government agencies, these agencies often record either

Figure 2: Transaction Price Histogram



Note: Transactions are censored at R100,000.

zero price or the value of the subsidy in the deeds, which produces substantial bunching at these values compared to non-project transactions. We find some prices scattered away from subsidy values consistent with some measurement error in deeds reporting or miscategorization of non-project properties as project properties.

3.2.3. Pre-Existing Formal Dwellings:

We exclude transactions that occur on land plots with at least one preexisting formal residential structure in the 2001 building census (31% of remaining records). This method not only helps to reduce error in matching seller identities to housing programs, but also works to distinguish new housing projects from titling, home loan, or other programs that may have been implemented by local housing agencies over the same time period.

3.2.4. Spatial Clustering:

Since housing projects are characterized by large plots of adjacent dwellings, we use a density-based clustering algorithm to group transactions that satisfy the above criteria according to their geographic proximity. By eliminating loosely clustered or singleton transactions, this method additionally helps to distinguish large housing projects from other small-scale housing programs.

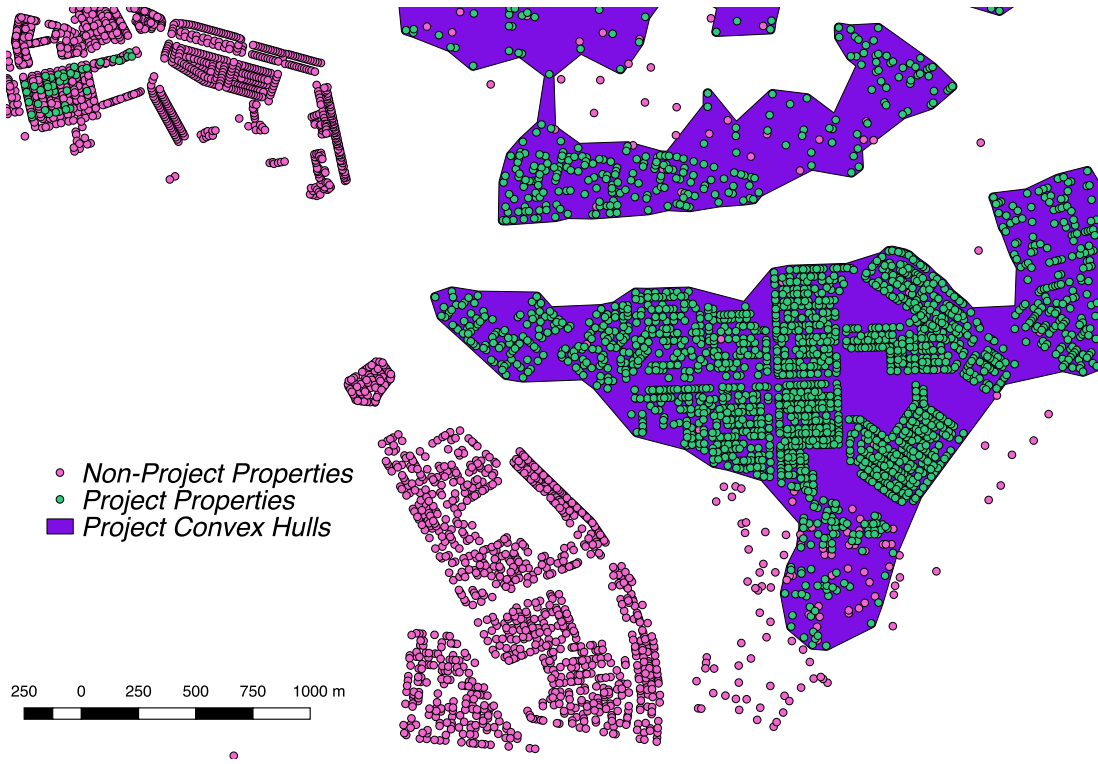
Figure 3 provides an example convex hulls formed around project housing transactions. The algorithm groups nearby transactions into two large projects on the right. Some non-project transactions (pink circles) are also included in the convex hulls, which are likely miscoded project houses or privately constructed houses within housing projects. The upper left collection of non-project houses contains a small group of houses coded as project houses (green circles). However, since there are very few of these houses and they are loosely clustered within a large group of non-project houses, they are not clustered into a convex hull. These more isolated cases are likely to be small-scale land-titling programs or home-financing programs launching by local NGOs in collaboration with the housing authorities.

3.2.5. Temporal Clustering:

Because the empirical design exploits the sudden timing of project construction, we assign a date for each cluster according to the modal year of transactions for houses within each project. We then include only clusters where over 50% of transactions occur during the modal year, which excludes half of total clusters. This approach rules out incremental projects as well as helps to exclude possible land titling programs, allowing us to leverage the sharp timing of large project construction in our identification.

Figure 4 indicates the distribution of transaction dates for properties within a 1.2 km buffer around clusters (above panel) and within the selected project areas (below panel). Project areas exhibit substantial bunching during a single month when projects were completed. There are also more transactions after the modal year than before the modal year, consistent with either a gradual roll-out for some project areas or immediate resale of projects houses after construction, which

Figure 3: Convex Hull Example



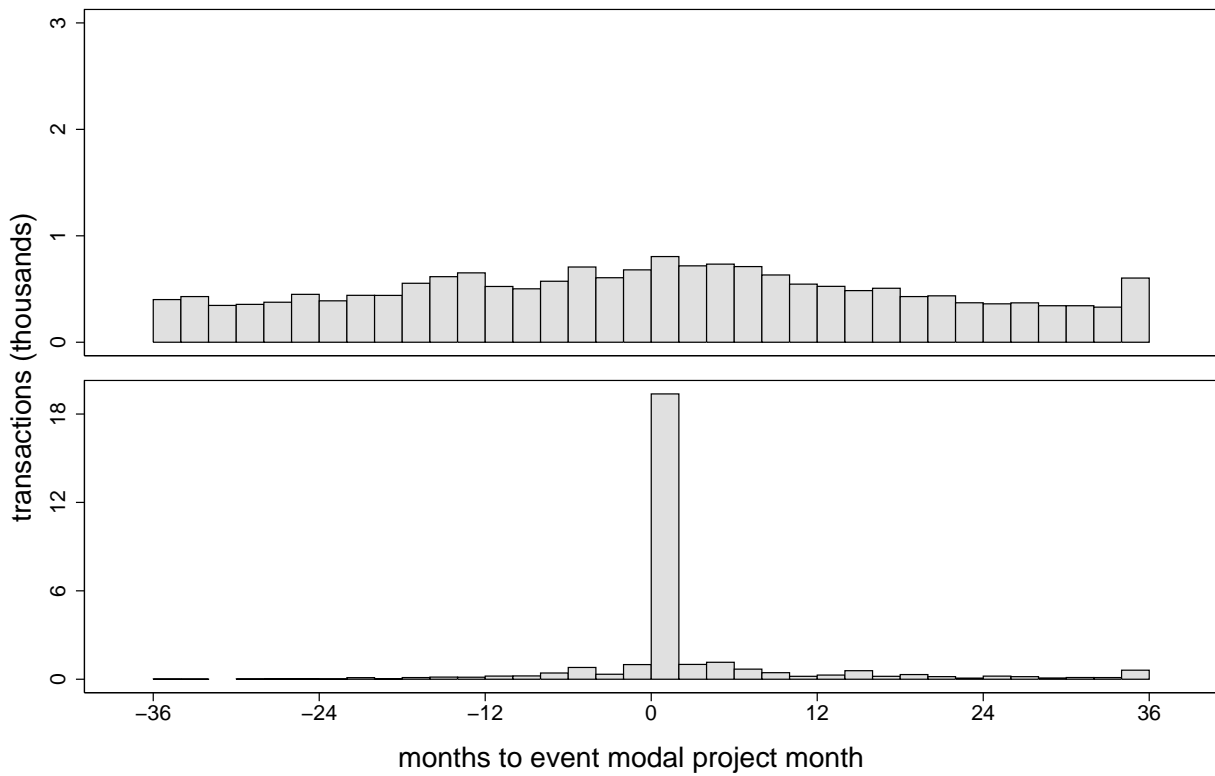
would be counter to housing regulations.

Evidence of similar bunching around the modal year for transactions coded as non-project transactions (above panel) would suggest that we may be miscoding project transactions as non-project transactions; instead, we find a smooth pattern relative to the modal year for these non-project transactions. The slight increase in density around and just following the mode year may also be consistent with housing projects having an immediate impact on local housing markets, which we will explore further below.

3.3. Defining Planned but Unconstructed Housing Projects

The administrative spatial data identifies many project areas that do not appear to have received housing projects, as measured by having fewer than 15 housing project transactions per square kilometer. We use these areas to create a sample of planned but unconstructed housing projects to serve as counterfactuals areas in our analysis.

Figure 4: Transaction Densities Relative to Modal Project Month



The above panel includes transactions within a 1.2 kilometer buffer around a project cluster relative to the modal year of that cluster. The below panel includes all transactions within the final sample of clusters.

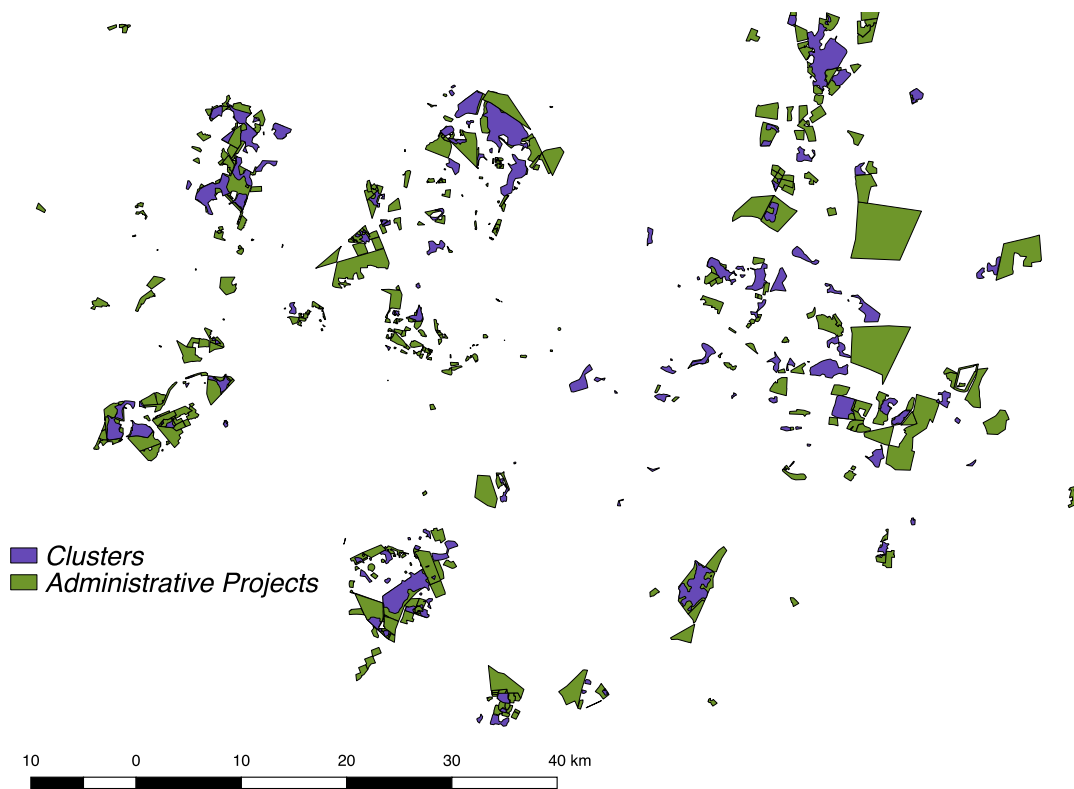
To construct an estimated completion date for planned but unconstructed housing projects, we digitized National Treasury budget reports that detail the start date, expected completion date, and cost of each housing project. We use a fuzzy-string matching algorithm with bigrams to successfully link project names from the budget reports for over 300 project names in our administrative spatial data (including both completed and uncompleted projects). Appendix 8 compares unmatched and matched projects finding that matched projects have a higher density of formal and informal houses although they are smaller in total area. We find that for completed projects, the mode transaction year observed in the deeds data falls an average of three years after the start date indicated in the budget reports. In other words, beneficiaries receive title to their new houses about three years after the housing program is announced in the budget. Therefore, we assign a expected completion date for unconstructed projects that is three years after the announced start-date in the project.

3.4. Assessing Project Measures

To assess the extent to which our definitions accurately measure housing projects, Figure 5 maps our cluster definitions on top of administrative data on housing projects. We see strong overlap between clusters and administrative projects, providing additional support for our deeds-based cluster measure. A few smaller clusters do not overlap with administrative definitions, which is consistent with the administrative data recording larger, higher-cost projects. Similarly, administrative boundaries that do not contain clusters are likely to be projects that were planned but were not completed or are scheduled to be completed in the future.

Table 1 provides descriptive statistics for the final sample of 56 completed and 51 uncompleted projects. At baseline in the 2001 building census, we find that completed and uncompleted project areas have similar formal building density although uncompleted projects have much higher densities of informal structures. Consistent with this finding, slum areas may face greater challenges in organizing funding as well as infrastructure provision necessary to complete large housing projects. As a result of project follow-through, completed areas experience dramatic increases in formal

Figure 5: Clusters and Administrative Project Boundaries



This figure includes the northern half of housing projects.

Table 1: Housing Project Descriptives

	Completed	Uncompleted
Formal Density: 2001	340.6	344.9
Formal Density: 2011	1,783.1	861.8
Informal Density: 2001	443.0	1,450.0
Informal Density: 2011	1,064.6	2,038.4
Median Year (est.)	2005	2006
Distance to CBD (km)	28.9	26.5
Total Projects	56	101

Density measures number structures per km².

Central Business Districts (CBD) are measured with respect to Johannesburg and Tshwane.

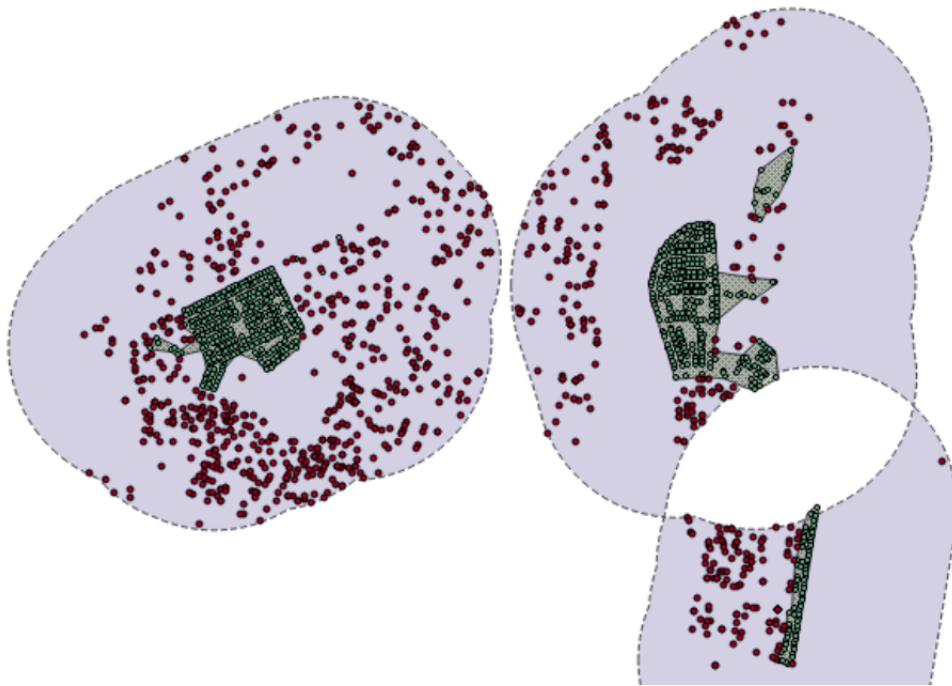
building density between 2001 and 2011 compared to uncompleted projects. Both areas also experience substantial and similar growth in their levels of informal settlement growth. The median year of completion as well as distance to Central Business District match well between completed and uncompleted projects

4. Descriptive Evidence for Spillover Areas

Examining the spillover effects of housing projects requires focusing the analysis on areas just around these housing projects, which may also limit the extent to which results may generalize more broadly. To identify the spillover effects of public housing projects, we focus on all outcomes within a 1.2 kilometer buffer around each cluster. Figure 6 shows an example. Areas where buffers overlap are excluded to ensure that the empirical exercise is able to recover the impacts due to the nearest housing project.

To assess the extent to which these project areas are representative, Table 2 average characteristics for transactions in buffer areas for completed projects, buffer areas for uncompleted projects, and all other transactions outside of all buffer areas. Average prices for completed projects closely resemble prices outside of buffer areas while uncompleted projects have much lower prices, supporting earlier evidence that these uncompleted projects are located in areas with greater slum

Figure 6: Buffer Design Properties Example



- Project Properties : red circles
- Non-Project Properties : dark green circles
- Buffer Area (1.2 km from clusters) : light blue area

Table 2: Property Prices within Buffer Areas

	Completed	Uncompleted	Other
Purchase Price (Rand)	248,181.0 [144,025.5]	230,410.1 [754,489.7]	243,484.9 [304,045.0]
Plot Size (m3)	819.2 [34,138.9]	865.2 [4,243.3]	1,888.5 [55,518.5]
Sold At Least Once	0.326	0.350	0.331
Median Purchase Year	2006	2006	2006
Observations	28,943	20,700	167,578

Standard errors are in parentheses.

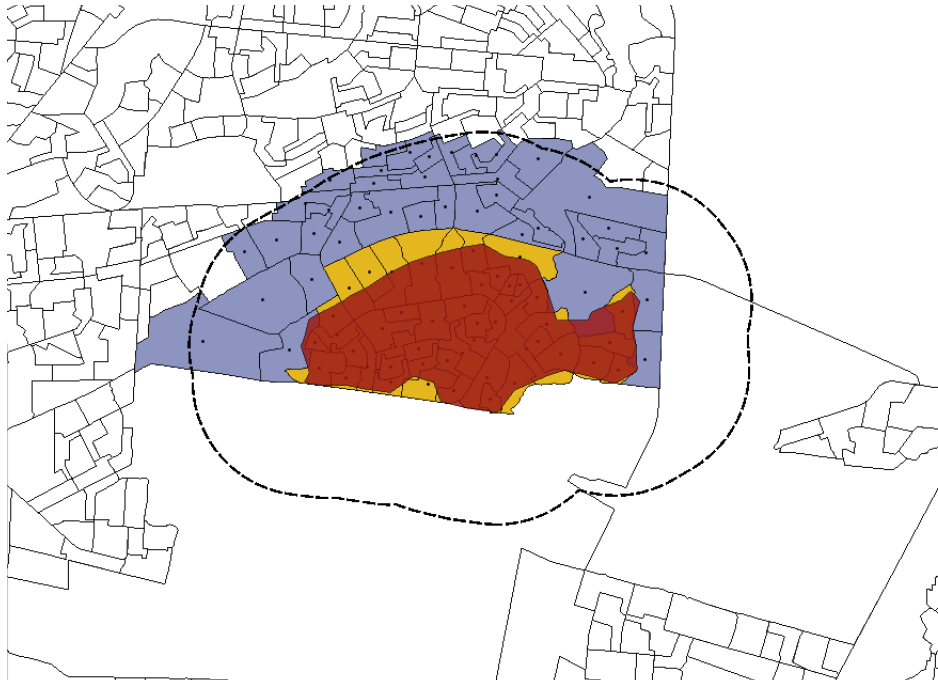
This figure excludes transactions within project areas.

density. Plot sizes for properties in completed and uncompleted project buffers are similar and much smaller than plots outside of these areas. This finding is consistent with housing projects being located nearby dense residential or slum areas instead of in rural locations throughout the province. Close to a third of all properties are sold multiple times and all share the same median purchase year.

We use a similar buffer approach to examine census characteristics by including census blocks with centroids that fall within 1.2 km buffers of projects. Figure 7 provides an example project that distinguishes between census blocks (in yellow) with greater than 30% overlapping area with project areas and census blocks (in blue) with less than 30% area overlap but whose centroids fall within the 1.2 km buffer. This criteria allows for separating the direct effects of the housing projects within their footprint from the spillover effects of housing projects on the surrounding areas. We focus on the 30% area overlap threshold to account for cases where housing project areas are small relative to the nearest census blocks.

To compare census characteristics for these areas at baseline, Table 3 provides averages for households within and nearby both completed and uncompleted projects. The first two columns find that relative to census blocks within uncompleted project areas, census blocks within completed projects have much better infrastructure (flush toilets, piped water, and electricity), but have

Figure 7: Buffer Design Census Block Example



- Census Blocks with $\geq 30\%$ project area overlap : yellow polygons
- Census Blocks with $< 30\%$ project area overlap : blue polygons
- Buffer Area (1.2 km from clusters) : dotted line
- Project Area : red polygon

Table 3: Census Characteristics in 2001
by Overlap with Project Areas

	Within Project (>30% Overlap)		Outside Project (<30% Overlap)	
	Completed	Uncompleted	Completed	Uncompleted
Flush Toilet	0.56	0.26	0.77	0.78
Piped Water	0.21	0.11	0.41	0.37
Owner	0.57	0.43	0.47	0.51
Elec. Cooking	0.58	0.24	0.68	0.63
Elec. Light	0.79	0.36	0.74	0.78
Single House	0.51	0.45	0.52	0.57
Number of Rooms	2.93	3.05	3.11	3.28
Household Size	3.59	3.54	3.27	3.50
Census Blocks	883	967	2,370	2,463
Households	59,460	75,768	213,061	212,005

All variables are dummies for having each household attribute.

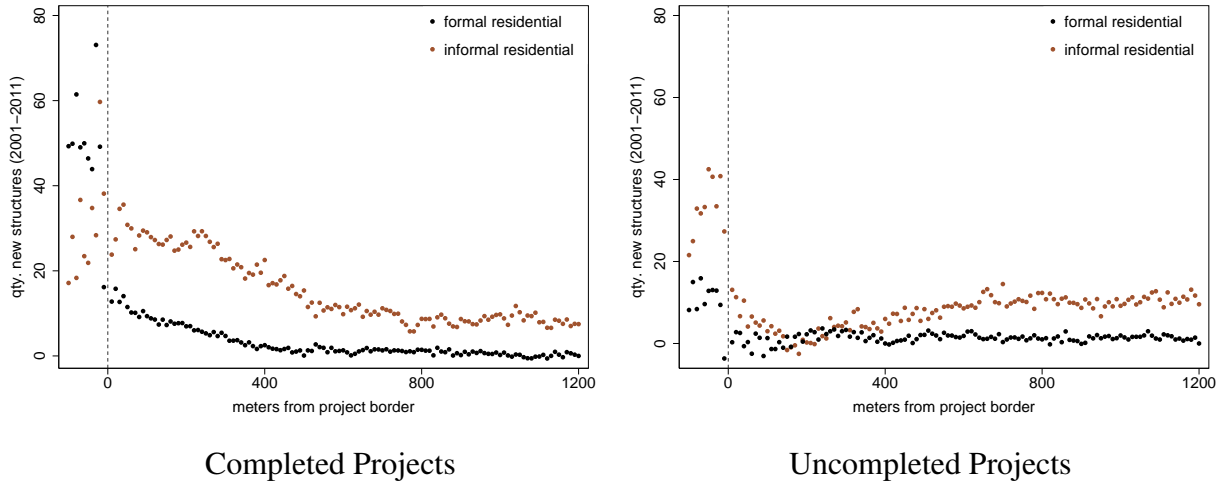
similar household sizes and similarly likely to own and reside in formal single houses. Lower service provision matches previous evidence that uncompleted projects are more often in areas with greater informal settlements. These differences between uncompleted and completed projects disappear for census areas located just outside of these project areas as shown in the last two columns. Taken together, these results suggest that despite differences in service provision within the project footprints at baseline, neighboring, spillover areas may serve as useful comparison groups.

4.1. Building Growth around Housing Projects

Given this descriptive evidence at baseline, we use the building structures survey to provide a first look at changes in development around completed and uncompleted projects. Figure 8 plots the average number of new structures between 2001 and 2011 in distance rings from the boundary of both completed and uncompleted project areas. Negative distances measure new structures within the project areas themselves.

Focusing on the left panel for completed projects, strong growth in formal residential structures at negative distances provides another measure of the direct impact of the housing projects. We

Figure 8: Growth in Structures between 2001 and 2011



also observe strong growth in informal residential structures within project areas. This finding is consistent with anecdotal evidence from fieldwork in South Africa pointing to the prevalence of backyard shacks where owners of project houses use the rest of their land plots to construct informal housing. At positive distances, the growth in formal structures immediately dissipates while the growth in informal structures remains substantial up to at least 400 meters from the project border. This sustained growth in nearby informal dwellings matches a common complaint of local housing authorities and NGOs who notice crowding-in of slum areas around project areas possibly to utilize the public services offered within the projects.

Results for uncompleted projects in the left panel indicate a similar pattern of informal structure growth in project areas, which is consistent with these areas remaining undeveloped by housing projects. Compared to completed projects, lower levels of formal structure growth occur within uncompleted project footprints before quickly dissipating outside of the projects. Unlike completed project areas, informal structure growth also decreases suddenly outside of the projects, providing evidence that completed housing projects may play an important role in generating nearby growth in informal settlements. The flat, smooth trends in structure growth past 500 meters for both completed and uncompleted projects support interpreting these regions as capturing the types of local development that we might expect in the absence of the housing projects.

5. Results: Price Spillovers

To identify the spillover effects of public housing on residential home prices in the formal market, we use a difference-in-differences approach comparing prices for areas close and far from project areas before and after project implementation. Our main empirical strategy takes the following specification:

$$P_{itp} = \alpha D_{tp} T_{ip} + \theta_1 D_{tp} + \theta_2 T_{ip} + X_i' \beta + \lambda_p + \eta_t + \varepsilon_{itp}$$

The outcome, P_{itp} , is measured in terms of the log-purchase price of property i sold at time t in the vicinity of project p . To capture changes in prices over time within project areas, D_{tp} is equal to one if date t is after the month of project implementation and zero otherwise. T_{ip} takes a value of one if property i is within 400m of the project boundary. The coefficient of interest, α captures the differential change in prices between near and far properties before and after project construction. X_i controls for a quadratic in lot size, which can affect prices over time. Additionally, λ_p includes a project fixed affect controlling for any fixed, unobserved drivers of house prices that vary between projects. Likewise, η_t controls for a calendar month (year \times month) fixed-effect to account for any factors such as shifts in aggregate housing demand that may be correlated with prices and the timing of housing projects.

Interpreting the coefficient, α , as the causal effect of housing projects on nearby home prices requires the following assumption:

$$E[\varepsilon_{itp} | X_i, T_{ip}, D_{tp}, \lambda_p, \eta_t] = 0$$

This assumption implies that there are no other factors occurring in the same time and place as the housing projects which may otherwise impact home prices. One possibility is that housing markets anticipate the construction of these projects so that transactions in the pre-period may be partially treated by the advent of a housing project. Anecdotal evidence suggests that completion dates for these projects are very uncertain due to the large coordination of stakeholders needed for each

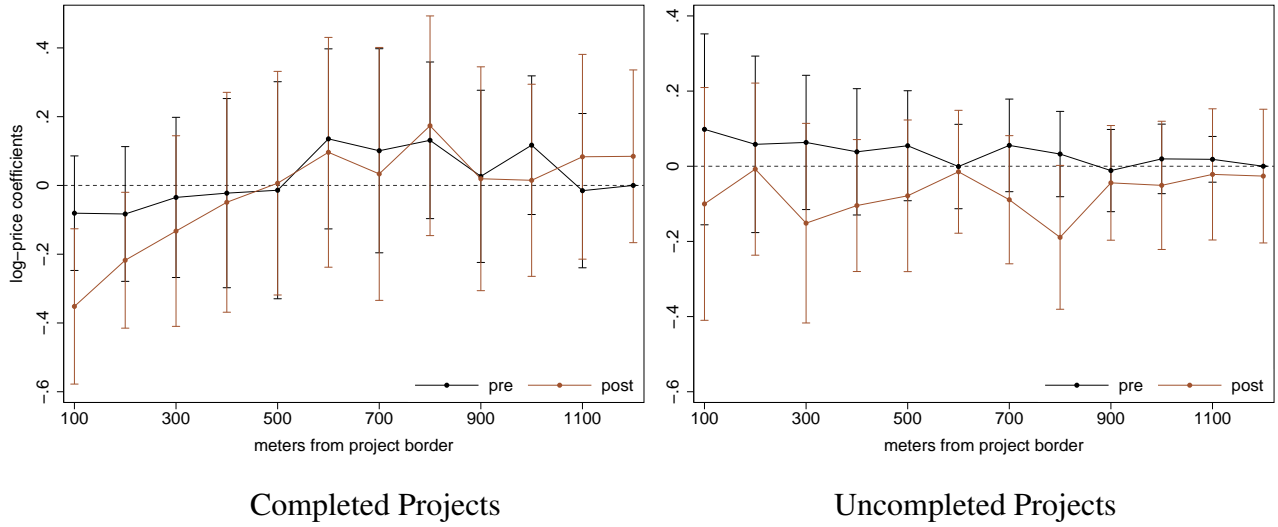
project, making it difficult to accurately anticipate implementation. Another concern would be that housing projects are accompanied by other social programs that would stimulate investments in neighborhoods near project areas. In order to isolate market anticipation or accompanying social programs from the actual impacts of housing projects, we estimate an identical model for planned but uncompleted projects to test the robustness of the results.

Similarly, in targeting housing projects to particular areas, governments may be responding to local trends in housing markets or economic conditions. To separate project impacts from secular market trends, we leverage the sudden roll-out of housing projects under the assumption that market trends are relatively smooth over space and time relative to the construction of housing projects. In order to non-parametrically assess identification in this way, we also estimate a more flexible model both in terms of distance to project, D_{tp} and time relative to project construction, T_{ip} . Specifically, we estimate separate treatment effects for each 100 meters of distance, $\sum_{d=100}^{1200} \alpha_d D_{tpd} T_{ip}$. We also allow effects to vary according to two-month intervals relative to construction, $\sum_{l=-36}^{36} \alpha_l D_{tp} T_{ipl}$. All regressions cluster standard errors at the project level in order to account for potentially correlation in prices within very localized housing markets.

Figure 9 separately plots coefficients for average home prices before construction, $\alpha_{d,pre}$ and after construction, $\alpha_{d,post}$ for each 100m distance ring from housing project boundaries. The left panel presents price gradients for completed projects. The pre-project gradient slopes slightly upwards consistent with housing projects being located in undeveloped land or preexisting informal settlements. After implementation, the price gradient slopes downwards sharply within 400 meters of the project area while remaining unchanged past this 400 meters threshold. This result is consistent both with the growth of informal settlements concentrated in the same spatial interval as observed in Figure 8.

Uncompleted projects in the right panel exhibit a flat price gradient before project implementation; since uncompleted projects are more often located in large slum neighborhoods, spillover areas may closely resemble these project areas. After the estimated date of project completion, the price gradient remains flat shifting slightly, but noisily downward at all distances from the project

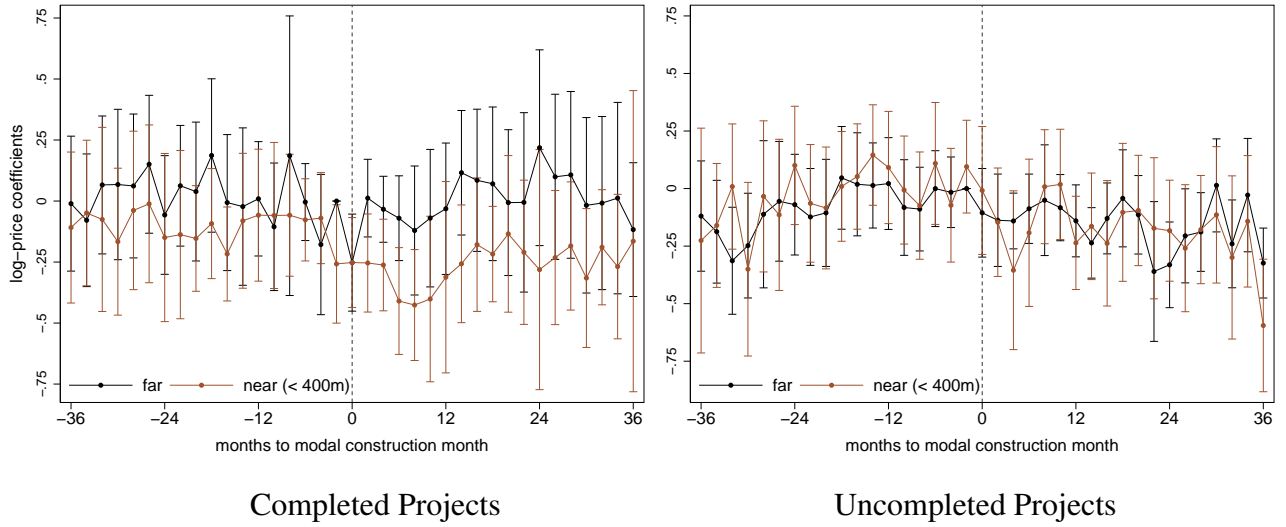
Figure 9: Price Estimates over Distance



boundary. These findings suggest little local spillovers from the uncompleted projects alongside a deterioration in the local housing market affecting all properties evenly. Additionally, these results provide suggestive evidence that exact location of housing projects does not appear to be strongly correlated with housing market trends at a very local level.

Complementing these distance gradients, Figure 10 plots coefficients for average price changes over time relative to the modal construction month for properties within and beyond 400 meters of a project boundary. Focusing on completed projects in the left panel, both areas far and near project boundaries follow a smooth, parallel evolution in prices over time. This parallel trend provides evidence against large shifts in policy or economic conditions that may be correlated with the onset of these housing programs. The price trajectories begin to deviate in the few months preceding the construction of the housing project. While further away properties maintain a steady price level following the housing project, nearby properties experience a sustained dip in prices that persists up to three years after the project is implemented. The sustained reduction in housing prices provides evidence for a lasting structural change in local housing market instead of a temporary resorting of households from nearby residences to project houses. The dip in prices begins in the couple months before the modal transaction year, which is consistent with nearby housing markets anticipating projects during their construction periods. One hypothesis is that given the

Figure 10: Price Estimates over Time



large observed decline in house prices, it would be rational for housing markets to anticipate construction, producing a declining pretrend in prices before completion date. The absence of an extended pretrend may speak to the unpredictability of the timing for these projects, especially given high coordination costs and frequent unplanned delays.

For uncompleted projects in the right panel, areas far and near project boundaries follow similar parallel trends to completed projects leading up to the expected completion date, but then remain indistinguishable for the entire duration after this date. This null result helps to exclude competing explanations that rely on the timing of housing projects such as targeting housing programs according to hyper-local housing market trends or pairing housing projects with other place-based policies.

Table 4 and Table 5 provide the difference-in-differences regression analogues to Figure 9 and Figure 10. In the first column of Table 4, we estimate a simple difference-in-differences specification and report the coefficient for properties within 400 meters of the project and 3 years of the start date. Since the outcome is measured in log-prices, the coefficient indicates that housing prices in this range decline by 23.8% as a result of the housing program. Including project fixed-effects in the second column attenuates this effect to 16.6% and reduces its statistical significance from 5% to 10% by accounting for fixed differences in the housing markets between clusters. The third

Table 4: Price Estimates for Completed Projects

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
3 yrs 0-400m	-0.238** (0.106)	-0.166* (0.0835)		
1st yr 0-400m			-0.147* (0.0754)	
2nd yr 0-400m			-0.180 (0.115)	
3rd yr 0-400m			-0.118 (0.0971)	
3 yrs 0-200m				-0.224** (0.0961)
3 yrs 200-400m				-0.0701 (0.0664)
Observations	28,701	28,701	28,701	28,701
R-squared	0.229	0.488	0.489	0.489
Project FE	NO	YES	YES	YES
Year-Month FE	YES	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All control for cubic in plot size. Standard errors are clustered at the project level.

column disaggregates the coefficients by years following project completion. Comparing coefficients, we only find a statistically significant coefficient for the first year of the project although effect sizes remain similar for all time intervals. With just over 150 projects, significant noise in the housing price data, and a rich set of fixed effects, the specification has difficulty precisely recovering a price estimate. In the spirit of Figure 9, the fourth column examines price effects at a finer spatial scale, finding an especially large and statistically significant price decline in the immediate 200 meters of a housing project which sharply attenuates past 200 meters.

Table 5 repeats the same differences-in-differences exercise in Table 4 but for uncompleted projects. Across all four columns, estimates are statistically insignificant and very small in magnitude compared to results in Table 4. In column three, we find one statistically significant coefficient at the 10% level for price effects in third year, which appears to be driven by an outlier in the 36th

Table 5: Price Estimates for Uncompleted Projects

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Price	(4) Log Price
3 yrs 0-400m	-0.0435 (0.0784)	-0.0664 (0.0597)		
1st yr 0-400m			-0.0687 (0.0662)	
2nd yr 0-400m			0.00256 (0.0729)	
3rd yr 0-400m			-0.159* (0.0841)	
3 yrs 0-200m				-0.0365 (0.0739)
3 yrs 200-400m				-0.0896 (0.0624)
Observations	24,562	24,562	24,562	24,562
R-squared	0.307	0.502	0.502	0.502
Project FE	NO	YES	YES	YES
Year-Month FE	YES	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All control for cubic in plot size. Standard errors are clustered at the project level.

month (according to the right panel in Figure 10). Taken together, these results provide additional evidence that price effects are primarily driven by the construction of housing projects rather than local trends in housing markets or economic conditions.

6. Results: Infrastructure and Demographic Changes

To provide additional evidence on how the demographics and infrastructure may be changing as a result of these projects, we use slightly modified differences-in-differences design where uncompleted project areas are directly treated as counterfactuals. The specification for analyzing the census data takes the following form:

$$Y_{hbt p} = \alpha_O D_{tp} T_{bp} O_{bp} + \alpha_S D_{tp} T_{bp} S_{bp} + \theta_1 D_{tp} O_{bp} + \theta_2 D_{tp} S_{bp} + \theta_3 T_{bp} S_{bp} + \theta_4 S_{bp} + \lambda_p + \varepsilon_{hbt p}$$

The outcome, Y_{btp} , includes demographic characteristics for household (or person) h living in census block b and project area p in year t . O_{bp} takes a value of one for census blocks with greater than 30% area overlap with housing projects (zero otherwise) while S_{bp} is the converse identifying “spillover” census blocks that have less than 30% overlap but whose centroids are still within a 1.2 km buffer of the housing project boundaries. To account for secular changes in outcomes between 2001 and 2011 affecting all census blocks, D_{tp} is equal to one if year t is equal to 2011 and zero otherwise. T_{ip} takes a value of one for census blocks near completed projects and a value of zero for census blocks near uncompleted projects. Controlling for differential time trends for overlap and spillover projects, $D_{tp}O_{bp}$ and $D_{tp}S_{bp}$ as well as separate means for both completed, $T_{bp}S_{bp}$, and uncompleted, S_{bp} , spillover areas leaves census blocks that overlap with uncompleted projects in the pre-period as the reference group. λ_p includes a project-level fixed effect to control for fixed differences in census characteristics between different projects. The coefficients of interest, α_O and α_S capture the differential change in outcomes between completed and uncompleted projects separately for overlapping and spillover census blocks.

To interpret the coefficients, α_O and α_S , as the causal effects of housing projects on both overlapping and spillover outcomes, we make the following assumption:

$$E[\varepsilon_{hitp} | T_{ip}, D_{bp}, S_{bp}, O_{bp}, \lambda_p] = 0$$

Intuitively, this assumption requires that there are no other factors occurring at the same time both within and around the completed housing projects that may drive changes in local demographics. By limiting the control group to only include uncompleted projects, we make the following parallel trends assumption: outcomes in areas with completed projects would have evolved in the same ways as outcomes near uncompleted projects in the absence of the housing projects. This design leverages the similarities in areas around completed and uncompleted projects instead of comparing project areas to the entire region of Gauteng. To the extent that locations of housing project may be targeted to local economic conditions, other areas in the province may be less likely to

satisfy this assumption.

At the same time, local economic and political conditions may determine which projects are completed and which are either canceled or delays. To the extent that these conditions may be correlated with changes in local demographics, the assumption of parallel trends would not hold in the analysis. Table 3 and Table 1 provide evidence that uncompleted project areas have more informal settlements and fewer services at baseline, consistent with policymakers placing less priority on finishing projects in these areas. At the same time, spillover areas appear very similar in Table 3 suggesting that differences in trends in these areas may be more plausibly associated with completion of housing projects.

Table 6 provides difference-in-differences estimates for a range of household level outcomes. The first two rows include the coefficients of interests, which measure the differential change in outcomes for completed and uncompleted projects. The first row reports effects for census blocks with greater than 30% overlap (Project) and the second row includes census blocks with less than 30% overlap (Spillover). Broadly, infrastructure and home quality measures (flush toilet, piped water, single house, and number of rooms) increase in overlapping areas while these same measures decrease in spillover areas. Many effects are statistically significant at the 5 to 1% level, and are large in magnitude. For example, the estimates predict a 21% increase in piped water for project areas with a corresponding decrease of 7% in spillover areas. While we observe large changes in infrastructure provision, estimates are small and insignificant for household size and home ownership.

The third and fourth rows capture secular changes in outcomes over time for uncompleted project areas. These measures indicate large overall increases in infrastructure and home quality measures in roughly similar magnitude for spillover and project areas.

Table 7 repeats the same exercise focusing on changes in demographic composition for households living in these areas. In the first row of column one, we find a statistically significant decrease in the share of high school (or greater) educated adults by around 3.6%. One explanation is that the housing policies are able to target recipients based on means and many of the original benefi-

Table 6: Census Household-level Estimates

VARIABLES	(1) Flush Toilet	(2) Piped Water Inside	(3) Electric Cooking	(4) Electric Lighting	(5) Single House	(6) Owns House	(7) No. Rooms	(8) Household Size
Project X Post X Complete	0.210** (0.0824)	0.202*** (0.0540)	0.0679 (0.0849)	-0.0482 (0.0998)	0.183*** (0.0451)	-0.0523 (0.0645)	0.286* (0.158)	0.0992 (0.0915)
Spillover X Post X Complete	-0.0723* (0.0411)	-0.0464 (0.0302)	-0.130*** (0.0449)	-0.0461 (0.0385)	-0.0957** (0.0374)	0.00820 (0.0501)	-0.102 (0.0923)	-0.00151 (0.0462)
Project X Post	0.115** (0.0448)	0.148*** (0.0298)	0.353*** (0.0760)	0.270*** (0.0822)	0.0535** (0.0238)	0.223*** (0.0498)	0.353*** (0.106)	-0.242*** (0.0771)
Spillover X Post	0.136*** (0.0313)	0.182*** (0.0205)	0.268*** (0.0347)	0.138*** (0.0285)	0.129*** (0.0312)	0.289*** (0.0245)	0.390*** (0.0670)	-0.236*** (0.0351)
Spillover X Complete	-0.205* (0.122)	-0.0732 (0.106)	-0.262*** (0.0885)	-0.450*** (0.0935)	-0.0540 (0.0731)	-0.217*** (0.0724)	-0.148 (0.222)	-0.202* (0.122)
Spillover	0.328*** (0.0873)	0.185** (0.0719)	0.373*** (0.0788)	0.429*** (0.0831)	0.147*** (0.0560)	0.128** (0.0624)	0.474*** (0.180)	0.150 (0.103)
Constant	0.498*** (0.0514)	0.217*** (0.0425)	0.389*** (0.0379)	0.537*** (0.0415)	0.439*** (0.0301)	0.468*** (0.0292)	2.746*** (0.0927)	3.334*** (0.0515)
Observations	1,544,285	1,544,285	1,544,285	1,544,285	1,479,342	1,496,636	1,459,677	1,532,866
R-squared	0.360	0.243	0.301	0.306	0.195	0.147	0.174	0.057
Project FE	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Standard errors are clustered at the project level.

Table 7: Census Person-level Demographic Estimates

VARIABLES	(1) Over HS Educ.	(2) HH Income	(3) Unemployed
Project X Post X Complete	-0.0360*** (0.0117)	-96.45 (395.0)	0.00604 (0.0222)
Spillover X Post X Complete	-0.00455 (0.0125)	22.31 (476.6)	0.0200 (0.0159)
Project X Post	0.0786*** (0.00946)	705.2*** (252.2)	-0.142*** (0.0110)
Spillover X Post	0.0298*** (0.00684)	2,231*** (288.1)	-0.134*** (0.00801)
Spillover X Complete	-0.0194 (0.0156)	-1,425*** (532.9)	0.0299 (0.0220)
Spillover	0.0236* (0.0138)	860.2* (455.4)	-0.0591*** (0.0159)
Constant	0.825*** (0.00600)	2,460*** (244.7)	0.505*** (0.00837)
Observations	2,191,516	972,949	1,145,850
R-squared	0.028	0.062	0.057
Project FE	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Standard errors are clustered at the project level.

ciaries continue to reside in project areas. For spillover areas and measures of household income and employment, we find no detectable effect of the housing program.

7. Discussion

Taken together, these census results suggest that ...

8. Conclusion

References

M. Busso, J. Gregory, and P. Kline. Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947, 2013.

- M. D. Cattaneo, S. Galiani, P. J. Gertler, S. Martinez, and R. Titiunik. Housing, health, and happiness. *American Economic Journal: Economic Policy*, 1(1):75–105, 2009.
- Department of Human Settlements. Breaking new ground: A comprehensive plan for the development of sustainable human settlements. Technical report, 2004. URL http://abahlali.org/files/Breaking%20new%20ground%20New_Housing_Plan_Cabinet_approved_version.pdf.
- Department of Human Settlements. Annual reports. Technical report, 2012,2015. URL <http://www.dhs.gov.za/content/annual-reports>.
- F. Devoto, E. Duflo, P. Dupas, W. Parienté, and V. Pons. Happiness on tap: piped water adoption in urban morocco. *American Economic Journal: Economic Policy*, 4(4):68–99, 2012.
- R. Diamond and T. McQuade. Who wants affordable housing in their backyard? an equilibrium analysis of low income property development. Technical report, National Bureau of Economic Research, 2016.
- P. Dlmini. R95m down the tubes as housing project picked apart brick by brick. 2017. URL <https://www.timeslive.co.za/news/south-africa/2017-09-12-r95m-down-the-tubes-as-housing-project-picked-apart-brick-by-brick/>.
- E. Durojaye, G. Mirugi-Mukundi, N. Munshi, and K. Tissington. Jumping the queue, waiting lists and other myths: Perceptions and practice around housing demand and allocation in south africa. Technical report, Socio-Economic Rights Institute of South Africa, Community Law Centre University of the Western Cape, 2013. URL http://www.seri-sa.org/images/Jumping_the_Queue_MainReport_Jul13.pdf.
- S. Franklin. Enabled to work: the impact of government housing on slum dwellers in south africa. 2016.
- S. Galiani, P. J. Gertler, R. Undurraga, R. Cooper, S. Martínez, and A. Ross. Shelter from the storm: Upgrading housing infrastructure in latin american slums. *Journal of Urban Economics*, 98:187–213, 2017.
- B. Marx, T. Stoker, and T. Suri. The economics of slums in the developing world. *Journal of Economic Perspectives*, 27(4):187–210, November 2013. doi: 10.1257/jep.27.4.187. URL <http://www.aeaweb.org/articles?id=10.1257/jep.27.4.187>.
- M. Mathoho. A case study of participation in the thinasonke ext 4 fedup pledge project. 2010. URL <http://www.planact.org.za/participation-and-development-from-the-perspective-of-the-poor-2008-2009/>.
- D. Matsena. Premier launches rdp housing project in north. 2018. URL <https://rekordnorth.co.za/121824/premier-launches-rdp-housing-project-in-north/>.
- A. Takeuchi, M. Cropper, and A. Bento. Measuring the welfare effects of slum improvement programs: The case of mumbai. *Journal of Urban Economics*, 64(1):65–84, 2008.
- K. Tissington. A resource guide to housing in south africa 1994-2010: Legislation, policy, programmes and practice. Technical report, Socio-Economic Rights Institute of South Africa, 2011. URL https://www.seri-sa.org/images/stories/SERI_Housing_Resource_Guide_Feb11.pdf.
- United Nations. Annex millennium development goals, targets and indicators, 2015: statistical tables. 2015. URL <http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2015/StatAnnex.pdf>.

Table 8: Assessing Name Matching between
Budget and Spatial Administrative Data

	Matched	Unmatched
Formal Density: 2001	230.5	171.5
Formal Density: 2011	814.1	444.0
Informal Density: 2001	1,055.6	1,401.0
Informal Density: 2011	1,613.2	2,147.0
Project House Density	125.0	66.0
Project Mode Year	2005	2005
Hectares	97.3	119.6
Observations	322	320

Density is measured in structures per km².

Appendix