\begin{document}

\begin{titlepage}

\title{{Public Housing Spillovers in a Developing Country}\thanks{We are grateful to Andrew Foster, Daniel Bjorkegren, Matthew Turner, Jesse Shapiro, Brian Knight, and John Friedman for their feedback and advice. We also thank Adelaide Steedley and the Centre for Affordable Housing and Finance in Africa as well as GeoTerraImage for generous research guidance and for providing data without which this project would not have been possible. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Federal Trade Commission or its Commissioners.}}

\author{\\[3em] Benjamin Bradlow\thanks{Dept. of Sociology, Brown University, Providence, RI E-mail: benjamin\textunderscore bradlow@brown.edu}\\

Stefano Polloni\thanks{Dept. of Economics, Brown University, Providence, RI E-mail: stefano\textunderscore polloni@brown.edu}\\

William Violette\thanks{Federal Trade Commission, Washington, DC. E-mail: william.j.violette@gmail.com} \\

\\

}

\vspace{30mm}

\date{\vspace{5mm}This Version: \today}

\maketitle

\begin{abstract}

We estimate the spillover effects of large public housing projects in South Africa using aerial measures of housing density, property transaction records, and census data. We compare (1) constructed projects versus planned but unconstructed projects, (2) before versus after scheduled construction, and (3) areas nearby versus far from project footprints. Projects increase the quantity and quality of formal housing within their footprints. Between 0 and 500 meters outside of their footprints, projects generate declines in informal housing quantity and formal housing prices especially in higher income neighborhoods. Estimates from a structural model of housing markets suggest that projects may generate net welfare losses on the order of 6.6 million USD per project.

\bigskip

\end{abstract}

\setcounter{page}{0}

\thispagestyle{empty}

\end{titlepage}

\pagebreak \newpage

\spacing{1.4}

\section{Introduction} \label{sec:introduction}

30\% of urban populations in developing countries live in slums often characterized by substandard dwellings, overcrowding, unsanitary conditions, and high crime rates \citep{mdg}. In response, many cities invest in replacing slums with large public housing projects. These projects aim not only to improve local housing conditions, but also to spur investment in surrounding neighborhoods.\footnote{For example, the South African Department of Housing promotes ``accelerating the delivery of housing as a key strategy for poverty alleviation,'' while also ``utilizing housing as an instrument for the development of sustainable human settlements, in support of spatial restructuring'' \citep{bng}.} By providing new formal houses and improved infrastructure, housing projects may generate positive neighborhood amenities. Projects may also have unintended consequences by attracting growth of low-quality, informal dwellings in the backyards of project houses \citep{Brueckner2018backyarding}. Dense backyard dwellings may recreate slum conditions and depress neighborhood property investment. The net spillover effects of housing projects remain an empirical question that has important implications for designing place-based policies to address slum growth.

This paper analyzes the spillover effects of large-scale public housing projects both within their footprints and within their surrounding neighborhoods. We focus on South Africa's National Housing Program, which is unique among developing countries for its massive scale having allocated over 3 million houses across hundreds of projects since 1994. This large volume of projects allows us to identify precise, spatial impacts, building on existing research which is largely limited to considering outcomes for direct project recipients.\footnote{\cite{cattaneo2009housing} estimate child health improvements from housing upgrades in Mexico, \cite{franklin2016enabled} finds greater women's employment for public housing recipients in South Africa, and \cite{galiani2017shelter} combine housing projects from throughout Latin America, finding large improvements in reported well-being of households.}

To measure these impacts, we compile detailed spatial data for the Johannesburg metro-area including aerial measures of formal and informal houses, census data, as well as deeds records of formal housing transactions. We find that housing projects more than double the density of formal housing within their footprints and improve access to basic services like piped water. While housing projects eradicate preexisting informal housing, they also crowd-in enough backyard houses to keep the total density of informal housing unchanged. Our estimates show that for neighboring areas between 0 and 500 meters from project boundaries, informal housing density declines and formal housing prices drop with the strongest reductions concentrated in relatively higher income neighborhoods.

In order to identify these effects, we combine three comparisons: First, we compare constructed projects versus planned but unconstructed projects to control for how governments strategically locate projects on undeveloped land with preexisting slums. Second, we compare projects before versus after scheduled construction to control for how neighborhoods with constructed projects often have more preexisting housing than neighborhoods with unconstructed projects. Third, we compare areas nearby versus far from a project footprint within the same neighborhood to control for housing market trends specific to each neighborhood. Taken together, these comparisons form a triple-differences estimator, which extends standard difference-in-differences approaches by allowing for the endogenous placement of projects within neighborhoods.\footnote{See \cite{rossi2010housing,hornbeck2017creative}, and \cite{diamond2016wants} for difference-in-differences estimators of place-based policies.}

We build an equilibrium model of housing markets that maps housing density effects into household preferences for living within and nearby housing projects. In the model, developers choose to build houses on land plots when the value to households from renting and enjoying local amenities exceeds the construction costs. Housing projects may affect both local construction costs as well as the quality of local amenities. By connecting the number of houses to welfare, this approach may have useful applications in developing country settings because price data are often scarce while housing quantity measures are increasingly available through remote sensing.\footnote{ \cite{marxthere} use satellite images to measure density and improvements in the quality of roofs in a Kenyan slum. \cite{Brueckner2018backyarding} use the same satellite data as this paper to study backyard houses in South Africa. \cite{donaldson2016view} review several research projects using building-level remote sensing data to map urban growth especially in developing countries. }

Structural estimates suggest that housing projects are likely to reduce aggregate welfare. By subsidizing home construction, projects generate welfare gains for formal housing markets within their footprints; however, these gains are overwhelmed by the costs of project implementation as well as negative spillover effects on nearby formal and informal housing markets. In total, simulation exercises estimate a net welfare loss of 6.6 million USD per project, which is economically large given implementation costs of 3.2 million USD per project.\footnote{Implementation costs are drawn from the National Treasury's Annual Budget Reports.} Since this framework assumes that developers are able to capture all rental surplus, the distributional consequences of these policies depend on which households are also developers. Without data on ownership over time especially for informal housing, we instead explore distributional consequences by examining effects separately by neighborhood income quartile.

We proceed by first providing background on public housing and the South African housing program in Section~\ref{section:background}. In Section~\ref{section:theory}, we build an equilibrium model of housing markets. Section~\ref{section:data} describes the data used to measure outcomes and details our approach to identifying housing projects. Motivated with descriptive evidence in section~\ref{section:descriptives}, Section~\ref{section:methodology} lays out the estimation strategy and identifying assumptions. Section~\ref{section:results} presents our main results, and Section~\ref{section:discussion} concludes.

\section{Public Housing Projects}\label{section:background}

European cities began investing in social housing programs in the early 20th century, and national governments in the West continued to grow these investments in the post-WWII era. After the wave of de-colonization in the 1950s, many developing nations also began public investments in housing, which were progressively phased out in order to secure international financing for national debt crises in the 1970s and 1980s. Beginning in the 1990s large middle-income countries took on subsidized housing programs that focused on channeling public funds to private developers to build full top-structure dwellings. Chile, China, South Africa, and Brazil are examples of this more recent approach.

\subsection{Public Housing in South Africa}

The housing projects studied in this paper were implemented as part of a large national housing scheme enacted in 1994. Though periodically revised and renamed,\footnote{Public housing in South Africa has been delivered under the {\it Reconstruction and Development Program} (RDP) starting in 1994 and subsequently by its successor, {\it Breaking New Grounds}, as of 2004.} the program has consistently sought to redress the economic and geographic legacy of apartheid by providing formal housing to low-income households.

This program allocates yearly funds to local housing authorities to construct and allocate 40m$^2$, single-story, two-room houses on individual plots with bulk services.\footnote{Local housing authorities include provincial housing departments and municipal housing departments for urban areas like Johannesburg \citep{dhsreports}.} Local authorities identify many large parcels of land that are able to each accommodate several hundred houses. To meet budgetary targets, authorities often identify inexpensive land located on the urban periphery \citep{beninterview,dhsreports}. Since project construction requires passing an environmental impact assessment, receiving zoning approval from local municipalities, and providing bulk service, authorities implement the most feasible subset of candidate projects each year.\footnote{Existing policy research characterizes South African housing as an opaque, top-down system where government authorities locate and allocate houses with minimal participation from local communities \citep{seriq}. Tissington [2012] documents how local community calls for housing improvements in Guateng were left unanswered for over a decade by local authorities. \cite{seriq} report how the Gauteng government refused to provide information about plans for housing subsidies and allocation to the Soweto Form, a local community based organization (p. 67).} In many cases, authorities combine housing projects with slum eradication goals by replacing preexisting informal settlements with new housing projects (\cite{hofmeyr2008risk}). Housing authorities subcontract with local housing developers for housing construction. According to government figures, upwards of 3 million housing units have been delivered nationally between 1994 and 2015.

Eligible households are required to sign up for an official waiting list maintained by the National Department of Housing.\footnote{According to the General Household Surveys, 2009 to 2013, the share of households reporting at least one member on the waiting list has remained stable at over 13\% from 2009 to 2013.} Eligibility requires South African citizenship, no previous property ownership, being married or having financial dependents, and having a monthly household income below R3,500 \citep{seriq}. Each project house is assigned to a beneficiary in a first-come, first-served basis according to the the be beneficiary's place in the waiting list for the province or municipality. When projects replace existing informal settlements, previous inhabitants have priority for receiving project houses. 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 within their first 7 years of ownership. %\footnote{The Gauteng Province has implemented their own waiting list since 2008 in order to exert greater control over the allocation process.}

In practice, these guidelines are loosely followed. Recent reports point to cases of corruption in the allocation of houses \citep{seriq}.\footnote{Research suggests that beneficiaries are often selected over the course of project construction and sometimes even after construction has finished \citep{seriq}.} Also, only 82\% of project houses are reported as being still occupied by their original beneficiaries within five years of construction.\footnote{This figure is calculated from the General Household Surveys from 2009 to 2013. Anecdotal evidence suggests that project managers are aware of active secondary markets but have difficulty policing these transactions \citep{resale}.}

\section{Modeling the Housing Market}\label{section:theory}

We develop a simple model of residential housing markets to capture the welfare effects of local housing projects within their neighborhoods. This model infers welfare changes not only through changes in rents, but also through changes in the quantities of housing supplied.

Each time period, developers choose how many formal and informal houses to construct and maintain on each plot of land as well as the level of rent for each house. Each house can serve up to one household, and houses have identical quality. Then households choose whether to rent a house receiving utility given by

\begin{align\*}

U &= \delta\_{hlt} - \lambda\_{h}(k\_{hjt}) - \theta R\_{hjt} + \epsilon\_{hjt}

\end{align\*}

\noindent where $h$ indexes housing sector (either formal or informal), $l$ indexes location, $j$ indexes the specific land plot within location $l$, and $t$ indexes the time period. $\delta\_{hlt}$ indicates the amenity value of location, $l$ to housing sector, $h$ in time period, $t$. $k\_{hjt}$ measures the number of houses of a sector on the land plot. $\lambda\_{h}(k\_{hjt})$ captures the congestion disutility from having multiple houses of the same sector on the same plot and is assumed to be increasing and convex in the number of houses per plot. $\lambda\_{h}(k\_{hjt})$ is also assumed to be independent of the number of houses from another sector on the same plot. $R\_{hjt}$ is the rent charged for each house on plot $j$ and $\theta$ measures the marginal disutility of rent. $\epsilon\_{hjt}$ is the plot and time-specific amenity shock. The distribution of $\epsilon\_{hjt}$ may be correlated across housing sectors to capture whether land plots that are suitable for formal housing may also be suitable for informal housing. Households can also choose to live outside the city and receive reservation value, $\overline{U}$.

This approach assumes that households have identical preferences across houses and locations, which implies that housing demand is perfectly elastic. This assumption is consistent with households having low moving costs within the city as well as no idiosyncratic preferences for particular neighborhoods. These assumptions may be reasonable in a context like South Africa where high levels of unemployment mean that different households face similar strong incentives to locate near employment centers.

Given household preferences, developers maximize profits for each housing sector by setting rents just high enough to ensure that households are indifferent between renting each house on plot $j$ and their reservation utility $\overline{U}$. Profit-maximizing rents as a function of the number of houses on a plot is given by

\begin{align}

\label{eq:rent}

R\_{hjt}^{\*}(k\_{hjt}) &= \frac{ \delta\_{hlt} - \lambda\_{h}(k\_{hjt}) + \epsilon\_{hjt} - \overline{U}}{\theta}

\end{align}

From a distributional perspective, this approach assumes that developers are able to extract all social surplus by setting rents relative to household reservation utility. Although some households may also be developers, welfare for non-developer households remains unaffected by shifts in local amenities because any shifts are immediately capitalized into rents. Without detailed measures of home ownership, this approach is unable to disentangle the distributional consequences of public housing policies across the population of households.

Given profit-maximizing rents, developers then choose how many houses to build on each plot by maximizing total profit per plot given by

\begin{align\*}

max\_{\,k\_{hjt}} \,\,\,\,\,\, k\_{hjt} \, \Big[ \, R\_{hjt}^{\*}(k\_{hjt}) - C\_{hjt} \, \Big ]

\end{align\*}

\noindent where the $C\_{hjt}$ measures the cost of constructing and maintaining each house, which may vary based on location.

%and may depend on plot-specific features such as land slope, $C\_{j} = c + \psi S\_{j}$ where land slope is given by $S\_{j}$.

When $\lambda(k\_{hjt})$ is convex, the profit-maximizing number of houses in each sector, $k^{\*}\_{hjt}$ can be expressed as a function of the random amenity shock $\epsilon\_{hjt}$ given by

\begin{align}

\label{eq:housingdemand}

\begin{split}

k\_{hjt}^{\*} &=

\begin{cases}

0 &\text{ if }\,\, \epsilon\_{hjt} + V\_{hjt} \leq \Lambda\_{h}(1) \\

% 1 &\text{ if }\,\, H + \lambda(1) < \epsilon\_j \leq H + 2\lambda(2) - \lambda(1) \\

k &\text{ if }\,\, \Lambda\_{h}(k) < \epsilon\_{hjt} + V\_{hjt} \leq \Lambda\_{h}(k+1)

\end{cases} \\

\text{ where }& \\

V\_{hjt} & = \delta\_{hlt} - \theta C\_{hjt} \\

\Lambda\_{h}(k) & = k\lambda\_{h}(k) - (k-1)\lambda\_{h}(k-1) + \overline{U} \\

\lambda(0) & = 0 \,\, \text{ and } \lambda(.) \text{ is increasing and convex }

\end{split}

\end{align}

The profit-maximizing number of plots with houses and the number of houses per plot by sector are increasing in amenity values, $\delta\_{hlt}$, decreasing in congestion disutility, $\Lambda\_{h}(k)$, and decreasing in construction costs, $C\_{hjt}$. According to this framework, observed increases in the number of plots with houses and/or the number of houses per plot indicate an improvement in welfare holding constant reservation utility. The variance of amenity shocks, $\epsilon\_{hjt}$ may be interpreted as a measure of the sector-specific housing market elasticity.

Given optimal rents in equation (\ref{eq:rent}) and optimal houses in equation (\ref{eq:housingdemand}), let the price of each plot equal profit for each plot as follows

\begin{align}

\label{eq:profits}

\begin{split}

P\_{hjt}^{\*} &=

k\_{hjt}^{\*} \Big[ \frac{ \delta\_{hlt} - \lambda\_{h}(k\_{hjt}^{\*}) + \epsilon\_{hjt} - \overline{U}}{\theta} - C\_{hjt}\Big]

\end{split}

\end{align}

% \begin{cases}

% 0 &\text{ if }\,\, \epsilon\_{hjt} + V\_{hjt} \leq \Lambda\_{h}(1) \\

% % 1 &\text{ if }\,\, H + \lambda(1) < \epsilon\_j \leq H + 2\lambda(2) - \lambda(1) \\

% k \Big[ \frac{ \delta\_{hlt} - \lambda\_{h}(k) + \epsilon\_{hjt} - \overline{U}}{\theta} - C\_{hjt}\Big] &\text{ if }\,\, \Lambda\_{h}(k) < \epsilon\_{hjt} + V\_{hjt} \leq \Lambda\_{h}(k+1)

% \end{cases}

Like the profit-maximizing number of houses per plot, plot prices are increasing with local amenities, decreasing in congestion disutility, and decreasing in construction costs. Changes in plot prices fully capture changes in social welfare because equilibrium rents ensure that household utility is always equal to reservation utility, leaving developers to enjoy all additional surplus.

% By contrast, increases in observed average rents have ambiguous welfare implications depending on whether these changes are driven by amenity improvements or construction cost increases. Equation (\ref{eq:rent}) indicates that optimal rents rise and fall monotonically with amenity quality, $\delta\_{lt}$. Construction costs affect optimal rents indirectly by affecting how many houses developers locate on plots. Higher construction costs mean that developers . Therefore, rents

% Predicts FEWER SALES?!?!?!?!??!?!! SHOW THE SALES REPORTS!!!!!!! RENTS VERSUS PRICES?! TRICKY!!!!!!

% Increases in construction costs lead to fewer developed plots, and .

% Optimal prices, ... as the change.....

% Price changes capture the net result of (1) the direct effects of local amenity quality as well as (2) the indirect effects of shifts in the extent of development and the quality of lots that are developed. Optimal rents given by equation (\ref{eq:rent}) indicate that rents respond directly to amenities, $\delta\_{lt}$ and indirectly to changes in the intensity of development, $k\_{jt}$, which affects prices through both the amount of crowding disutility, $\lambda(k\_{jt})$ as well as the unobserved quality of plots that are developed, $\epsilon\_{jt}$. In this framework, decreases in prices can be generated not only by decreases in amenity values, but also by decreases in construction costs as lower construction costs induce development on low quality plots as well as greater congestion

\subsection{Modeling Public Housing Projects}

Public housing projects enter the model by shifting housing costs and amenity values both within as well as nearby project footprints. Let location index $l$ indicate whether plots fall within project footprints, $w$ or nearby spillover areas, $s$. Similarly, let housing sector index $h$ indicate whether amenities/costs apply to formal, $f$ or informal, $i$ housing sectors.

Housing projects allocate subsidies to local developers to construct houses within project footprints. These subsidies directly lower formal housing construction costs in project footprints, $C\_{fwt}$. These houses are then allocated to beneficiary households who can choose to live in these houses or rent them at efficient prices. Housing projects also remove preexisting informal housing and rezone land for formal housing development, leaving limited land for new informal developments. These practices may increase costs for informal housing, $C\_{iwt}$ in project footprints.

Housing projects may also affect local amenity values for formal housing, $\delta\_{fwt}$ and informal housing, $\delta\_{iwt}$ within project footprints. Housing projects may boost local amenities by increasing access to piped water, electricity, and sewerage services while providing better roads and community centers.

Housing projects may also have effects on amenities for neighboring areas just outside of project footprints. On one hand, improvements in infrastructure in project footprints as well as greater densities of formal housing may provide positive amenities for surrounding neighborhoods, $\delta\_{hst}$. On the other hand, greater densities of backyard housing within projects as well as shifts toward lower income demographics in project areas may be perceived as negative amenities by surrounding housing markets, $\delta\_{hst}$.

Collapsing these effects into amenity terms, $\delta\_{hlt}$ provides a reduced-form method of capturing the spillover effects of housing projects; however, this approach implies a series of restrictions on behavior. First, the amenity term, $\delta\_{hlt}$ enters household utility linearly and independent of the number of houses per plot, which nests the assumption that the all households enjoy local amenities equally and independently of the density of neighboring housing. This assumption does not allow for congestion costs from overusing public services.

Second, this approach imposes a limited structure for amenity externalities to extend across space: amenities in project areas are allowed to affect amenities for spillover areas while amenities in all other areas are assumed to be spatially independent of each other. This assumption is consistent with housing projects generating large, first-order changes that precipitate across neighborhoods while each individual's development decisions produce negligible externalities across neighborhoods. Alternative approaches that explicitly model housing externalities across space would likely suffer from multiple equilibria as well as challenges for identification in this setting.\footnote{\cite{rossi2010housing} develop a model of spatial housing externalities over a single dimension, ``housing services'' where identification draws from the functional form of the housing externalities. Our setting would require households to consider externalities across housing quantity as well as sectors.}

\section{Data}\label{section:data}

% Our analysis focuses on the South African province of Gauteng, both the smallest geographically and most populated province in the country. Gauteng's boundaries roughly correspond to the greater metropolitan areas of Johannesburg and Pretoria.

Understanding the local development impacts of public housing requires (1) a precise measure of the location, timing, and size of housing projects, and (2) outcomes measured at high spatial resolutions.

We locate projects using a map of housing projects obtained from the Gauteng City Regional Observatory, a research unit composed of the Gauteng Provincial Government and two Johanesburg universities.\footnote{\href{url}{www.gcro.ac.za/}} The map describes \input{figures/total\_gcro}housing projects as of 2008, including notes on their completion status. We classify \input{figures/total\_con}projects with descriptions including ``current,'' ``under implementation,'' or ``complete'' as constructed projects. We then classify \input{figures/total\_uncon}projects with descriptions including ``proposed,'' ``under planning,'' ``future,'' ``investigating,'' and ``uncertain'' as planned but unconstructed projects. The remaining \input{figures/total\_other}projects either do not include descriptions or their descriptions are difficult to classify as either constructed or unconstructed. Appendix Table~\ref{table:projectdescriptions} includes tabulations of the full list of descriptions.

%Unclassified projects are more likely to be located in the northern region of the Gauteng province according to the map in Figure~\ref{figure:map}, which suggests that housing authorities in northern areas like Pretoria area may have followed different record keeping practices compared to those in the Johannesburg metro area.

% \input{figures/informal\_other}of the unclassified projects are described as ``informal,'' which may correspond to government efforts to identify informal settlements rather than.

% Additional information for some projects allows us to examine the impacts of different types of housing projects. We identify \input{figures/total\_insitu}in-situ upgrading projects as those with descriptions containing the phrase ``Essential Services,'' which are linked to an aggressive demolition and rebuilding campaign launched by the Gauteng Province between 2000 and 2008 (\cite{hofmeyr2008risk}). Likewise, we identify \input{figures/total\_greenfield}greenfield projects as those with descriptions containing the phrase ``Mixed Housing Developments,'' which were designed to modernize greenfield projects by combining project houses with investments in local services (\cite{greenfield}).

For constructed projects, we assign a completion date to each project according to the date when recipients received deeds to their project houses. To recover this date, we overlay project locations onto deeds data from the South African National Deeds Office. The deeds cover the universe of housing transactions from 2001 to 2011 in {\it affordable areas}, which are defined as census enumeration areas with 2010 mean house prices below R500,000.\footnote{These data were provided by the Affordable Land and Housing Data Centre, which tracks affordable housing markets.} Deeds include price, GPS location, plot size, buyer name, and seller name. Although the data do not identify whether deeds belong to housing projects, we are able to infer project status according to whether the seller name includes a government or municipality when first transacted. This project definition also excludes deeds flagged as large buildings used for commercial purposes (less than 2\% of transactions) as well as purchase prices more than R50,000 above the yearly nominal subsidy values (less than 4\% of remaining transactions), resulting in a sample of over 48,000 deeds. Appendix Figure \ref{figure:transactionhist} plots the histogram of sale prices according to this project definition, finding substantial bunching around subsidy values for project deeds and a smooth distribution of prices for non-project deeds. We assign a completion date according to the modal year and month for deeds within each project. Within projects, most government-sponsored properties are transacted in the same month.\footnote{Appendix \ref{appendix:histfreq} plots the distribution of subsidized and non-subsidized transactions around the modal transaction month.}

For planned but unconstructed projects, we approximate expected completion dates by matching project names to National Treasury budget reports, which list expected project completion dates. We digitized data for \input{figures/unique\_budget\_report\_names}projects from budget reports spanning 2004 to 2009, which detail the name, start date, expected completion date, and cost of each housing project. We then use a string-matching algorithm to link project names from the budget reports to the administrative maps. Out of the 642 total project names from the administrative map, we are able to match 322 projects, including 126 constructed projects and 73 planned but unconstructed projects. We provide further details about the digitization and string-matching algorithm in Appendix~\ref{appendix:stringmatch}. Because of the low rate of matches, we use expected completion dates only for a subsection of the empirical analysis that focuses on estimating the immediate price impacts of housing projects. The majority of the analysis compares outcomes in 2001 and 2011 for constructed and unconstructed projects under the assumption that these projects either were constructed or would have been constructed between 2001 and 2011.

%%% MORE DISCUSSION OF THIS LIMITATION?

\begin{figure}[t!]

\centering

\caption{Housing Project Map}\label{figure:map}

% \tcbox[colback=white,boxrule=.35mm,boxsep=0mm]{\includegraphics[scale=.098,trim={.9cm .4cm .9cm .4cm}]{figures/explanmap.jpg}}

\tcbox[colback=white,boxrule=.35mm,boxsep=0mm]{\includegraphics[scale=.6,trim={.9cm .4cm .9cm .4cm}]{figures/newprojectmap.pdf}}

\end{figure}

Figure \ref{figure:map} maps the sample of constructed and unconstructed projects as polygons within the Gauteng province. Importantly, Figure \ref{figure:map} also shows that the affordable areas -- the coverage area for our deeds data -- contain nearly every project boundary. While constructed and unconstructed projects are often adjacent to each other, possibly indicating cases where authorities were unable to complete different phases of planned projects, there are also many examples of isolated projects of both types. Projects are generally located at relatively great distances from central business districts (CBDs), suggesting that vacant or inexpensive land plots are targeted by housing authorities. Despite their distance from the CBDs, housing projects are often next to arterial highways, potentially easing commuting costs for recipients.

% We note that our approach is not without limitations, and may introduce measurement error insofar as we are misattributing deeds to housing projects (false-positives), or assuming a project is unconstructed when it was actually constructed (false-negatives).

% To provide some validation for our classification, we tabulate in Table~\ref{table:projectdescriptions} project descriptions from the 2008 administrative policy maps according to whether projects are classified as constructed or unconstructed by 2011. As of 2008, we find that constructed projects are more likely to be classified as ``completed" or ``under implementation", while unconstructed projects are more likely to fall into ``proposed'' or ``planning'' categories. Among constructed projects, we find 5 ``proposed'' and 8 ``planning'' indicators, which could imply either (1) these projects are false-positives or (2) while planned at some point prior to 2008, these projects were eventually completed by 2011.\footnote{We are unable to assess the frequency at which these descriptions are updated in the data.} We also identify a single ``complete'' project as unconstructed, likely because its houses do not appear in our deeds records. Figure \ref{fig:forchange} in section \ref{section:descriptives} further validates our sample, showing how formal housing increases disproportionately in projects classified as constructed.

% \subsection{Nearby housing transactions}

To examine housing price spillovers nearby projects, we include properties that are located outside of the project areas and are not sold by government housing agencies or large developers identified as sellers with more than 30 transactions. Excluding these sellers limits the external validity of our results only to small developers and individual homeowners. We focus on properties located within 1.5 kilometers of constructed and unconstructed housing projects, forming a sample of over 67,000 transactions. We exclude the top 1\% of prices as well as prices below 2,500 Rand, which are likely to be composed of mismeasured prices or titles exchanged between family members.

% \subsection{Census Data}

We measure impacts on dwelling and household characteristics with the full 1996, 2001, and 2011 Censuses of Population and Housing. We analyze average outcomes at the smallest available census geography, which is referred to as {\it small-areas} by the South African census and which we refer to as census blocks for the purposes of this research. The province of Gauteng is divided between approximately 17,000 blocks in 1996, 11,000 blocks in 2001, and 17,000 blocks in 2011, which are not constant over time periods. On average, each block contains 170 households.

% Though we observe responses from every surveyed household in both census waves, the data does not allow for linking households across time periods.

% We analyze household-level responses describing the quality of their living quarters.

% Our outcomes are mainly binary indicators and pertain to the household's access to services (flush toilets, water tap, electricity access), housing durability, and tenure arrangements.

% We identify households

\begin{figure\*}[t!]

\centering

\caption[ Building Data: Example Housing Project]

{\small Building Data: Example Housing Project }

\vspace{2mm}

\includegraphics[width=\textwidth,trim={0cm 0cm 0cm 0cm}]{figures/newbblumap1.png}

\label{fig:bblumaps}

\end{figure\*}

% \subsection{Building Based Land Use}

% \label{section:data:bblu}

To measure housing growth, we analyze hand-coded building data derived from high-resolution aerial and satellite imagery. We obtain these data from GeoTerraImage (Pty) ltd., a local remote-sensing specialist.\footnote{\href{http://www.geoterraimage.com/}{\tt http://www.geoterraimage.com/}} The data differentiate structures across 30 categories, including formal and informal residential dwellings. Informal housing structures are easily identified from their temporary nature, often made of materials such as recycled wood and corrugated metal. In contrast, formal housing structures are permanent, generally made out of brick, and may have a pitched or a flat roof with tiles, zinc panels, or other materials. Importantly for our analysis, the data encodes backyard shacks as informal structures, but distinguishes between backyard and other types of informal buildings. We track changes in residential development by using two available survey waves in Gauteng: 2001 and 2012. As a validation exercise, we compute aggregated building counts at the census block level, and check the resulting figures against the number of households reporting to live in formal/informal housing in the census. The two data sources paint a consistent picture of housing in Gauteng, with high correlations ($\geq$ 0.85) between formal and informal building quantities in both available comparison years.\footnote{We correlate the 2001 and 2012 building surveys with the 2001 and 2011 censuses, respectively.} To measure building density, we transform this data into 25$\times$25m grid cell aggregates, creating five measures of housing density: (1) total residential structures split between (2) formal residential structures and (3) informal residential structures, which can be further decomposed into (4) backyard informal and (5) non-backyard informal structures. We exclude grid cells where development is infeasible because these cells intersect with rivers, lakes, recreational locations (tennis courts, pools, basketball courts, etc.), and mining excavation areas (10\% of total grid cells). In Figure \ref{fig:bblumaps}, we provide an example of the raw data and a depiction of our griding procedure, using the 2012 data wave.

% These grid cells will be the main units of observations for our housing density regressions, described in section \ref{section:bbluestimates}.

We also construct a measure of construction costs as a function of land slope. First, recent research by the Center for Affordable Housing and Finance (CAHF) suggests that construction costs account for the majority of the value of an average property, which is consistent with a competitive market in property development (\cite{cahfcosts}). The average property value in our data is R231,000. To benchmark this value, a 2016 study by AECOM engineering estimates construction costs to be around R3,500 per square meter of floor space for the lower end of the housing market (\cite{aecom}). Dividing construction costs by costs per square meter implies that the floor space of the average house in the low end of the housing market is around 66 m2, which is comparable to the 40 m2 of floor space available for each project house. The CAHF also estimates that the average construction cost for a fully serviced, 46 m2 house (with a 9 m2 balcony on a 120 m2 plot) in Pretoria is R336,000, which is also consistent with our estimate given that our data includes the bottom 20\% of the housing market.

Second, South Africa's National construction guidelines indicate that constructing homes on slopes with 6\% to 12\% gradients increases infrastructure costs by 25\% and building costs by 5\%, and constructing on slopes with greater than 12\% gradients increases infrastructure costs by 50\% and building costs by 15\% (\cite{redbook}). Recent research from the CAHF in Pretoria estimates that out of total construction costs, around 62\% is attributable to building costs while around 12\% is attribute to infrastructure costs (\cite{cahfcosts}). To measure land slope, we use 200m elevation contour lines provided by the National Geo-Spatial Information component of the Department of Rural Development and Land Reform of South Africa. We calculate a proxy for slope by finding the highest and lowest points in a 500$\text{m}^{2}$ square and dividing their elevation difference by the euclidean distance between them. The dataset includes 9,770 squares of which 93\% have slopes of less than 6\%, 5\% have slopes between 6\% and 12\%, and the remaining 2\% have slopes greater than 12\%.

\section{Descriptive Statistics}\label{section:descriptives}

% As mentioned in section \ref{section:data},

% Validate measures :

% \begin{itemize}

% \item delivered houses are way higher in constructed, but there's a little room for slippage

% \end{itemize}

% Balance/identification discussion :

% \begin{itemize}

% \item everything else is really balanced, which is good for identification

% \end{itemize}

% Spatial targeting

% Unconstructed and constructed Differences

% Which types of neighborhoods receive unconstructed and constructed projects and where are projects located within neighborhoods

% Note! :

% \begin{itemize}

% \item first, hh 2001 subplaces do not intersect perfectly with project area, so average incomes include project areas in the pre-period, which actually might be fine

% \item second, quartiles defined for prices may be different than quartiles defined for

% \item it is possible to do the nearby analysis conditioning on nearby incomes... maybe its fine as is....

% \end{itemize}

We provide descriptive evidence to validate our measures of constructed and unconstructed projects. We then examine differences between neighborhoods with constructed projects and neighborhoods with unconstructed projects. We also examine where projects are located within neighborhoods. Finally, we document how project areas and nearby areas evolve differentially over time.

Table~\ref{table:projectdescriptives} compares average attributes for constructed and unconstructed projects. The first row counts the number of project houses recorded in the deeds data within constructed and unconstructed projects, providing an independent measure of which projects were constructed. Projects categorized as constructed receive a large influx of 374 deeds project houses between 2001 and 2011, indicating that many of these projects were successfully completed in this interval. By contrast, projects categorized as unconstructed receive just 11 deeds project houses on average, which suggests that completed projects are unlikely to be misclassified as unconstructed projects.

Housing prices within 1 km of projects are lower than the average price of 252,000 Rand for areas over 1 km from projects, indicating that projects are located in neighborhoods with less expensive land. With average land areas exceeding one $\text{km}^{2}$, housing projects represent a significant change in local neighborhoods. Size, timing, location, and nearby house prices are broadly similar between unconstructed and constructed projects.

% All other attributes of housing

% the location, size, and surrounding housing stock of the \input{figures/total\_con\_uncon}constructed and unconstructed projects. shows that constructed projects deliver an average of 374 houses as measured by the deeds data. By contrast, projects flagged as unconstructed only produce 11 houses on average, suggesting that

% validated also by the descriptive changes in housing projects...

% Constructed projects are also substantially larger in area than unconstructed projects. these discrepancies may be due to many unconstructed projects representing smaller planned extensions to previously implemented projects.

% It may also reflect the fact that larger land plots are first targeted by housing authorities to maximize delivery.

% Though unconstructed projects are located in slightly more remote locations, both types of projects impose long commutes on residents, as measured by the average distance to central business districts.

% These distances are consistent with housing authorities targeting inexpensive, vacant land for these projects. Finally, prices for non-subsidized housing in the vicinity of project boundaries are slightly lower for unconstructed projects, perhaps driven in part by their more distant locations.

% Table~\ref{table:projectdescriptives} further divides projects into city and suburb according to whether their centroids are above or below median distance from a CBD respectively. City projects are smaller, but receive more houses than suburb projects. We also find higher housing prices in formal markets surrounding project areas consistent with these areas providing better access to employment opportunities. Similar patterns emerge between constructed and unconstructed projects when we disaggregate by distance to CBD.

\vspace{0mm}

\begin{table}[h!]

\centering

\caption{Housing Project Areas Description}\label{table:projectdescriptives}

\vspace{0mm}

\begin{tabular}{l\*{1}{cc}}

\toprule

&Constructed & Unconstructed \\

\midrule

\input{figures/descriptives\_tablet\_4}

\bottomrule

%\multicolumn{3}{l}{\scriptsize Const. refers to constructed projects and unconst. refers to unconstructed projects.}\\[-.5em]

%\multicolumn{3}{l}{\scriptsize $^\*$Calculated from {\it expected} completion dates using Gauteng National Treasury budget reports.}\\[-.5em]

\multicolumn{3}{l}{\scriptsize $^\dagger$ The USD averaged around 7.70 Rands during the 2001-2011 period.}\\[-.5em]

\multicolumn{3}{l}{\scriptsize $^\ddagger$Measured as the average minimum distance with respect to Johannesburg and Pretoria CBDs. } \\[-.5em]

\end{tabular}

\end{table}

% Next, we turn to household census data to assess how both types of projects differ in terms of dwelling characteristics prior to project construction.

% \footnote{Our census geography selection method is depicted in more detail in appendix \ref{appendix:bufferdesign}.}

To further assess where projects are located, Table~\ref{table:projectdescriptivescensus} uses pooled 1996 and 2001 census data to examine preexisting housing quality in project footprints before construction. Since census blocks are often larger than housing projects, Table~\ref{table:projectdescriptivescensus} considers census blocks as pertaining to a housing project when over 50\% of the land area of the block overlaps with a project footprint. Table~\ref{table:projectdescriptivescensus} compares means for household indicators aggregated at the block level based on whether the census boundaries overlap with constructed (first column) or planned but unconstructed projects (second column). As a reference, the third column provides averages for all census blocks in the Gauteng province.

\begin{table}[h!]

\centering

\caption{Mean Housing Characteristics from the 1996 and 2001 Censuses}\label{table:projectdescriptivescensus}

\vspace{-2mm}

\begin{tabular}{l\*{1}{ccc}}

\toprule

& Constructed & Unconstructed & All blocks \\

\midrule

\input{figures/census\_at\_baseline\_AGG\_4}

\bottomrule

\multicolumn{4}{l}{\scriptsize ``Constructed'' and ``Unconstructed'' include census blocks with over 50\% } \\ [-.5em]

\multicolumn{4}{l}{\scriptsize area overlap with constructed and unconstructed projects respectively. } \\ [-.5em]

\multicolumn{4}{l}{\scriptsize ``All'' includes all blocks. Means are weighted by land area.}

\end{tabular}

\end{table}

Table \ref{table:projectdescriptivescensus} finds that before project construction, preexisting houses in project footprints are smaller and less formal with worse access to services compared to other houses in the province. This finding is consistent with efforts by the Gauteng government to locate housing projects in inexpensive, poorer neighborhoods to both upgrade these neighborhoods as well as save costs. Comparing preexisting houses in constructed and unconstructed projects, we do not observe systematic differences in housing attributes. While constructed project footprints are more likely to have flush toilets, electricity, and large household sizes, households in unconstructed project footprints report having better access to formal houses and piped water in their homes. With a small number of blocks, average characteristics are not statistically different between constructed and unconstructed project footprints except for the formal housing measure.\footnote{Statistical significance is calculated at 5\% confidence level weighted by land area.}

\begin{figure\*}[h!]

\centering

\caption[ 2001 Housing Densities in Constructed and Unconstructed Projects]

{\small 2001 Housing Densities in Constructed and Unconstructed Projects}

%\vspace{2mm}

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_for\_pre\_means\_4\_spk.pdf}

\caption[Network2]%

{{\small Formal density}}

\label{fig:prefor\_raw}

\end{subfigure}

\hfill

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_inf\_pre\_means\_4\_spk.pdf}

\caption[]%

{{\small Informal density}}

\label{fig:preinf\_raw}

\end{subfigure}

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_for\_fe\_pre\_means\_4\_spk.pdf}

\caption[Network2]%

{{\small Formal density within neighborhood}}

\label{fig:prefor\_fe}

\end{subfigure}

\hfill

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_inf\_fe\_pre\_means\_4\_spk.pdf}

\caption[]%

{{\small Informal density within neighborhood}}

\label{fig:preinf\_fe}

\end{subfigure}

\label{fig:rawbblumeans\_het}

%\vspace{-6mm}

{\scriptsize Negative distances indicate areas within project boundaries while positive distances measure areas away increasingly outside of project boundaries. Within neighborhood densities are calculated by subtracting mean densities \\[-.5em] at the 2001 census subplace level from each 25$\times$25m grid cell.}

\end{figure\*}

% Using the building-based land-use measures described in section \ref{section:data:bblu},

% we then examine housing densities across constructed and unconstructed project areas.

% To inspect how densities change with proximity to housing projects,

% Probably worth commenting on this gap in the pre-period inside the boundaries.

The building-based land-use data complement the census data by providing a spatially precise measure of preexisting development in areas prior to project construction. We calculate euclidean distances from the centroid of 25$\times$25m grid-cells to the nearest project boundary, assigning negative distances to grid-cells inside project footprints. We then compute the average density at 25m intervals, reported in houses per square kilometer. Greater negative distances exclude smaller projects in computing average densities, which may account for some of the variation in building density within projects, particularly for informal housing. Appendix~\ref{appendix:projectcounts} illustrates this compositional change in more detail.

Figures \ref{fig:prefor\_raw} and \ref{fig:preinf\_raw} respectively plot average preexisting densities of formal and informal houses in 2001. We find low densities of formal housing and very high densities of informal housing within both unconstructed and constructed project footprints (at negative distances). Crossing just outside of project footprints (to positive distances), formal housing densities quickly increase while informal housing densities sharply jump downwards. These trends are consistent with housing projects being precisely targeted within poorer neighborhoods with few existing formal dwellings. Formal and informal housing densities gradually decay moving further from project footprints. On average, constructed projects are located in neighborhoods with greater housing densities than unconstructed projects although housing densities following similar trends with respect to distance from project boundaries. Informal densities within project footprints are significantly higher for constructed projects compared to unconstructed projects, which is consistent with housing authorities prioritizing project completion in neighborhoods with large slums (\cite{hofmeyr2008risk}).

Given that housing projects appear to be located in inexpensive neighborhoods with high preexisting densities of informal housing, we next examine whether housing projects are further targeted to particular locations within neighborhoods. Figures \ref{fig:prefor\_fe} and \ref{fig:preinf\_fe} are calculated by first subtracting the average housing density in each neighborhood from the density in each grid cell, then calculating averages over distances to project footprints. Neighborhoods are defined according to 1,001 census 2001 subplaces, which identify suburbs, wards, villages, or informal settlements (\cite{censusmeta}). Figures \ref{fig:prefor\_fe} and \ref{fig:preinf\_fe} can be interpreted as the difference in housing density from the neighborhood average associated with each distance from a project boundary.

Accounting for neighborhood differences removes much of the variation in preexisting housing densities across distances to housing projects, especially outside of project footprints. Yet, low levels of formal housing and high levels of informal housing persist just within project footprints, consistent with housing authorities targeting projects toward inexpensive, poorer plots of land even within neighborhoods. Similar patterns within neighborhoods for constructed and unconstructed areas suggest that both project types are located using similar criteria. Accounting for neighborhood densities collapses average differences in densities between constructed and unconstructed projects although constructed projects have higher building densities within project footprints.

% Appendix Table~\ref{fig:rawbblumeans\_het} graphs pre-period density of formal and informal housing separately for projects in city centers and projects in suburbs. Unconstructed project areas near city centers have slightly higher densities of informal housing and lower densities of formal housing. In suburbs, the opposite pattern holds with constructed project areas having somewhat higher densities of formal housing and lower densities of informal housing. Outside of project boundaries, constructed and unconstructed projects have similar informal and formal housing levels in both city and suburb areas.

\begin{figure\*}[t!]

\centering

\caption[ Changes in Housing Densities in Constructed and Unconstructed Projects Areas]

{\small Changes in Housing Densities in Constructed and Unconstructed projects }

%\vspace{2mm}

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_for\_rawchanges\_4\_spk}

\caption[]%

{{\small formal housing density change}}

\label{fig:forchange}

\end{subfigure}

\hfill

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/bblu\_inf\_rawchanges\_4\_spk}

\caption[]%

{{\small informal housing density change}}

\label{fig:infchange}

\end{subfigure}

\label{fig:rawbblumeanschange}

\vspace{-6mm}

\end{figure\*}

To examine how project areas evolve over time, Figures \ref{fig:forchange} and \ref{fig:infchange} are constructed by computing the change in building density for each grid cell from 2001 to 2012 and then averaging these changes according to distances to project boundaries. Both formal and informal housing densities increase over this time period, reflecting rapid urbanization throughout the Guateng province.

Growth in formal housing is concentrated within constructed project footprints, dropping abruptly at project boundaries (at zero distance). This trend is consistent with housing projects producing large amounts of new, formal housing (\cite{dhsreports}). Sharp changes in formal housing growth at project boundaries are also consistent with having a precise measure of housing project boundaries.\footnote{In contrast, an imprecise measure that misclassifies some non-project areas as project areas (and vice versa) would predict a smooth average trend in formal housing changes across footprint boundary.} Compared to constructed projects, unconstructed projects experience much less growth in formal housing within their footprints. This finding confirms their status as planned, but unconstructed. Moving from within footprints to outside of footprints, unconstructed projects observe declining formal housing growth. Since projects are often located on undeveloped land, private housing development may crowd into these areas in the absence of housing project construction. Moving further from project boundaries, formal housing growth decays similarly between constructed and unconstructed areas.

% Alternatively, this growth in formal housing may represent measurement error where some projects that are classified as unconstructed were actually finished between 2001 and 2012.

Growth in informal housing density is higher within than outside of project footprints as shown by Figure~\ref{fig:infchange}. This result is unsurprising for unconstructed projects: informal settlements grow in underdeveloped, inexpensive land in the absence of project construction. Yet, for constructed projects, this result suggests that even after authorities have cleared the existing informal housing stock to make way for housing projects, total informal housing growth still exceeds growth in unconstructed project areas. Moving further from footprints, growth in informal housing decays with distance at a faster rate for unconstructed projects than for constructed projects.

% Informal housing now!!! We might think that informal housing should decline as areas are bulldozed! BUT

% DO DENSITIES TO THE CLOSEST PROJECT! ( TO SEE THE DECAY EVIDENT IN THE REGRESSIONS!!! )

% The effect of the policy is clearly evidenced in figure \ref{fig:forchange}, where tall black bars at negative distances show increases in the density of formal housing of around 1,000 structures per $\text{km}^{2}$ within constructed project areas. Compared to a baseline mean density of about 400 structures per $\text{km}^{2}$ in Figure \ref{fig:prefor}, this shift represents a nearly 250\% increase. In contrast, unconstructed areas (red bars) experience increases in formal housing density at much smaller rates.

% %Importantly, density changes at negative distances for unconstructed areas resemble changes outside of project areas in magnitude, consistent with a broader secular expansion in housing.

% At positive distances, both series are almost overlapping. This result previews our empirical analysis by showing little evidence of spillover effects on neighboring formal housing markets.

% In figure \ref{fig:infchange}, we observe large increases in informal density within project boundaries for both constructed and unconstructed project areas. At some distances, unconstructed projects appear to overtake constructed areas in informal housing growth. Outside of project boundaries, constructed areas experience larger increases in informal density than unconstructed areas. With the exception of 0-200m, these increases do not appear to be systematically different between the treated and control ares.

% %This absence of differential trends with distance previews a similar absence of spillover effects for informal housing densities explored further in the empirical analysis.

% Relative to baseline levels in Figure \ref{fig:preinf}, informal housing increases are large, increasing by at least 50\% across all distances and nearly doubling in some cases.

\begin{figure\*}[t!]

\centering

\caption[ House prices outside constructed and unconstructed projects ]

{\small House prices outside constructed and unconstructed projects }

%\vspace{2mm}

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.9cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_pre\_means\_4}

\caption[Network2]%

{{\small pre-period log-prices }}

\label{fig:preprice}

\end{subfigure}

\hfill

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.9cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_rawchanges\_4}

\caption[]%

{{\small change in log-prices}}

\label{fig:changeprice}

\end{subfigure}

\label{fig:rawpricemeans}

\vspace{-6mm}

\end{figure\*}

Figure~\ref{fig:rawpricemeans} performs an analogous exercise with property transaction data from the deeds records. These records are likely to be composed of formal houses because informal housing transactions are rarely recorded. Since there are very few non-subsidized housing transactions within project footprints, we focus only on transactions at positive distances. Using the date of each transaction, we group prices before and after the expected construction dates for both constructed and unconstructed projects.

% To make comparisons within a normalized timeline, we first residualize log purchase prices from year$\times$month and project fixed effects.

Figure~\ref{fig:preprice} plots average log purchase prices before the expected construction date for both constructed and unconstructed areas with respect to distance to project footprints. Moving from 0 to 1,500 meters from project footprints, prices increase by about 0.3 log-points or roughly 30\% for both constructed and unconstructed projects. Low prices near project boundaries are consistent with reports of public housing officials targeting low-cost land for housing development. Figure~\ref{fig:changeprice} graphs average changes in log-purchase prices before and after the expected construction of projects with respect to distance to project footprints. While prices increase overall, prices near unconstructed footprints increase more than prices for constructed footprints, especially within 500 meters from footprints.

\begin{figure\*}[t!]

\centering

\caption[ Average house prices before and after expected construction ]

{\small Average house prices before and after expected construction }

%\vspace{2mm}

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.9cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_pretrends\_close\_\_4}

\caption[Network2]%

{{\small log-prices 0 to 500m from projects}}

\label{fig:pretrend\_close}

\end{subfigure}

\hfill

\begin{subfigure}[b]{0.495\textwidth}

\centering

\includegraphics[width=\textwidth,trim={0.9cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_pretrends\_far\_\_4}

\caption[]%

{{\small log-prices 500 to 1,500m from projects}}

\label{fig:pretrend\_far}

\end{subfigure}

\label{fig:pretrend}

\vspace{-6mm}

\end{figure\*}

Deeds data contain the exact date of each housing transaction, which allows us to examine how local housing prices vary with respect to the completion date for constructed projects and the expected completion date for unconstructed projects. Figure~\ref{fig:pretrend\_close} plots average log-prices for housing transactions between 0 and 500 meters outside project footprints from two years before and after expected construction. Within 500 meters, prices for constructed and unconstructed projects follow roughly parallel trends prior to expected construction. Prices diverge after expected construction with prices near unconstructed projects rising more quickly than prices near constructed projects. Figure~\ref{fig:pretrend\_far} includes the analogous plot for housing transactions located 500 to 1,500 meters from project footprints. Again, both constructed and unconstructed prices follow parallel trends prior to expected construction, following similar slopes to their corresponding plots in Figure~\ref{fig:pretrend\_close}. After expected construction, price growth slows for areas near unconstructed projects, shrinking the gap between prices near constructed and unconstructed projects. Taken together, Figure~\ref{fig:pretrend\_close} and Figure~\ref{fig:pretrend\_far} do not provide strong evidence for pretrends in prices in anticipation of project construction.

% Prices within 500 meters of unconstructed projects maintain a steady, increasing trend two years before and after expected project completion.

% Prices for constructed and unconstructed projects also follow similar levels and trends before construction.

% We document an increase in prices of about 0.2 log-points or roughly 20\% for unconstructed project areas. The picture is less clear for constructed projects, which show an increasing price gradient within the first 500m, but exhibit noisy patterns further away. Overall, both series show large variations in house prices, which make it difficult to identify distinct patterns.

% The steady increase observed for unconstructed project areas is consistent with findings in table \ref{table:projectdescriptivescensus} that project areas (both constructed and unconstructed) fare worse in terms of indicators of housing quality.

% we observe some evidence that unconstructed areas experience a boom in nearby housing prices that was not mirrored by constructed areas. Yet, this gap quickly closes, further suggesting that housing projects play a minimal role in driving nearby housing markets.

\section{Empirical Methodology}\label{section:methodology}

To estimate the local impacts of housing projects, we implement a difference-in-differences-in-differences (or triple-differences) strategy comparing changes in outcomes nearby projects to areas further away from projects across constructed project areas and planned but unconstructed project areas. This triple differences strategy overcomes limitations of two standard differences-in-differences approaches: (1) comparing changes in outcomes nearby housing projects to areas further way from projects and (2) comparing changes in outcomes for constructed projects to planned but unconstructed projects.

First, previous research identifies housing project impacts by comparing changes in outcomes nearby projects to changes in outcomes further away from projects under two identifying assumptions: (1) project impacts dissipate with distance, and (2) areas nearby projects would have evolved in the same way as areas further away from projects in the absence of project construction (\cite{diamond2016wants,harari2018slum}). In our context, projects are specifically located in areas with inexpensive land (Table \ref{table:projectdescriptives}) and dense, preexisting slums (Table~\ref{table:projectdescriptivescensus} and Figure~\ref{fig:preinf\_raw}). Therefore, economic shifts such as changes in labor market demand for informal workers or demand for informal housing would likely produce differential impacts for these nearby areas, driving differential trends in development that are unrelated to housing project construction. Moreover, local governments may select areas for housing projects according to local trends in development, using projects to either reinvigorate deteriorating areas or build on local growth. These factors suggest that a difference-in-differences estimator that relies on distances to projects may generate biased estimates in this context.

Second, an alternative approach would be to compare changes in areas with constructed projects to areas with planned but unconstructed projects under the assumption that constructed project areas would have evolved in the same way as unconstructed project areas in the absence of construction. Constructed and unconstructed project areas appear to be located according to similar criteria: both have sharply higher levels of informal housing and lower levels of formal housing compared to their immediate neighborhoods (Figures \ref{fig:prefor\_fe} and \ref{fig:preinf\_fe}). Controlling for changes in outcomes for unconstructed projects may help to capture any location-specific trends that constructed projects would have experienced in the absence of construction. However, differences in neighborhood characteristics between constructed and unconstructed projects suggest that these projects may be exposed to differential trends at the neighborhood-level. For example, constructed project areas have higher preexisting densities of both informal and formal housing on average (Figures \ref{fig:prefor\_raw} and \ref{fig:preinf\_raw}), which may make these areas more sensitive to broad housing market fluctuations than unconstructed project areas. In fact, constructed projects experience greater growth in informal housing density than unconstructed projects, even at distances over 1 km from project boundaries where any spillover effects of project construction are likely to have dissipated (Figures \ref{fig:forchange} and \ref{fig:infchange}). Therefore, comparing changes in outcomes between constructed and constructed project areas may conflate differential neighborhood trends with the true effect of housing construction.

We propose combining near and far comparisons with constructed and unconstructed comparisons in the following triple-differences specification:

\begin{align}\label{eq:main}

\begin{split}

\quad y\_{ipnt} \, =& \, \textsc{\small Proj}\_{ip} \,\, \Big( \alpha\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \alpha\_2 \, \textsc{\small Post}\_{t} \, + \, \alpha\_3 \, \textsc{\small Const}\_{p}\, +\, \alpha\_4 \Big) \, + \\[.2em]

& \, \textsc{\small Spill}\_{ip} \, \Big( \beta\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \beta\_2 \, \textsc{\small Post}\_{t} \, + \, \beta\_3 \, \textsc{\small Const}\_{p} \, +\, \beta\_4 \Big) \, + \\[.2em]

& \, \gamma\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \gamma\_2 \,\textsc{\small Post}\_{t} \, + \, \gamma\_3 \, \textsc{\small Proj}\_{p} \, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align} % \footnote{ DO THE ROBUSTNESS EXCLUDING THESE AREAS!!!!!!! }

\noindent where $y\_{ipnt}$ is the outcome for unit $i$ in vicinity of project $p$ in neighborhood $n$ observed at time $t$ with idiosyncratic error $\varepsilon\_{ipnt}$. We consider three units of observation: a 25$\times$25m grid-cell, a census block, and a formal housing property. Units are weighted according to their spatial area to ensure that project effects are representative across space. Each unit is assigned to its nearest project according to the euclidean distances to project boundaries. $\textsc{\small Post}\_{t}$ equals one in the period after construction or expected construction and zero otherwise. $\textsc{\small Const}\_{p}$ equals one if project $p$ is constructed and zero if project $p$ remains unconstructed. $\lambda\_{n}$ includes neighborhood fixed effects and $X\_{ipnt}$ includes a vector of additional controls. The terms $\textsc{\small Post}\_{t}$, $\,\textsc{\small Const}\_{p}$, and $\textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p}$ represent a standard differences-in-differences framework, controlling for differential changes for constructed and unconstructed projects before and after construction.

We augment this standard framework by also estimating differential changes nearby project footprints (0 to 500m) relative to further away from footprints (500m to 1,500m). $\,\textsc{\small Proj}\_{ip}$ measures the percentage of unit $i$'s area that overlaps with the footprint of project $p$. When unit $i$ is measured in terms of GPS points such as the deeds data, $\,\textsc{\small Proj}\_{ip}$ is either one (if unit $i$ is within footprint $p$) or zero (if unit $i$ is outside footprint $p$). Other outcomes are measured at geographical areas such as census blocks, which are often larger than project footprints or only partially overlap with project footprints. In these cases, $\,\textsc{\small Proj}\_{ip}$ takes a value between zero and one depending on the percentage of unit $i$'s area that overlaps with the project footprint. Likewise, $\,\textsc{\small Spill}\_{ip}$ measures the percentage of unit $i$'s area that overlaps with a 0 to 500m buffer around the footprint of project $p$.

While our choice of buffer area is arbitrary, we aim to select a buffer area that is sufficiently wide to capture any potential spillover effects of housing projects. With average areas of 3 $\text{km}^{2}$, buffers from 0 to 500m are 2.5 times larger than average project footprints. To the extent that spillover effects dissipate across space, these buffers are likely to include spillover effects.\footnote{As shown in Figures~\ref{fig:prefor\_fe} and \ref{fig:preinf\_fe}, pre-period differences in building density within neighborhoods are uncorrelated with distance to project footprint at 500m and greater. This finding suggests that any preexisting spillover effects from living near project areas disappear by 500m.} Similarly, we limit the control group to areas between 500m to 1,500m under the assumption that spillover effects from housing projects have fully dissipated by 500m from project footprints. This control group provides a useful way to control for any economic or housing market trends that may affect these neighborhoods. In this framework, since each project includes two treatment groups (project and spillover) and one control group, we cluster all standard errors at the project by group level.

Interpreting coefficients on $\,\textsc{\small Proj}\_{ip}$ and $\,\textsc{\small Spill}\_{ip}$ terms as capturing the effect of moving from 0\% to 100\% spatial exposure requires two additional assumptions: (1) the spatial effects are proportional to spatial area, and (2) the distribution of project and spillover areas within outcome units is uncorrelated with the distribution of outcomes within these units. %\footnote{TEST THESE ASSUMPTIONS?! PROVIDE MORE EXPLANATION!?}

Our coefficients of interest, $\alpha\_1$ and $\beta\_1$, capture differential changes in outcomes for constructed projects relative to unconstructed projects in footprints and buffer areas ($<$500m) respectively relative to areas further away from footprints (500 to 1,500m). A causal interpretation requires the assumption that absent the completion of a housing project, changes within and nearby constructed and unconstructed areas would have evolved similarly relative to areas far away from projects. This triple-difference framework allows for constructed and unconstructed projects to experience different trends in economic development over time as long as these trends follow similar trajectories near and far from projects. This framework also permits differential economic trends in areas near and far from projects given that these differential trends are parallel between constructed and unconstructed projects.

Despite these advantages, this approach also has several limitations. A central concern is that forward-looking households anticipate project construction and alter their housing investment decisions accordingly. For example, squatter households may be more likely to construct temporary, informal structures or locate elsewhere rather than invest in long-term dwellings in areas that are likely to receive housing projects. In this case, comparisons between constructed and uncontructed areas would overestimate the role of project construction in boosting formal housing development. Similarly, nearby housing markets may anticipate projects by either appreciating or depreciating in advance (depending on the anticipated amenity value of the projects). Qualitative evidence suggests that these effects may be limited by the substantial uncertainty in project location and timing due to difficulties coordinating stakeholders and sources of funding as well as a lack of transparency among housing authorities \citep{serihistory}.\footnote{\cite{diamond2016wants} also leverage uncertainty in project timing to study affordable housing in the US.}

We are able to empirically examine anticipatory effects to some extent by observing dynamics in housing prices around the the time of construction. Descriptive Figures \ref{fig:pretrend\_close} and \ref{fig:pretrend\_far} find little evidence of differential trends in the years leading up to project construction. Moreover, the majority of our empirical approaches compare changes in outcomes between 2001 and 2011 so that our baseline time-period is likely occur well before households are aware of future project construction. This long-term comparison helps to ensure that any possible anticipatory effects play a minimal role in biasing results.

Other concerns include local economic and housing market trends correlated both with project location within neighborhoods and with whether projects are constructed. For example, if housing authorities target constructed projects to both declining or improving neighborhoods and declining or improving areas within these neighborhoods, then the triple-differences estimator would be unable to disentangle these trends from the effects of housing projects. We are unable to empirically test for the presence of these trends.

\subsection{Estimating Heterogeneity by Income}

We also examine how effects may vary by neighborhood income quartile according to the following specification:

\begin{align}\label{eq:mainhet}

\begin{split}

\quad y\_{ipnt} \, =& \sum\_{q=1}^{4} \, Q\_{n}^{q} \, \Bigg[ \, \, \textsc{\small Proj}\_{ipq} \,\, \Big( \alpha\_1^q \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \alpha\_2^q \, \textsc{\small Post}\_{t} \, + \, \alpha\_3^q \, \textsc{\small Const}\_{pq}\, +\, \alpha\_4^q \Big) \, + \\[.2em]

& \, \textsc{\small Spill}\_{ipq} \, \Big( \beta\_1^q \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \beta\_2^q \, \textsc{\small Post}\_{t} \, + \, \beta\_3^q \, \textsc{\small Const}\_{pq} \, +\, \beta\_4^q \Big) \, + \\[.2em]

& \, \gamma\_1^q \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \gamma\_2^q \,\textsc{\small Post}\_{t} \, + \, \gamma\_3^q \, \textsc{\small Proj}\_{pq} \, \Bigg] \, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align}

\noindent where $q$ indexes neighborhood income quartiles, and $Q\_s^q$ takes a value of one if average household income for neighborhood $n$ falls within quartile $q$. Average income is measured by taking average household incomes across census subplaces as reported in the 2001 Census of Population and Housing.

\subsection{Estimating Housing Demand}

The triple-difference approach also provides a useful framework for estimating residential housing demand given by equation (\ref{eq:housingdemand}). Our empirical setting allows us to identify the causal effect of housing projects on local amenities net of unobserved housing costs. Variation in housing costs also identify the marginal utility of income, which helps translate amenity effects of housing projects into monetary welfare estimates.

To leverage housing cost variation, we divide housing costs into an unobserved portion, $c^{u}\_{hlt}$ that varies at the location level and an observed portion, $c^{o}\_{j}$ so that total costs are $C\_{hjt} = c^{u}\_{hlt} + c^{o}\_{j}$. To measure the welfare effects of local housing projects, let $ \overline{V}\_{hlt}$ equal the amenity value net of unobserved housing costs, $(\delta\_{hlt}- \theta c^{u}\_{hlt} )$, which may vary over time, project construction status, and location in the following triple-differences framework

\begin{align\*}

\overline{V}\_{hlt} &= \textsc{\small Proj}\_{lp} \,\, \Big( \alpha^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \alpha^{Vh}\_2 \, \textsc{\small Post}\_{t} \, + \, \alpha^{Vh}\_3 \, \textsc{\small Const}\_{p}\, +\, \alpha^{Vh}\_4 \Big) \, + \\[.2em]

& \, \textsc{\small Spill}\_{lp} \, \Big( \beta^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \beta^{Vh}\_2 \, \textsc{\small Post}\_{t} \, + \, \beta^{Vh}\_3 \, \textsc{\small Const}\_{p} \, +\, \beta^{Vh}\_4 \Big) \, + \\[.2em]

& \, \gamma^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \gamma^{Vh}\_2 \,\textsc{\small Post}\_{t} \, + \, \gamma^{Vh}\_3 \, \textsc{\small Proj}\_{p}

\end{align\*}

$\alpha^{Vh}\_1$ and $\beta^{Vh}\_1$ measure the effect of housing projects on local amenities net of unobserved housing costs by housing sector in project and spillover areas respectively. Without direct measures of housing costs and amenities, we are unable to disentangle the relative contributions of cost and amenity changes; however, changes in amenities net of housing costs summarize the total local welfare effects of housing projects.

We assume that idiosyncratic amenity shocks, $\epsilon\_{hjt}$ for formal and informal housing sectors follow a bivariate normal distribution with a two-by-two variance matrix $\Sigma$ with diagonal terms capturing the variance of formal housing, $\sigma\_{for}^{2}$ and informal housing, $\sigma\_{inf}^{2}$ and off-diagonal terms both capturing the correlation, $\rho$ between formal and informal amenity shocks in the same plot. We also assume that amenity shocks are uncorrelated with all other determinants of household utility: (1) local fixed amenities net of housing costs, $\overline{V}\_{hlt}$, (2) observed housing costs, $c^{o}\_{j}$, (3) disamenities from crowding, $\Lambda(k)$, as well as (4) reservation utility, $\overline{U}$. Under these assumptions, $\alpha^{Vh}\_1$ and $\beta^{Vh}\_1$ measure the causal effects of housing projects in an analogous way to equation (\ref{eq:main}). $\theta$ measures the marginal utility of income from observed variation in housing costs. As described in Section~\ref{section:data}, we proxy for observed housing costs using average land slope because steep land slopes increase the costs of home construction. Identification of $\theta$ requires assuming that land slope only affects the number of buildings through its effect on construction and maintenance costs. This assumption may be violated particularly in cases where households have particular preferences for living on sloped land. Disamenities from crowding, $\Lambda(k)$ are identified from the distribution of the number of houses per plot given the bivariate normal distribution of amenity shocks. The frequency of plots with both formal and informal housing identify the correlation term, $\rho$. All estimates are normalized with respect to $\sigma\_{for}^{2}$ and $\sigma\_{inf}^{2}$, which are not separately identified in this framework.

Under these assumptions, the probability of observing a given number of formal and informal houses per plot according to equation (\ref{eq:housingdemand}) can be written in terms of a seemingly unrelated bivariate ordered probit model as follows \citep{sajaia2008maximum}

\begin{align\*}

Pr[ k\_{for,jt}^{\*} =k\_{for} , k\_{inf,jt}^{\*} =k\_{inf} ] \, = \, & \Phi\_2\Bigg(\dfrac{\Lambda\_{for}(k\_{for}+1) - \overline{V}\_{for,jt} - \theta c^{o}\_{j}}{\sigma\_{for}}, \dfrac{\Lambda\_{inf}(k\_{inf}+1) - \overline{V}\_{inf,jt} - \theta c^{o}\_{j}}{\sigma\_{inf}} , \rho \Bigg) \\

- \, &\Phi\_2\Bigg(\dfrac{\Lambda\_{for}(k\_{for}) - \overline{V}\_{for,jt} - \theta c^{o}\_{j}}{\sigma\_{for}}, \dfrac{\Lambda\_{inf}(k\_{inf}+1) - \overline{V}\_{inf,jt} - \theta c^{o}\_{j}}{\sigma\_{inf}} , \rho \Bigg) \\

- \, &\Phi\_2\Bigg(\dfrac{\Lambda\_{for}(k\_{for}+1) - \overline{V}\_{for,jt} - \theta c^{o}\_{j}}{\sigma\_{for}}, \dfrac{\Lambda\_{inf}(k\_{inf}) - \overline{V}\_{inf,jt} - \theta c^{o}\_{j}}{\sigma\_{inf}} , \rho \Bigg) \\

+ \, &\Phi\_2\Bigg(\dfrac{\Lambda\_{for}(k\_{for}) - \overline{V}\_{for,jt} - \theta c^{o}\_{j}}{\sigma\_{for}}, \dfrac{\Lambda\_{inf}(k\_{inf}) - \overline{V}\_{inf,jt} - \theta c^{o}\_{j}}{\sigma\_{inf}} , \rho \Bigg) \\

\end{align\*}

\noindent where $\Phi\_{2}(.)$ indicates a standard normal bivariate distribution with correlation terms, $\rho$. This expression can be estimated by maximizing the following log-likelihood function

\begin{align}

\label{eq:demandll}

LL = \sum\_{t=1}^{T} \sum\_{j=1}^{J} \sum\_{m\_{for}=1}^{K\_{for}} \sum\_{m\_{inf}=1}^{K\_{inf}} \mathbbm{1} \{k\_{for,jt} = m\_{for}, k\_{inf,jt} = m\_{inf}\} \,\,\, log \Bigg( Pr[ k\_{for,jt}^{\*} =k\_{for} , k\_{inf,jt}^{\*} =k\_{inf} ] \Bigg)

\end{align}

\noindent where $J$ is the total number of plots observed, $T$ is the total number of time periods, and $K\_{for}$ and $K\_{inf}$ are the maximum observed number of formal and informal houses observed per plot.

% Using variation across time we are only able to estimate...

%%% HOW BOUT THE NEAREST NEIGHBOR ASSUMPTION!!?!?!?!

% driven by changes in local political or economic conditions that may also affect local housing development. For example, growing neighborhoods may generate greater political resistance to housing projects, delaying project construction despite improvements in local housing. Alternatively, worsening infrastructure quality may both increase the costs of finishing projects while also worsening local housing conditions.

% To the extent that these trends are reflected in average levels of local characteristics, we can evaluate potential sources of endogeneity by comparing baseline characteristics across constructed and unconstructed areas as in Section~\ref{section:descriptives}. However, we are unable to test for these factors if changes in local economic and political conditions happen at the same time as project construction.

% How is this assumption weaker than standard differences in differences?

% We interpret the coefficients of interest $\alpha^d$ as the causal impacts of subsidized housing on outcome $y$ at distance $d$ from a housing project. Specifically, $\alpha^d$ measures the differential change in outcomes for constructed projects, relative to unconstructed projects. We discuss caveats below.

% By comparing unconstructed versus constructed project areas, this method builds on previous studies of housing spillovers, which instead focus on comparing areas nearby to areas just further away from projects (\cite{diamond2016wants}). The advantage of our method is that it does not require assuming how spillover effects dissipate with distance ex-ante; instead, we directly estimate these spillovers. It is particularly helpful to relax this assumption for evaluating South African housing programs since (1) the large sizes of projects suggest that spillover effects may extend for long distances and (2) program recipients may be drawn from wide areas around housing projects.

\section{Estimation Results}\label{section:results}

% To draw statistical inference on the patterns depicted in section \ref{section:descriptives}, we proceed to estimate variants of equation (\ref{eq:main}) for each outcome of interest.

% d is 25 by 25, and t = 2001 and 2011

\subsection{Housing Density}\label{section:bbluestimates}

We first investigate how housing projects impact housing density within and around their footprints. We estimate the reduced form equation (\ref{eq:main}) where the outcomes are building density measures for 25$\times$25m grid cells in 2001 and 2012 and $\textsc{\small Post}\_{t}$ equals one for observations in 2012. Table~\ref{table:bbluDDD} reports the triple-difference coefficients of interest for (1) within project footprints and (2) 0-500m outside of project footprints.\footnote{Appendix Table~\ref{table:bbluDDDfull} includes the full set of coefficients.} Results are reported in terms of houses/$\text{km}^{2}$. Appendix Table~\ref{table:bbluDDDhet} reports triple-difference coefficients separately by neighborhood income quartile following the specification in equation (\ref{eq:mainhet}).

Column (1) of Table~\ref{table:bbluDDD} finds that total housing density within project footprints statistically significantly increases by 618 houses/$\text{km}^{2}$, which is over twice the baseline average of 579 houses/$\text{km}^{2}$. This finding suggests that government housing projects are able to substantially out-pace any private housing development in similar but unconstructed project footprints. At the same time, total housing density 0-500m outside of projects decreases by around 217 houses/$\text{km}^{2}$ with statistical significance at the 10\% level. Given an average project area of 1.2$\text{km}^{2}$ and an average buffer area of 3$\text{km}^{2}$, the magnitudes of these estimates suggest that each project generates around 742 new houses inside project footprints while crowding out around 651 houses in nearby areas.\footnote{742 new houses per footprint is much larger the 374 houses identified as project houses within footprints using the deeds transaction data in Table~\ref{table:projectdescriptives}. This discrepancy is likely due to under-reporting of housing projects in deeds records \citep{seriq}.} Appendix Table~\ref{table:bbluDDDhet} includes corresponding estimates by neighborhood income quartile. Column (1) of Table~\ref{table:bbluDDDhet} finds that quartiles with larger increases in total housing (ie. Q2 and Q3) are also accompanied by greater decreases in housing density 0-500m outside of housing projects. This finding provides further suggestive evidence that housing projects crowd-out nearby housing growth.

% Backyard housing is a common phenomenon in South Africa where households erect informal dwellings within the plots of formal dwellings \citep{Brueckner2018backyarding}. Project houses may be well-suited for backyarding because each house receives an individual landplot that is often much larger than the house itself. put this up top!! with a discussion of the maps!!

To examine the composition of housing growth, Columns (2) through (5) of Table~\ref{table:bbluDDD} disaggregate total housing into formal housing, informal non-backyard housing, and informal backyard housing. Column (2) finds that the increase in total housing within projects is almost entirely driven by formal housing with a statistically significant, positive coefficient of 620 houses/$\text{km}^{2}$. This finding matches earlier descriptive evidence from Figure \ref{fig:rawbblumeanschange}. Large reductions in informal non-backyard housing (Column 4) within projects are consistent with qualitative evidence that housing authorities cleared preexisting slums before constructing new projects \citep{hofmeyr2008risk}. Column (4) documents a statistically significant decrease of 573 informal non-backyard houses/$\text{km}^{2}$ within projects, which suggests that these slum clearance efforts were particularly aggressive given low preexisting densities of around 184 houses/$\text{km}^{2}$. Column (5) indicates that project areas also experience substantial growth in backyard informal housing on the order of 542 houses/$\text{km}^{2}$, which almost exactly offsets reductions in informal non-backyard houses.

Appendix Table~\ref{table:bbluDDDhet} finds similarly zero net changes in informal housing within projects across all neighborhood income quartiles. Yet, compared to richer quartiles (Q3 and Q4), poorer quartiles (Q1 and Q2) experience larger decreases in non-backyard informal housing that are offset by larger increases in backyard informal housing.

According to Columns (2) through (5) of Table~\ref{table:bbluDDD}, total reductions in housing density 0-500m outside of project footprints are driven by decreases in informal non-backyard housing and informal backyard housing although these coefficients are statistically insignificantly different from zero. Column (2) shows a statistically insignificant decrease in formal housing nearby housing project boundaries, which appears to contrast with descriptive evidence in Figure \ref{fig:rawbblumeanschange}. This discrepancy is likely due to differences in weighting observations: while Figure \ref{fig:rawbblumeanschange} computes average changes in formal housing according to distance bins, Table~\ref{table:bbluDDD} reports differential changes that are spatially representative, which results in applying relatively greater weights to areas at further distances from project boundaries since further distances cover greater land areas. These findings suggest that housing projects primarily crowd-out nearby informal housing supply with little impact on nearby formal housing supply.

% On net, the quantity of informal housing within projects remains roughly unchanged.

% discuss appendix table, how the substitution is concentrated in lower income quartiles

% Meanwhile, backyard housing growth (Column (3)) replaces informal non-backyard housing (Column (4)) that is removed as part of well-documented slum clearance programs in many of these projects \citep{hofmeyr2008risk}.

% Formal housing growth drives almost all of the increase in density within project footprints while informal housing generates most of the decline in density outside of project areas. To the extent that formal housing has higher quality than informal housing, these results suggest that housing project generate an overall upgrade in local housing quality. Within project areas, informal housing density remains nearly unchanged despite well-documented slum clearance programs in many of these projects \citep{hofmeyr2008risk}.\footnote{Cite APPENDIX WITH HETEROGENEITY OF RESULTS ACCORDING TO PROJECT TYPE< AND DESCRIBE PROJECT TYPE ANALYSIS!!!!}

%Little change in total informal housing masks significant shifts in the type of informal housing within project footprints. Our density measure differentiates informal housing between backyard and non-backyard housing. Backyard housing is a common phenomenon in South Africa where households erect informal dwellings within the plots of formal dwellings \citep{Brueckner2018backyarding}. Project houses are especially well-suited for backyarding because each house receives an individual landplot that is often much larger than the house itself. Columns (4) and (5) include results for informal backyard and non-backyard housing respectively. In project footprints, backyard housing almost completely replaces non-backyard informal housing as households erect new informal houses within the plots of project houses. Both backyard and non-backyard housing decline 0-500m nearby projects although the coefficients are statistically insignificant. These results suggest that bulldozing slums and replacing them with new housing projects does not change the resulting density of informal housing within project footprints. However, these projects may have spillover effects on neighboring areas by crowding out growth of nearby informal housing.

\begin{table}[hbt!]

\small

\centering

\caption{Housing Density Estimates}\label{table:bbluDDD}

\vspace{-2mm}

\begin{tabular}{lCCCCC}

\toprule

& \small (1) & \small (2) & \small (3) & \small (4) & \small (5) \\

& \small Total Housing & \small Total Formal Housing & \small Total Informal Housing & \small Informal Non-Bkyd. Housing & \small Informal Backyard Housing \\ \midrule

\input{figures/b\_ksp\_o0\_d250\_500\_top}

\bottomrule\\[-.8em]

\multicolumn{6}{l}{\footnotesize Standard errors clustered at the project level in parenthesis. \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 }\\[-.3em]

% \multicolumn{6}{l}{\footnotesize Project and spillover areas are 25$\times$25m grid cells.}\\[-.3em]

\multicolumn{6}{l}{\footnotesize Outcomes are measure in terms of houses per $\text{km}^{2}$. }\\[-.3em]

\multicolumn{6}{l}{\footnotesize Estimates control for 2001 census sub-place fixed effects.}\\[-.3em]

\multicolumn{6}{l}{\footnotesize Coefficients are interactions with constructed, $\textsc{Const}\_{p}$ and post, $\textsc{Post}\_{t}$ indicators.}\\[-.3em]

\multicolumn{6}{l}{\footnotesize Appendix Table~\ref{table:bbluDDDfull} includes all corresponding triple-difference coefficients.}

\end{tabular}

\end{table}

\subsection{Housing Demand}\label{section:demandestimates}

Structural estimates of housing demand provide a mapping from reduced-form changes in housing density into estimates of the local welfare effects of housing projects. We estimate equation (\ref{eq:demandll}) where the unit of analysis is a 25m$\times$25m plot in 2001 and 2012 and the outcome variables are the counts of formal and formal houses on each plot. Outcome variables count the number of buildings from 0 to 9 with a last category indicating 10 or more buildings. 86\% of plot observations have zero formal buildings and 91\% have zero informal buildings. Less than 0.05\% of plots have more than 10 total buildings. Appendix Table~\ref{figure:buildhist} provides a histogram of building counts for plots with at least zero formal or informal houses.

Table~\ref{table:housingdemandestimates} includes estimates for the parameters of interest while Appendix Table~\ref{table:housingdemandestimates\_full} includes the full set of estimates. According to Table~\ref{table:housingdemandestimates}, estimates of the marginal utility of income, $\theta$, are similar across formal and informal sectors suggesting that higher costs of housing from steeply sloped land have a similar impact on housing construction across sectors. These estimates allow us to convert net amenity effects from units of utility into units of income by dividing net amenity coefficients by $\theta$. For this exercise, we refer to the estimate of $\theta$ for formal housing since the variation in housing costs underlying this estimate are relevant to formal housing construction (\cite{redbook}). This exercise shows that for the formal housing sector, average net amenities increase by R\input{figures/for\_proj}per plot within project areas and decrease by R\input{figures/for\_spill}within spillover areas. For the informal housing sector, average net amenities decrease by R\input{figures/inf\_proj}per plot within project areas and decrease by R\input{figures/inf\_spill}within spillover areas. Given average housing prices of R231,000, these results translate into price changes of between -\input{figures/inf\_proj\_per}\unskip\% for informal housing in project areas to +\input{figures/for\_proj\_per}\unskip\% for formal housing in project areas.

Positive estimates of $\rho$ in Table~\ref{table:housingdemandestimates} indicate that unobserved plot-specific amenities for formal and informal housing are positively correlated, which is likely driven by prevalent backyard housing. Appendix Table~\ref{table:housingdemandestimates\_full} includes estimates for cut-points that determine breaks in the distribution of unobserved amenities that correspond to each equilibrium number of houses. The lowest cut point for formal housing is slightly more negative than the lowest cut point for informal housing while the highest cut point for formal housing is substantially more positive than the highest cut point for informal housing. These cut point estimates parallel the histogram of building counts in Appendix~\ref{appendix:histfreq}: plots are more likely to have any formal housing than any informal housing; however, conditional on having any housing, plots are more likely to have many informal houses than many formal houses. These findings provide suggestive evidence that the supply of informal housing is more elastic than the supply of formal housing in this context.

Results for formal housing in Table~\ref{table:housingdemandestimates} mirror reduced-form results from Table~\ref{table:bbluDDD}. Project areas promote statistically significant growth of formal housing within their footprints with little change in spillover areas. Structural estimates for informal housing contrast with reduced form results: both specifications suggest declines in informal housing nearby project footprints, but structural estimates further predict significant declines in informal housing within project areas while reduced form estimates find little change in nearby informal housing. One explanation for this discrepancy is that structural estimates may prioritize extensive margins of whether plots have buildings while reduced-form estimates simply measure the average change in building density. Therefore, the structural approach may interpret the compositional shift away from non-backyard toward backyard informal housing in project areas as a decline in the local net amenities enjoyed by the informal housing sector.

We also allow for heterogeneity in housing demand by neighborhood income quartile with results presented in Appendix Table~\ref{table:housingdemandestimatesinc}. While effects for formal housing do not appear correlated with neighborhood income quartile, informal housing markets exhibit a clear negative gradient with respect to neighborhood income: lower income neighborhoods experience improvements in local amenities while high income neighborhoods experience steep declines in local amenities in both project and spillover areas.

Changes in local net amenities are likely to overstate the full welfare effects of housing projects as renters substitute between project areas, spillover areas, and outside areas. To compute the full welfare effects associated with changes in local amenities, we use estimates from Table~\ref{table:housingdemandestimates} to simulate the effect of housing projects on local building growth. Given simulated choices, we calculate the change in net rents extracted by landlords, which fully summarize total welfare effects. Appendix~\ref{section:appendixwelfaresim} includes details on the simulation exercise and Table~\ref{table:housingdemandsim} includes predicted net rental changes by housing sector, project and spillover areas, and neighborhood income quartile.

Within housing projects, this exercise predicts that net rents from formal housing increase by R\input{figures/for\_proj\_w}on average per plot; at the same time, net rents from informal housing decrease by R\input{figures/inf\_proj\_w}on average per plot. Together, these findings suggest that average rents per plot increase by a total of R\input{figures/full\_proj\_w}within project footprints. High average densities of informal housing magnify the effect of declines in local informal housing amenities on welfare, almost offsetting welfare gains from formal housing. Patterns by neighborhood income quartile reflect housing demand coefficient estimates in Appendix Table~\ref{table:housingdemandestimatesinc}.

In spillover areas, housing projects drive a slight decrease in nearby net rents from formal housing of R\input{figures/for\_spill\_w}per plot and a more substantial decrease in nearby net rents from informal housing of R\input{figures/inf\_spill\_w}per plot, which together lead to a total decline of R\input{figures/full\_spill\_w}\unskip. Effects for spillover areas are smaller per plot than corresponding effects for project areas likely because project areas have higher overall densities of housing.

To compute the overall average change in net rents per plot as a result of housing projects, we weight effects for project areas and spillover areas by their corresponding land areas (1.2$\text{km}^{2}$ and 3$\text{km}^{2}$ respectively). We compute an average decline of R\input{figures/full\_w}per plot. This change represents around a \input{figures/full\_w\_per}\unskip\% decline in average housing prices.

To provide a rough estimate of the aggregate welfare effects of housing projects, we compare changes in net rents at the project-level to project implementation costs. Given around 6,720 plots surrounding project areas, the project-level decline in rents totals 26 million Rand (3.4 million USD). National budget reports suggest an average cost per project of 25 million Rand (3.2 million USD). In total, these results suggest that housing projects generate an aggregate decline in welfare of 51 million Rand (6.6 million USD).

% These welfare effects contrast with Table~\ref{table:housingdemandestimates}, which predicts that net amenity increases for formal housing more than offset decreases in net amenities for informal housing in project areas.

\begin{table}[h]

\centering

\caption{Housing Demand Estimates}\label{table:housingdemandestimates}

\vspace{-2mm}

\begin{tabular}{lDD}

\toprule

& Formal Housing & Informal Housing \\ \midrule

Net amenities: $( \delta\_{hl} - \theta c^{u}\_{hlt} )$ \\[.3em]

\input{figures/forimpcmp10\_1}[.3em]

\input{figures/forimpcmp10\_2}

\\

Correlation of amenity shocks: $\rho\_{for,inf}$ & $0.620^{\text{a}}$ & \\[-.3em]

& (0.018) & \\[.3em]

N & 6,143,756 & \\

\bottomrule\\[-.8em]

% Implied amenity value net of housing costs in Rands: $\frac{1}{\theta\_{for}}(\delta\_{hl}- \theta\_{for}c^{u}\_{hlt})$ & & \\

% inside project &\input{figures/for\_proj} & \input{figures/inf\_proj} \\

% 0-500m outside &\input{figures/for\_spill} & \input{figures/inf\_spill} \\

% \bottomrule\\[-.8em]

\multicolumn{3}{l}{ \footnotesize Standard errors clustered at the project level in parenthesis. \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 } \\[-.3em]

\multicolumn{3}{l}{ \footnotesize Outcome categories by plot and year include categories for formal and informal} \\[-.3em]

\multicolumn{3}{l}{ \footnotesize building counts from 0 to 9 and a category at least 10. Table~\ref{table:housingdemandestimates\_full} includes the full set} \\[-.3em]

\multicolumn{3}{l}{ \footnotesize of coefficients. Net amenities equal local amenities minus local housing costs. } \\[-.3em]

\end{tabular}

\end{table}

\subsection{Infrastructure and Demographics}\label{section:resultscensus}

Given large shifts in quantities and types of housing within and around projects, we next use census data to understand whether housing projects also affect infrastructure access, housing quality, and residents' demographics. We reestimate equation (\ref{eq:main}) using census data from 1996, 2001, and 2011. Outcomes include housing and demographic characteristics averaged across households in each census block and year. $\textsc{\small Post}\_{t}$ equals one for observations from the 2011 census. $\textsc{\small Proj}\_{i}$ and $\textsc{\small Spill}\_{i}$ measure the fraction of spatial overlap of each census block with project areas and spillover areas (0-500m) respectively.\footnote{3,118 blocks partially overlap with project areas while 1,138 blocks fully overlap with project areas. Likewise, 7,456 blocks partially overlap with spillover areas while 1,402 blocks fully overlap with spillover areas.} Specifications control for the spatial area of census blocks up to a cubic term and interacted with year in order to account for the possibility that household sampling may vary according to census block size. All estimates are weighted by the spatial area of census blocks so that the estimates are representative of the spatial impacts of the projects. Weighting according to spatial areas further ensures that the census results are directly comparable to the building density results. Table~\ref{table:censusestimates} reports the triple-difference coefficients of interest for (1) within project footprints and (2) 0-500m outside of project footprints.\footnote{Appendix Tables \ref{table:censusestimatesfull} and \ref{table:censusestimatesfull2} include the full set of coefficients. Census blocks lack the spatial precision needed to estimate effects separately by quartile of neighborhood income.}

% Appendix Tables \ref{table:censusestimateshet} and \ref{table:censusestimateshet2} report triple-difference coefficients for spillover areas separately by neighborhood income quartile following the specification in equation (\ref{eq:mainhet}).\footnote{Census blocks lack the spatial precision to estimate within-project effects separately by quartile of neighborhood income.}

\begin{table}

\small

\centering

\caption{Census Household-level Estimates}\label{table:censusestimates}

\vspace{-2mm}

\begin{tabular}{lDDDDDD}

\toprule

& \small (1) & \small (2) & \small (3) & \small (4) & \small (5) & \small (6) \\

& \small Is a Formal House & \small Has a Flush Toilet & \small Has Piped Water Indoors & \small Has Electricity & \small Total Rooms & \small Owns House \\ \midrule

\input{figures/ch1\_ksp\_o0\_d250\_500\_bb1\_top} \midrule

\\

& \small (7) & \small (8) & \small (9) & \small (10) & \small (11) & \small (12)\\

& \small HH Size & People per $\text{km}^{2}$ & Age HoH & HoH is African & HoH is Employed & HH Log Income \\ \midrule

\input{figures/ch2\_ksp\_o0\_d250\_500\_bb1\_top\_1}

% \input{figures/census\_hh\_DDregs\_AGG}

\bottomrule\\[-.6em]

\multicolumn{7}{l}{\footnotesize Standard errors clustered at the project level in parenthesis.\textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Outcomes in Columns (1) to (4), (6), and (9) to (11) are indicator variables.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Specifications include census 2001 subplace fixed effects.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Results are weighted by unit area and control for up to a cubic in unit area by time. }\\[-.3em]

\multicolumn{7}{l}{\footnotesize ``HH'' stands for household and ``HoH'' stands for head of household.}

\end{tabular}

\end{table}

Columns (1) to (6) of Table~\ref{table:censusestimates} capture changes in housing characteristics within and around housing project footprints. Inside projects, the share of households living in formal housing jumps by 61 percentage points consistent with the large growth in formal housing density documented in Table~\ref{table:bbluDDD}. The shares of households in project areas with flush toilets, piped water indoors, and/or electricity also increase by economically meaningful magnitudes although only the coefficient for access to piped water indoors is statistically significant. The average number of rooms per house increases by 1.9 likely because new project houses are often larger than informal backyard and non-backyard houses. The coefficient for the share of households in project areas that report owning their home is positive but statistically insignificant.

In spillover areas (0-500m outside of projects), the share of households that report living in formal houses (Column (1)) increases by a statistically insignificant 12 percentage points. Any increase in the ratio of formal to informal housing nearby projects may reflect crowding-out of nearby informal housing as described by housing density estimates from Table~\ref{table:bbluDDD}. Economically small and statistically insignificant coefficients for infrastructure access, number of rooms, and homeownership in spillover areas in Columns (2) to (5) suggest little scope for infrastructure and housing stock investments in project areas to spillover to surrounding neighborhoods.

Complementing changes in housing characteristics, Columns (7) to (12) of Table~\ref{table:censusestimates} estimate changes in demographics within and around project footprints. Within projects, household (``HH'') sizes increase by 0.59 people on average given a baseline average of 3.25 people per household, which is consistent with recipients of project houses inviting additional family members to live in their new houses. Population density increases by 1,610 people per $\text{km}^{2}$, which is statistically insignificant, but economically meaningful --- this effect almost doubles the baseline average population density of 1,793. Coefficients for average age, race, and employment status of the head of households (``HoH'') as well as household log-income are economically small and statistically insignificant, suggesting little demographic change as a result of these housing projects. The point estimate for household income (Column (12)) suggests a 12\% increase in average household incomes in project areas although this result is statistically insignificant. Taken together, these results are consistent with housing authorities ensuring that preexisting slum residents are able to benefit from new project houses. The results are less consistent with the hypothesis that political corruption steers project houses toward wealthier and more politically connected households \citep{seriq}. %%% maybe reinterpret in light of the new model...

Spillover areas report positive, statistically insignificant coefficients for both household size (Column (7)) and population growth (Column (8)) that are much smaller in magnitude and significance than corresponding estimates inside projects. Spillover estimates for household demographics in Columns (8) to (11) parallel estimates within projects by having low statistical significance and small magnitudes. Estimates predict a large, statistically insignificant increase of 21\% in average household income.

\subsection{Prices}\label{section:resultsprices}

We examine how housing projects may affect the prices of nearby formal housing using deeds data on housing transactions. Because few private housing transactions occur within project footprints and because it is possible that project house allocations may be miscoded as private housing transactions, we restrict our analysis to properties located outside of project footprints. We estimate the following modification of equation (\ref{eq:main}):

\begin{align}\label{eq:mainprice}

\begin{split}

\quad y\_{ipnt} \, =& \,\, \textsc{\small Spill}\_{ip} \, \Big( \beta\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \beta\_2 \, \textsc{\small Post}\_{t} \, + \, \beta\_3 \, \textsc{\small Const}\_{p} \, +\, \beta\_4 \Big) \, + \\[.2em]

& \, \gamma\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \gamma\_2 \,\textsc{\small Post}\_{t} \, + \, \gamma\_3 \, \textsc{\small Proj}\_{p} \, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align}

\noindent where the unit of analysis, $i$ is a property transaction, time, $t$ is monthly, and the outcome, $y\_{ipnt}$ is the log-price of each property transaction. To utilize monthly transaction data, we limit the sample to include constructed housing projects where the month of completion is available and unconstructed housing projects where we are able to project an expected month of construction. We refer to the completion month for constructed projects and the expected completion month for unconstructed projects jointly as the scheduled construction month. $\textsc{\small Post}\_{t}$ takes a value of one for transactions occurring after the scheduled construction month. $\textsc{\small Spill}\_{ip}$ takes a value of one for properties between 0 and 500 meters outside of their nearest project boundary. Some specifications include additional controls, $X\_{ipnt}$ for the calendar month of each transaction and up to a cubic polynomial in the plot size of each property. All specifications control for neighborhood fixed effects, $\lambda\_n$. % \footnote{EVALUATE THIS SAMPLE LIMITATION?}

To examine how price effects may vary by neighborhood income quartile, we also estimate the following modification of equation (\ref{eq:mainhet}):

\begin{align}\label{eq:mainhetprice}

\begin{split}

\quad y\_{ipnt} \, =& \sum\_{q=1}^{4} \, Q\_{n}^{q} \, \Bigg[ \,\, \textsc{\small Spill}\_{ipq} \, \Big( \beta\_1^q \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \beta\_2^q \, \textsc{\small Post}\_{t} \, + \, \beta\_3^q \, \textsc{\small Const}\_{pq} \, +\, \beta\_4^q \Big) \, + \\[.2em]

& \, \gamma\_1^q \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \gamma\_2^q \,\textsc{\small Post}\_{t} \, + \, \gamma\_3^q \, \textsc{\small Proj}\_{pq} \, \Bigg] \, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align}

\noindent where $Q\_n^q$ indexes neighborhood income quartile.

Table~\ref{table:priceDDD\_het} reports estimation results for all neighborhoods following equation (\ref{eq:mainprice}) and separately by neighborhood quartile following equation (\ref{eq:mainhetprice}). Table~\ref{table:priceDDD\_het} includes triple difference coefficients, $\beta\_1$ which capture the differential price changes for properties between 0 and 500 meters from project boundaries before and after scheduled construction for constructed versus unconstructed projects. With log-prices as an outcome, coefficients can be interpreted as in terms of percent of transaction prices.

\begin{table}

\small

\centering

\caption{Triple-Difference Estimates of Log-Prices 0 to 500 Meters Outside of Housing Projects}\label{table:priceDDD\_het}

\vspace{-2mm}

\begin{tabular}{lCC}

\toprule

& \small (1) & \small (2) \\ \midrule

\input{figures/price\_temp\_Tester\_3d\_spill\_top\_no\_lab}\\

By Neighborhood Income Quartile \\[.3em]

\input{figures/price\_temp\_Tester\_3d\_q\_top}

\bottomrule

\multicolumn{3}{l}{\footnotesize Estimates capture differential price effects for properties } \\[-.3em]

\multicolumn{3}{l}{\footnotesize 0 to 500 meters outside of project boundaries relative to } \\[-.3em]

\multicolumn{3}{l}{\footnotesize properties 500 to 1,500 meters from boundaries before and } \\[-.3em]

\multicolumn{3}{l}{\footnotesize after scheduled construction for constructed and unconstructed projects. } \\[-.3em]

% \multicolumn{3}{l}{\footnotesize } \\[-.3em]

\multicolumn{3}{l}{\footnotesize Income quartiles are from average household 2001 incomes in } \\[-.3em]

\multicolumn{3}{l}{\footnotesize census 2001 subplaces. All estimates include neighborhood fixed effects.} \\[-.3em]

\multicolumn{3}{l}{\footnotesize Standard errors clustered at the project level in parenthesis.} \\[-.3em]

\multicolumn{3}{l}{\footnotesize \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 } \\[-.3em]

\end{tabular}

\end{table}

Column (1) of Table~\ref{table:priceDDD\_het} documents a statistically insignificant 7\% decrease in housing prices, which shrinks to 6\% in column (2) after accounting for year-month fixed effects as well as up to a cubic polynomial in plot size. Examining price effects by quartile of neighborhood income, we find that price effects become increasingly negative in higher quartile neighborhoods. The poorest quartile (Q1) exhibits a small, statistically insignificant 3\% change in prices (column (2)) while the richest quartile (Q4) experiences a 17\% decrease in prices, which is statistically significant at the 5\% level. Comparing columns (1) and (2), we find that additional controls have little impact on the magnitudes or statistical significance of the estimates.

Appendix~\ref{section:appendixpricedist} estimates differential changes in prices before and after scheduled construction for constructed versus unconstructed projects non-parametrically according to distance to the nearest project boundary. The strongest decreases in prices appear concentrated closest to project boundaries --- especially for the richest income quartile (Q4) --- although not all price gradients are monotonically increasing with distance to project boundaries.

To examine the dynamics of these price effects, Appendix~\ref{section:appendixpricetime} provides corresponding estimates of differential changes in prices for properties near versus far from project boundaries for constructed versus unconstructed projects non-parametrically by months to scheduled construction. Estimates indicate some evidence of pretrends with prices declining in anticipation of project completion particularly for the bottom income quartile. Pretrends are close to zero for the top two income quartiles. Declines in prices occur sharply after scheduled construction and are most pronounced in the 48 months following scheduled construction for the top two income quartiles.

One concern with this approach is that our method of imputing the expected month of project completion for planned, but unconstructed may introduce noise into the estimates. As a robustness exercise, Appendix~\ref{section:appendixpricedd} provides differences-in-differences estimates of price effects using only constructed projects and comparing properties near and far from project boundaries before and after construction. Results from this exercise (Table~\ref{table:priceDD\_het}) closely match triple-difference estimates (Table~\ref{table:priceDDD\_het}).

% Paragraph

% COMPARE PRICE ESTIMATES TO DIAMOND PLEASE!

\section{Discussion and Conclusion}\label{section:discussion}

Sections

\begin{enumerate}

\item Policymaker goals (especially in South Africa, but elsewhere as well)

\begin{itemize}

\item Goal 1: Improve quality of dwellings in project footprints [quote]

\begin{itemize}

\item This was largely a success

\item It did not have to be: for example, areas without projects could have developed even faster (and without big backyards that invite informal settlements)

\end{itemize}

\item Goal 2: Crowd-in local housing investment [quote]

\begin{itemize}

\item Not a success; little increase in quantity or prices of formal housing outside of footprints

\item Possibly because formal markets interpret housing projects as negative amenities

\end{itemize}

\item Goal 3: Reduce slum growth [quote]

\begin{itemize}

\item Mixed results: no reduction in footprints because of backyarding, but decreases outside of project footprints

\item Welfare effects are negative so governments would have to believe in very large negative externalities from slums in order to still justify this policy

\end{itemize}

\end{itemize}

\item Lessons from comparing developing and developed country evaluations of place-based policies

\begin{itemize}

\item Formal markets behave similarly between developing and developed contexts (similar results to \cite{diamond2016wants})

\item Informal housing grows within project areas, changing the spatial definition of ``treatment'' areas

\item The quantity of informal housing is also elastic and can adjust rapidly in response to place-based policies (compare to Glaeser and Gyorko?)

\end{itemize}

\item Limitations

\begin{itemize}

\item We don't have much to say about individual household outcomes and therefore, about the distributional consequences of these policies

\item We don't look at whether these big projects also crowd-in investment/jobs (although partially in identification)

\item We don't micro-found externalities well, especially for informal settlements

\end{itemize}

\end{enumerate}

We could also go with stuff from this ending:

Our results present a puzzle: large improvements in housing quality and infrastructure access within projects do not appear to generate new investments in formal housing development, influxes of high-income residents, or increases in housing prices nearby these projects. Additionally, we find no evidence that low-income households differentially sort into project areas or that projects increase the total supply of housing --- both mechanisms which may counteract any positive amenity effects. We offer two partial explanations to account for this puzzle.

In making housing decisions, people in South Africa may simply have weak preferences for neighborhood quality. Given very high rates of unemployment especially in rural areas, many South Africans migrate to cities solely for employment, sending large shares of their income home to their families. In the census data, we observe that over 43\% of our sample reports being born outside the province of Gauteng. Young adults regularly engage in circular migration, returning periodically to rural villages to spend their leisure time. Rather than investing in the local community and developing a strong social network, people may view their housing decision in terms of short-term access to employment. Therefore, people may place little value on nearby neighborhood quality or other amenities, instead prioritizing low rents and short commutes.

Another explanation is that households value neighborhood quality, but higher densities of informal backyard shacks offset any improvements in formal housing quality and infrastructure. \cite{Brueckner2018backyarding} document how households in backyard shacks often share water, electricity, and sanitation services with their landlords. Congestion in using shared services may lead households to consume less of these services and exacerbate externalities associated with these low levels of usage. \cite{Brueckner2018backyarding} also emphasize that backyard shacks often house young-adult, migrant workers who may be less inclined to make long-term social or economic investments in their communities.

\begin{comment}

% Family members of project beneficiaries may also move locally to maintain social ties (\cite{franklin2016enabled}).

These results contrast with similar findings in both developed and developing country settings. - what papers to discuss?

Our interpretation is relatively neutral from a policy perspective: housing projects can produce enough positive amenities to mediate any housing supply effects. Additionally, these projects effectively replace informal settlements with high quality formal houses, achieving a signature goal of South Africa's policy. At the same time, there is little evidence supporting housing policy as an important engine for local economic growth and development. These results may at least mediate any concerns policymakers may have for adversely affecting communities with housing projects, despite sometimes loud complaints from nearby community members.

It is important to emphasize that our analysis does not currently speak to the overall welfare implications of these policies. Since project houses cannot be sold, we have difficulty measuring the extent to which households value these houses.

Similarly, we do not measure the general equilibrium effects of housing policy outside of project neighborhoods; public housing may lead to an overall net reduction in slums in the city, which may generate total welfare gains through fewer congestion externalities or other mechanisms.

In light of South Africa's emphasis on using housing policy to stimulate neighborhood development, we hope that this paper will be useful for policymakers in considering the range of informal housing market responses as they design new housing policies.

% We find that South African housing projects greatly improve the housing conditions within their footprints. Housing projects successfully deliver large numbers of new formal structures (Figure~\ref{figure:dddformal}) and much better access to bulk services (Table~\ref{table:censusestimates}). At the same time, formal project housing appears to crowd-out growth in informal settlements so that the total density of housing remains the same as in unconstructed project areas (Figure~\ref{figure:dddtotal}). Moreover, results in Table~\ref{table:censusestimates} indicate that the average quality of new project houses are at least as good as neighboring houses at baseline. Therefore consistent with the model of amenity effects in \cite{diamond2016wants}, we would expect to find that these programs increase nearby home prices; however, results in Table~\ref{table:priceest} report if anything the opposite trend. Similarly, we do not observe large new developments in either informal or formal housing just nearby these projects.

% The lack of significant spillover effects in price as well as in quantity is consistent with countervailing amenity and housing supply effects. While an influx of new houses threatens to decrease prices and housing supply nearby, positive amenities provided by housing projects may also attract greater housing investment, increasing prices. While both amenity and supply effects predict increases in local population, there may be spatial or geographic constraints on population growth given that these settlements in South Africa are almost exclusively single-story and informal settlements often fill any remaining surface area.

%\section{Conclusion}\label{section:conclusion}

%This paper serves as a cautionary tale for low-density public housing in developing countries. While these projects may be designed to improve local housing markets and reduce local slum growth, they may instead support local slum growth through a mix of available land and improved infrastructure. It is important to emphasize that our analysis does not currently speak to the overall welfare implications of these policies. For example, public housing may lead to an overall net reduction in slums in the city, which may generate total welfare gains through fewer congestion externalities or other mechanisms. According to our model, the optimal housing policy depends crucially on the shape of these congestion externalities. We look forward to using variation in the South African housing program to estimate these externalities in future work.

% (1) construction costs, (2) value to recipients, (3) value to non-recipients within project footprints, and (4) spillover effects on neighboring areas.

\end{comment}

\pagebreak

\nocite{\*}

\singlespacing

\setlength{\bibsep}{7pt}

\bibliographystyle{abbrvnat}

\bibliography{ref}

\pagebreak

% APPENDIX

\appendix

\doublespacing

\section{Appendix}

\titleformat\*{\subsection}{\centering\normalsize\bfseries}

\begin{figure}[t]

\caption{Transaction Price Histogram}\label{figure:transactionhist}

\centering

\includegraphics[scale=.4,trim={0cm 0cm -1cm 0cm}]{figures/summary\_pricedist.pdf}

\\

\footnotesize{Note: Transactions are censored at R100,000.}

\end{figure}

\subsection{Project Descriptions}\label{appendix:projectdescriptions}

\begin{table}[ht!]

\centering

\caption{Project Descriptions}\label{table:projectdescriptions}

\vspace{-2mm}

\begin{tabular}{l\*{1}{c}}

\toprule

&Counts \\

\midrule

\textbf{Constructed} & \\[.5em]

\input{figures/proj\_con\_table}[.5em]

\input{figures/proj\_con\_table\_total}[.5em]

\textbf{Unconstructed} & \\[.5em]

\input{figures/proj\_uncon\_table}[.5em]

\input{figures/proj\_uncon\_table\_total}[.5em]

\textbf{Other} \\[.5em]

\input{figures/other\_table}[.5em]

\input{figures/other\_table\_total}[.5em]

\midrule

\input{figures/proj\_table\_total}[.5em]

\bottomrule

\end{tabular}

\end{table}

\subsection{String Matching for Project Names and Delivery Dates }

\label{appendix:stringmatch}

We then use a fuzzy-string matching algorithm with bigrams to link project names from the budget reports to the administrative maps. We keep all matches with over 60\% similarity, with 37 projects matching exactly. Appendix~\ref{table:stringmatch} compares unmatched and matched projects finding that matched projects have much higher numbers of project houses, lower nearby housing prices, and are further from the central business district compared to unmatched projects. One reason may be that the budget reports only include larger, more expansive projects, which are likely to take place further from city centers.

\vspace{0mm}

\begin{table}[ht!]

\centering

\caption{Assessing Name Matching between Budget Reports and Administrative Project Records}\label{table:stringmatch}

\vspace{0mm}

\begin{tabular}{l\*{1}{cccc}}

\toprule

& \multicolumn{2}{c}{\textbf{Matched}}& \multicolumn{2}{c}{\textbf{Unmatched}} \\

&Const. & Unconst. &Const. & Unconst. \\

\midrule

\input{figures/score\_tablet\_4}

\bottomrule

\multicolumn{5}{l}{\scriptsize Const. refers to constructed projects and unconst. refers to unconstructed projects.}\\[-.5em]

\multicolumn{5}{l}{\scriptsize $^\*$Calculated from {\it expected} completion dates using Gauteng National Treasury budget reports.}\\[-.5em]

\multicolumn{5}{l}{\scriptsize $^\dagger$ The USD averaged to about 7.70 Rands during the 2001-2011 period.}\\[-.5em]

\multicolumn{5}{l}{\scriptsize $^\ddagger$Measured as the average minimum distance with respect to Johannesburg and Pretoria CBDs. } \\[-.5em]

\end{tabular}

\end{table}

\subsection{Transaction Frequency}

\label{appendix:histfreq}

\vspace{-5mm}

\begin{figure}[ht!]

\centering

\includegraphics[scale=.4 , trim={.2cm 0.2cm .2cm 0.2cm},clip]{figures/summary\_densitytime.pdf}

\end{figure}

\subsection{Project counts}

\label{appendix:projectcounts}

\vspace{-5mm}

\begin{figure}[ht!]

\centering

\includegraphics[scale=1.2,trim={.1cm 1.2cm .1cm 1.2cm},clip]{figures/projectcounts\_4.pdf}

\end{figure}

\begin{table}[ht!]

\small

\centering

\caption{Triple Difference Estimates of Housing Density }\label{table:bbluDDDfull}

\vspace{-2mm}

\resizebox{1\linewidth}{!}{

\begin{tabular}{lCCCCC}

\toprule

& \small (1) & \small (2) & \small (3) & \small (4) & \small (5) \\

& \small Total Housing & \small Formal Housing & \small Informal Housing & \small Informal Backyard Housing & \small Informal Non-Bkyd. Housing \\ \midrule

\input{figures/b\_ksp\_o0\_d250\_500}

\bottomrule

\multicolumn{6}{l}{\footnotesize Standard errors clustered at the project level in parenthesis. \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 }\\[-.3em]

\multicolumn{6}{l}{\footnotesize Project areas are 25$\times$25m grid cells within 1,500m of project footprints.}\\[-.3em]

\multicolumn{6}{l}{\footnotesize Housing densities are measure in terms of houses per $\text{km}^{2}$.}\\[-.3em]

\multicolumn{6}{l}{\footnotesize ``inside'' means within project footprint. ``constr'' means constructed.}

\end{tabular}

}

\end{table}

\begin{table}[ht!]

\caption{Building Density by Neighborhood Income Quartile}\label{table:bbluDDDhet}

\resizebox{1\linewidth}{!}{

\begin{tabular}{lDDDDD}

\toprule

& \small (1) & \small (2) & \small (3) & \small (4) & \small (5) \\

& Total Housing & Formal Housing &Informal Housing &Informal Non-Bkyd. Housing & Informal Bkyd. Housing \\ \midrule

\input{figures/b\_q\_ksp\_o0\_d250\_500\_top}

\bottomrule

\end{tabular}

}

\end{table}

\begin{table}[ht!]

\small

\centering

\caption{Full Census Household-level Estimates}\label{table:censusestimatesfull}

\vspace{-2mm}

\resizebox{1\linewidth}{!}{

\begin{tabular}{lDDDDDD}

\toprule

& \small (1) & \small (2) & \small (3) & \small (4) & \small (5) & \small (6) \\

& \small Is a Formal House & \small Has a Flush Toilet & \small Has Piped Water Indoors & \small Has Electricity & \small Total Rooms & \small Owns House \\ \midrule

\input{figures/ch1\_ksp\_o0\_d250\_500\_bb1} \midrule

% \input{figures/census\_hh\_DDregs\_AGG}

\bottomrule\\[-.6em]

\multicolumn{7}{l}{\footnotesize Standard errors clustered at the project level in parenthesis.\textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Outcomes in Columns (1) to (4), (6), and (9) to (11) are indicator variables.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Specifications include census 2001 subplace fixed effects.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Results are weighted by unit area and control for up to a cubic polynomial in unit area by time. }\\[-.3em]

\multicolumn{7}{l}{\footnotesize ``HH'' stands for household and ``HoH'' stands for head of household.}

\end{tabular}

}

\end{table}

\begin{table}[ht!]

\small

\centering

\caption{Full Census Household-level Estimates Cont.}\label{table:censusestimatesfull2}

\vspace{-2mm}

\resizebox{1\linewidth}{!}{

\begin{tabular}{lDDDDDD}

\toprule

& \small (7) & \small (8) & \small (9) & \small (10) & \small (11) & \small (12) \\

& \small HH Size & People per $\text{km}^{2}$ & Age HoH & HoH is African & HoH is Employed & HH Log Income \\ \midrule

\input{figures/ch2\_ksp\_o0\_d250\_500\_bb1}

% \input{figures/census\_hh\_DDregs\_AGG}

\bottomrule\\[-.6em]

\multicolumn{7}{l}{\footnotesize Standard errors clustered at the project level in parenthesis.\textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Outcomes in Columns (1) to (4), (6), and (9) to (11) are indicator variables.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Specifications include census 2001 subplace fixed effects.} \\[-.3em]

\multicolumn{7}{l}{\footnotesize Results are weighted by unit area and control for up to a cubic polynomial in unit area by time. }\\[-.3em]

\multicolumn{7}{l}{\footnotesize ``HH'' stands for household and ``HoH'' stands for head of household.}

\end{tabular}

}

\end{table}

\subsection{Full Housing Demand Estimates}

\begin{table}[h]

\centering

\caption{Full Housing Demand Estimates}\label{table:housingdemandestimates\_full}

\vspace{-2mm}

\resizebox\*{!}{.97\textheight}{

\begin{tabular}{lDD}

\toprule

& Formal & Informal \\ \midrule

net amenities: $( \delta\_{hl} - \theta c^{u}\_{hlt} )$ \\[.3em]

\input{figures/forimpcmp10\_1\_full}[.3em]

\input{figures/forimpcmp10\_2\_full}[.5em]

\input{figures/forimpcmp10\_3\_1\_full}

\input{figures/forimpcmp10\_3\_2\_full}

\input{figures/forimpcmp10\_3\_3\_full}

\input{figures/forimpcmp10\_3\_4\_full}

\input{figures/forimpcmp10\_3\_5\_full}

\input{figures/forimpcmp10\_3\_6\_full}

\input{figures/forimpcmp10\_3\_7\_full}

\input{figures/forimpcmp10\_3\_8\_full}

\input{figures/forimpcmp10\_3\_9\_full}

\input{figures/forimpcmp10\_3\_10\_full}[.3em]

% \input{figures/forimpcmp10\_3\_6\_full}

correlation of amenity shocks: $\rho\_{for,inf}$ & $0.620^{\text{a}}$ & \\

& (0.018) & \\[.3em]

N & 6,143,756 & \\

\bottomrule\\[-.6em]

\multicolumn{3}{l}{ \footnotesize Standard errors clustered at the project level in parenthesis. \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 } \\[-.3em]

\multicolumn{3}{l}{ \footnotesize Outcome categories by plot and year include categories for formal and informal} \\[-.3em]

\multicolumn{3}{l}{ \footnotesize building counts from 0 to 9 and a category at least 10. Table~\ref{table:housingdemandestimates\_full} includes the full set} \\[-.3em]

\multicolumn{3}{l}{ \footnotesize of coefficients. Net amenities equal local amenities minus local housing costs. } \\[-.3em]

\end{tabular}

}

\end{table}

\begin{table}[h]

\centering

\caption{Housing Demand Estimates by Income Quartile}\label{table:housingdemandestimatesinc}

\vspace{-2mm}

\begin{tabular}{lGG}

\toprule

& Formal Housing & Informal Housing \\ \midrule

Net amenities: $( \delta\_{hl} - \theta c^{u}\_{hlt} )$ \\[.3em]

\input{figures/forimpcmp10\_1\_q1}[.3em]

\input{figures/forimpcmp10\_1\_q2}[.3em]

\input{figures/forimpcmp10\_1\_q3}[.3em]

\input{figures/forimpcmp10\_1\_q4}[.3em]

\input{figures/forimpcmp10\_2\_q}[.3em]

\\

Correlation of amenity shocks: $\rho\_{for,inf}$ & $0.620^{\text{a}}$ & \\[-.3em]

& (0.018) & \\[.3em]

N & 6,143,756 & \\

\bottomrule\\[-.8em]

% Implied amenity value net of housing costs in Rands: $\frac{1}{\theta\_{for}}(\delta\_{hl}- \theta\_{for}c^{u}\_{hlt})$ & & \\

% inside project &\input{figures/for\_proj} & \input{figures/inf\_proj} \\

% 0-500m outside &\input{figures/for\_spill} & \input{figures/inf\_spill} \\

% \bottomrule\\[-.8em]

\multicolumn{3}{l}{ \footnotesize Standard errors clustered at the project level in parenthesis. \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 } \\[-.3em]

\multicolumn{3}{l}{ \footnotesize Outcome categories include formal and informal building counts from 0 to 9 } \\[-.3em]

\multicolumn{3}{l}{ \footnotesize and at least 10 by plot and year. Table~\ref{table:housingdemandestimates\_full} includes the full set of coefficients. } \\[-.3em]

\multicolumn{3}{l}{ \footnotesize Net amenities equal local amenities minus local housing costs. } \\[-.3em]

\end{tabular}

\end{table}

\begin{table}[h]

\centering

\caption{Simulated Effects Housing Projects on Net Prices by Neighborhood Income Quartile}\label{table:housingdemandsim}

\vspace{-2mm}

\begin{tabular}{llDD}

\toprule

& Area & Formal Housing & Informal Housing \\ \midrule

All & inside project & \input{figures/for\_proj\_w} & -\input{figures/inf\_proj\_w} \\[.3em]

& 0-500m outside & -\input{figures/for\_spill\_w} & -\input{figures/inf\_spill\_w} \\[1em]

% By Neighborhood Income Quartile & & \\[.5em]

Quartile \\

\midrule

Q1 & inside project & \input{figures/for\_proj\_w\_q1} & \input{figures/inf\_proj\_w\_q1} \\[.3em]

& 0-500m outside & \input{figures/for\_spill\_w\_q1} & \input{figures/inf\_spill\_w\_q1} \\[.8em]

Q2 & inside project & \input{figures/for\_proj\_w\_q2} & \input{figures/inf\_proj\_w\_q2} \\[.3em]

& 0-500m outside & \input{figures/for\_spill\_w\_q2} & \input{figures/inf\_spill\_w\_q2} \\[.8em]

Q3 & inside project & \input{figures/for\_proj\_w\_q3} & \input{figures/inf\_proj\_w\_q3} \\[.3em]

& 0-500m outside & \input{figures/for\_spill\_w\_q3} & \input{figures/inf\_spill\_w\_q3} \\[.8em]

Q4 & inside project & \input{figures/for\_proj\_w\_q4} & \input{figures/inf\_proj\_w\_q4} \\[.3em]

& 0-500m outside & \input{figures/for\_spill\_w\_q4} & \input{figures/inf\_spill\_w\_q4} \\

\bottomrule\\[-1em]

\multicolumn{4}{l}{ \footnotesize All results are in Rands per plot. Average housing price } \\[-.3em]

\multicolumn{4}{l}{ \footnotesize is 231,000 Rands. Quartiles are measured with respect to } \\[-.3em]

\multicolumn{4}{l}{ \footnotesize average neighborhood household income in the 2001 census. } \\[-.3em]

\multicolumn{4}{l}{ \footnotesize Appendix~\ref{section:appendixwelfaresim} describes the simulation in more detail. } \\[-.3em]

\end{tabular}

\end{table}

\subsection{Price Estimates by Distance to Project Boundary}\label{section:appendixpricedist}

We examine how prices change differentially over time for constructed relative to unconstructed projects at different distances from project boundaries using the following specification:

\begin{align\*}

\quad \text{Log-Price}\_{ipnt} \, =& \, \beta\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \beta\_2 \, \textsc{\small Post}\_{t} \, + \, \beta\_3 \, \textsc{\small Const}\_{p}\, \\

&\, + \sum\_{d=1}^5 \,D\_i^{d} \, \Big( \, \alpha\_1^{d} \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \alpha\_2^{d} \, \textsc{\small Post}\_{t} \, + \, \alpha\_3^{d} \, \textsc{\small Const}\_{p}\, +\, \alpha\_4^{d} \Big) \, \\

&\, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad \\

\text{where}& \\

& D\_i^{d} = \mathbbm{1} \Big\{ - 250 \leq dist\_i -(250\*d) \leq 250 \Big\}

\end{align\*}

\noindent where distance to project boundary, $d$, ranges from 0 to 1,200 meters in intervals of 250 meters. $\textsc{\small Const}\_{p}$ is an indicator for being near a constructed project, and $\textsc{\small Post}\_{t}$ is an indicator that takes a value of one for properties sold after scheduled construction date for each project. Coefficient, $\beta\_1$ captures average differential changes in log-prices for constructed and unconstructed projects pre and post scheduled construction located in neighborhood income quartile $q$. $D\_i^{d}$ takes a value of one for properties, $i$, in 250 meter intervals, $d$, from 0 to 1,250 meters. The coefficients of interest, $\alpha\_1^{q,d}$ capture the differential changes in log-prices between constructed and unconstructed projects pre and post scheduled construction for each distance bin $d$ relative to similar changes for properties located in the omitted distance category, which is the furthest distance bin (1,250 to 1,500 meters). $\alpha\_1^{d}$ coefficients track how log-price changes evolve over distance to project boundaries. Additional controls $X\_{ipnt}$ include up to a cubic polynomial in the lot size of properties as well as a fixed effects for the year-month of each transaction. The specification also includes neighborhood fixed effects $\lambda\_n$. Figure~\ref{figure:pricesoverdistance} plots $\alpha\_1^{d}$ coefficients with respect to distance to the nearest project boundary.

\begin{figure}[!htb]

{ \centering

\caption[ Log-Price Estimates by Distance to Nearest Project Boundary ]

{\small Log-Price Estimates by Distance to Nearest Project Boundary }\label{figure:pricesoverdistance}

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_dist\_3d\_ctrl\_q\_one.pdf}

}

Estimates include differential changes in prices before and after (construction) for constructed relative to unconstructed project areas. Estimates are normalized with respect to price changes in the furthest distance group --- 1,250 to 1,500 meters from project boundaries. The y-axis is in terms of changes in log-prices.

\end{figure}

In Figure~\ref{figure:pricesoverdistance}, we observe low price changes nearby project boundaries, which rise moving to 750 meters from project boundaries before falling back to zero price change at further distances. Overall, this pattern does not provide evidence of a clear, monotonic gradient in housing price changes with respect to distance to housing projects.

We also explore how these effects vary according to neighborhood income quartile in the following specification:

\begin{align\*}

\quad \text{Log-Price}\_{ipnt} \, =& \sum\_{q=1}^{4} \, Q\_{n}^{q} \, \Bigg[ \, \, \beta\_1^{q} \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \beta\_2^{q} \, \textsc{\small Post}\_{t} \, + \, \beta\_3^{q} \, \textsc{\small Const}\_{pq}\, \\

&\, + \sum\_{d=1}^{5} \,D\_i^{d} \, \Big( \, \alpha\_1^{q,d} \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{pq} \, + \, \alpha\_2^{q,d} \, \textsc{\small Post}\_{t} \, + \, \alpha\_3^{q,d} \, \textsc{\small Const}\_{pq}\, +\, \alpha\_4^{q,d} \Big) \, \Bigg] \\

&\, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad \\

\text{where}& \\

& D\_i^{d} = \mathbbm{1} \Big\{ - 250 \leq dist\_i -(250\*d) \leq 250 \Big\}

\end{align\*}

\noindent where $q$ indexes quartiles from 1 to 4 and $Q\_{n}^{q}$ takes a value of one if average income in neighborhood $n$ falls within quartile $q$. Figure~\ref{figure:pricesoverdistancehet} plots $\alpha\_1^{d,q}$ coefficients with respect to distance to the nearest project boundary across income quartiles.

\begin{figure}[!htb]

{ \centering

\caption[ Log-Price Estimates by Distance to Nearest Project Boundary by Neighborhood Income Quartile ]

{\small Log-Price Estimates by Distance to Nearest Project Boundary by Neighborhood Income Quartile}\label{figure:pricesoverdistancehet}

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_dist\_3d\_ctrl\_q\_comb.pdf}

}

Estimates include differential changes in prices before and after scheduled construction relative to unconstructed project areas. Estimates are normalized with respect to price changes in the furthest distance group --- 1,250 to 1,500 meters from project boundaries. Q1 to Q4 index neighborhood income quartiles measured using the 2001 census of population and housing. The y-axis is in terms of changes in log-prices and the x-axis measures meters outside of the nearest housing project boundary.

\end{figure}

In Figure~\ref{figure:pricesoverdistancehet}, the poorest quartile (Q1) exhibits a decreasing gradient consistent with an appreciation in prices close to project boundaries. The second and third quartiles (Q2 and Q3) provide little evidence of a clear pattern betwene prices and distance. The richest quartile (Q4) finds an increasing gradient consistent with large drops in prices nearby constructed projects.

\subsection{Price Estimates by Months to Project Construction}\label{section:appendixpricetime}

We examine how prices change differentially for properties less than 500 meters outside of boundaries compared to projects between 500 and 1,500 meters outside of boundaries and between constructed and unconstructed project areas. We further plot these differences with respect to months to the (expected) construction month of each project across neighborhood income quartiles using the following specification:

\begin{align\*}

\quad \text{Log-Prices}\_{ipnt} \, =& \, \beta\_1 \, \textsc{\small Spill}\_{i}\times\textsc{\small Const}\_{p} \, + \, \beta\_2 \, \textsc{\small Spill}\_{i} \, + \, \beta\_3 \, \textsc{\small Const}\_{p}\, \\

&\, + \sum\_{l=1}^{5} T\_{l} \, \Big( \, \alpha\_1^{l} \, \textsc{\small Spill}\_{i}\times\textsc{\small Const}\_{p} \, + \, \alpha\_2^{l} \, \textsc{\small Spill}\_{i} \, + \, \alpha\_3^{l} \, \textsc{\small Const}\_{p}\, +\, \alpha\_4^{l} \Big) \, \\

&\, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad \\

\text{Where }& \\

&T\_1 = \mathbbm{1} \Big\{\, t \leq -48 \,\Big\} \,, \,\,\, T\_2 = \mathbbm{1} \Big\{ -24 \leq t < -48 \Big\} \,\,, \\

&T\_3 = \mathbbm{1} \Big\{ 0< t<=24 \Big\} \,\, ,\,\,T\_4\,=\,\mathbbm{1} \Big\{ 24< t <=48 \Big\} \, , \,\,\, T\_5 = \mathbbm{1} \Big\{ 48 \leq t \Big\}

\end{align\*}

\noindent where $\textsc{\small Spill}\_{i}$ is an indicator for properties within 500 meters outside of a project boundary and $\textsc{\small Const}\_{p}$ is an indicator for being nearby a project that is successfully constructed. $t$ indexes months to scheduled construction with the scheduled construction month normalized to zero. Differential prices near and far from constructed projects compared to unconstructed projects are estimated separately according to (1) all transactions more than 48 months before scheduled construction, (2) transactions between 48 and 24 months prior to scheduled construction, (3) transactions within 24 months after scheduled construction, (4) transactions within 24 months to 48 months after scheduled construction, and (5) all transactions over 48 months after scheduled construction. This specification leaves transactions occurring in the 24 months immediately before scheduled construction as the omitted category. Coefficient, $\beta\_1$ captures differential changes in log-prices near and far from constructed and unconstructed projects across all time periods. The coefficients of interest, $\alpha\_1^{l}$ track how prices change differentially in the 5 time windows relative to the omitted category --- 24 months before scheduled construction. Additional controls $X\_{ipnt}$ include up to a cubic polynomial in the lot size of properties as well as a fixed effects for the year-month of each transaction. The specification also includes neighborhood fixed effects $\lambda\_n$. Figure~\ref{figure:pricesovertime} plots $\alpha\_1^{l}$ coefficients with respect to months to scheduled construction.

Figure~\ref{figure:pricesovertime} finds a negative price differential 48 or more months before construction, which jumps above zero 48 to 24 months before scheduled construction. The results support a declining pretrend with prices dropping by 10\% in the 48 months before scheduled construction. This trend continues with a similar slope in the 48 months after scheduled construction, reaching a 20\% decline in prices between 24 and 48 months after construction. Over 48 months after construction the price difference remains negative.

\begin{figure}[!htb]

{ \centering

\caption[ Log-Prices By Months to Scheduled Construction ]

{\small Log-Prices By Months to Scheduled Construction }\label{figure:pricesovertime}

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_time\_3d\_ctrl\_q\_one.pdf}

}

Estimates capture differential changes in log-prices for properties within 500m of project boundaries compared to properties located between 500m to 1,500m from project boundaries for constructed relative to unconstructed projects. These estimates are computed at different time periods relative to the scheduled construction month of each project with price changes in the 24 months before scheduled construction normalized to zero. The y-axis is in terms of changes in log-prices.

\end{figure}

Figure~\ref{figure:pricesovertime} finds a negative price differential 48 or more months before construction, which recovers to above zero in 48 to 24 months before scheduled construction. In the 48 months after scheduled construction, the price difference dips increasingly below zero. Over 48 months after construction the price difference remains negative.

We also explore how these effects vary according to neighborhood income quartile in the following specification:

\begin{align\*}

\quad \text{Log-Prices}\_{ipnt} \, =& \sum\_{q=1}^{4} \, Q\_{n}^{q} \, \Bigg[ \, \, \beta\_1^{q} \, \textsc{\small Spill}\_{iq}\times\textsc{\small Const}\_{pq} \, + \, \beta\_2 \, \textsc{\small Spill}\_{iq} \, + \, \beta\_3 \, \textsc{\small Const}\_{pq}\, \\

&\, + \sum\_{l=1}^{5} T\_{l} \, \Big( \, \alpha\_1^{q,l} \, \textsc{\small Spill}\_{iq}\times\textsc{\small Const}\_{pq} \, + \, \alpha\_2^{q,l} \, \textsc{\small Spill}\_{iq} \, + \, \alpha\_3^{q,l} \, \textsc{\small Const}\_{pq}\, +\, \alpha\_4^{q,l} \Big) \, \Bigg] \\

&\, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad \\

\text{Where }& \\

&T\_1 = \mathbbm{1} \Big\{\, t \leq -48 \,\Big\} \,, \,\,\, T\_2 = \mathbbm{1} \Big\{ -24 \leq t < -48 \Big\} \,\,, \\

&T\_3 = \mathbbm{1} \Big\{ 0< t<=24 \Big\} \,\, ,\,\,T\_4\,=\,\mathbbm{1} \Big\{ 24< t <=48 \Big\} \, , \,\,\, T\_5 = \mathbbm{1} \Big\{ 48 \leq t \Big\}

\end{align\*}

\noindent where $q$ indexes quartiles from 1 to 4 and $Q\_{n}^{q}$ takes a value of one if average income in neighborhood $n$ falls within quartile $q$. Figure~\ref{figure:pricesovertimehet} plots $\alpha\_1^{d,q}$ coefficients with respect to time to scheduled construction across income quartiles.

Figure~\ref{figure:pricesovertimehet} provides evidence of a strong decreasing pretrend in prices for the first quartile, dropping from price increases of around 100\% over 48 months before scheduled construction. By contrast, the fourth quartile exhibits little pretrend in anticipation of scheduled construction. Following scheduled construction, prices following a slight upward trend in the first quartile. Moving from the second to the fourth quartiles, trends in prices after scheduled construction become more and more negative. Prices for the third and fourth quartiles drop substantially in the 48 months reaching just after scheduled construction before partially rebounding more than 48 months after scheduled construction.

\begin{figure}[!htb]

{ \centering

\caption[ Log-Prices By Months to Scheduled Construction by Neighborhood Income Quartile ]

{\small Log-Prices By Months to Scheduled Construction by Neighborhood Income Quartile }\label{figure:pricesovertimehet}

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/price\_time\_3d\_ctrl\_q\_comb.pdf}

}

Estimates capture differential changes in log-prices for properties within 500m of project boundaries compared to properties located between 500m to 1,500m from project boundaries for constructed relative to unconstructed projects. These price effects are computed at different time periods relative to the scheduled construction month of each project with price effects in the 24 months before scheduled construction normalized to zero. Q1 to Q4 index neighborhood income quartiles measured using the 2001 census of population and housing. The y-axis is in terms of changes in log-prices and the x-axis measures months to scheduled construction.

\end{figure}

\subsection{Difference-in-Differences Estimates on Log-Prices}\label{section:appendixpricedd}

As a robustness exercise, we estimate a differences-in-differences specification comparing properties close and far from project boundaries before and after scheduled construction only for projects that were successfully constructed. We implement the following specification:

\begin{align}\label{eq:mainpricedd}

\begin{split}

\quad y\_{ipnt} \, =& \,\, \beta\_1 \textsc{\small Spill}\_{ip}\times\textsc{\small Post}\_{t} \, + \, \beta\_2 \, \textsc{\small Post}\_{t} \, + \, \beta\_3 \, \textsc{\small Spill}\_{ip} \, + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align}

We also estimate this differences-in-differences specification separately according to neighborhood income quartile with the following specification:

\begin{align}\label{eq:mainpricehetdd}

\begin{split}

\quad y\_{ipnt} \, =& \,\, \sum\_{q=1}^{4} Q\_{n}^{q} \Bigg[ \beta\_1^{q} \textsc{\small Spill}\_{ipq}\times\textsc{\small Post}\_{t} \, + \, \beta\_2^{q} \, \textsc{\small Post}\_{t} \, + \, \beta\_3^{q} \, \textsc{\small Spill}\_{ipq} \, \Bigg] + \, \theta \, X\_{ipnt} \, + \, \lambda\_n \, + \, \varepsilon\_{ipnt} \quad

\end{split}

\end{align}

\noindent where $q$ indexes quartiles from 1 to 4 and $Q\_{n}^{q}$ takes a value of one if average income in neighborhood $n$ falls within quartile $q$.

Table~\ref{table:priceDD\_het} provides estimates across all neighborhoods and by income quartile separately, finding similar results to triple-difference results in Table~\ref{table:priceDDD\_het}.

\begin{table}

\small

\centering

\caption{Differences-in-Differences Estimates of Log-Prices 0 to 500 Meters Outside of Housing Projects}\label{table:priceDD\_het}

\vspace{-2mm}

\begin{tabular}{lCC}

\toprule

& \small (1) & \small (2) \\ \midrule

\input{figures/price\_temp\_Tester\_2d\_spill\_top\_no\_lab} \\

By Neighborhood Income Quartile \\

\input{figures/price\_temp\_Tester\_2d\_q\_top}

\bottomrule

\multicolumn{3}{l}{\footnotesize Estimates capture differential price effects for properties } \\[-.3em]

\multicolumn{3}{l}{\footnotesize 0 to 500 meters outside of project boundaries relative to } \\[-.3em]

\multicolumn{3}{l}{\footnotesize properties 500 to 1,500 meters from boundaries before and } \\[-.3em]

\multicolumn{3}{l}{\footnotesize after scheduled construction. } \\[-.3em]

% \multicolumn{3}{l}{\footnotesize } \\[-.3em]

\multicolumn{3}{l}{\footnotesize Income quartiles are from average household 2001 incomes in } \\[-.3em]

\multicolumn{3}{l}{\footnotesize census 2001 subplaces. All estimates include neighborhood fixed effects.} \\[-.3em]

\multicolumn{3}{l}{\footnotesize Standard errors clustered at the project level in parenthesis.} \\[-.3em]

\multicolumn{3}{l}{\footnotesize \textsuperscript{c} p$<$0.10,\textsuperscript{b} p$<$0.05,\textsuperscript{a} p$<$0.01 } \\[-.3em]

\end{tabular}

\end{table}

\subsection{Building Histogram}\label{section:appendixbuildhist}

\begin{figure}[!htb]

{ \centering

\caption[ Histogram of Buildings per Plot with at least 1 Building ]

{\small Histogram of Buildings per Plot with at least 1 Building }\label{figure:buildhist}

\includegraphics[width=\textwidth,trim={0.3cm .3cm 0.1cm 0cm}, clip=true]{figures/building\_hist.pdf}

Plots with zero buildings are not shown. 86\% of plots have zero formal houses, and 91\% of plots have zero informal houses.

This figure also excludes plots with over 15 buildings, which compose less than 0.001\% of the sample.

}

\end{figure}

\subsection{Welfare Simulation}\label{section:appendixwelfaresim}

To simulate counterfactual housing choices, we begin by taking 10,000 random draws from a standard normal distribution, which represent plot-specific random amenity shocks. The simulation sets housing costs, $C$ equal to the average housing price, R231,000, and sets $\theta$ equal to 0.0000094, which is the marginal utility of income estimate for formal housing. Estimated cutpoints, $\Lambda\_{h}(k)$ allow us to recover the implied congestion cost $\lambda\_{h}(k)$, assuming a reservation utility, $\overline{U}$ equal to zero. Net amenities are are given by

\begin{align\*}

\overline{V}\_{hlt} &= \textsc{\small Proj}\_{lp} \,\, \Big( \alpha^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \alpha^{Vh}\_2 \, \textsc{\small Post}\_{t} \, + \, \alpha^{Vh}\_3 \, \textsc{\small Const}\_{p}\, +\, \alpha^{Vh}\_4 \Big) \, + \\[.2em]

& \, \textsc{\small Spill}\_{lp} \, \Big( \beta^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \beta^{Vh}\_2 \, \textsc{\small Post}\_{t} \, + \, \beta^{Vh}\_3 \, \textsc{\small Const}\_{p} \, +\, \beta^{Vh}\_4 \Big) \, + \\[.2em]

& \, \gamma^{Vh}\_1 \, \textsc{\small Post}\_{t}\times\textsc{\small Const}\_{p} \, + \, \gamma^{Vh}\_2 \,\textsc{\small Post}\_{t} \, + \, \gamma^{Vh}\_3 \, \textsc{\small Proj}\_{p}

\end{align\*}

\noindent where the change in welfare is computed by shifting $\alpha^{Vh}\_1$ and $\beta^{Vh}\_1$ from zero to their empirical estimates for project and spillover areas respectively. Given the full set of parameters, we are able to simulate optimal building choices, $k^{\*}\_{h}$ as well as prices, $P^{\*}\_{h}$ before and after project construction.

\end{document}